Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: A systematic review

Daniel Sanchez-Morillo¹, Miguel A Fernandez-Granero¹ and Antonio Leon-Jimenez²

Abstract
Major reported factors associated with the limited effectiveness of home telemonitoring interventions in chronic respiratory conditions include the lack of useful early predictors, poor patient compliance and the poor performance of conventional algorithms for detecting deteriorations. This article provides a systematic review of existing algorithms and the factors associated with their performance in detecting exacerbations and supporting clinical decisions in patients with chronic obstructive pulmonary disease (COPD) or asthma. An electronic literature search in Medline, Scopus, Web of Science and Cochrane library was conducted to identify relevant articles published between 2005 and July 2015. A total of 20 studies (16 COPD, 4 asthma) that included research about the use of algorithms in telemonitoring interventions in asthma and COPD were selected. Differences on the applied definition of exacerbation, telemonitoring duration, acquired physiological signals and symptoms, type of technology deployed and algorithms used were found. Predictive models with good clinically reliability have yet to be defined, and are an important goal for the future development of telehealth in chronic respiratory conditions. New predictive models incorporating both symptoms and physiological signals are being tested in telemonitoring interventions with positive outcomes. However, the underpinning algorithms behind these models need be validated in larger samples of patients, for longer periods of time and with well-established protocols. In addition, further research is needed to identify novel predictors that enable the early detection of deteriorations, especially in COPD. Only then will telemonitoring achieve the aim of preventing hospital admissions, contributing to the reduction of health resource utilization and improving the quality of life of patients.

Keywords
Algorithms, asthma, chronic obstructive pulmonary disease, decision support systems, exacerbations, hospitalization/statistics, machine learning, prediction, predictive analytics, physiological measurements, predictive analytics, pulmonary disease, telemedicine, telemonitoring

Introduction
Asthma and chronic obstructive pulmonary disease (COPD) have attracted research interest as a major public health problem of increasing concern to healthcare systems worldwide because of high prevalence¹ and rising socioeconomic burden.²,³

Reducing the impact of exacerbations through the early recognition of symptoms and prompt treatment may reduce the risk of hospitalization, improve

¹ Biomedical Engineering and Telemedicine Research Group, University of Cádiz, Puerto Real, Cádiz, Spain
² Pulmonology, Allergy and Thoracic Surgery Unit, Puerta del Mar University Hospital, Cádiz, Spain

Corresponding author: Daniel Sánchez-Morillo, University of Cádiz, Puerto Real, Cádiz, Spain.
Email: daniel.morillo@uca.es
health-related quality of life (HRQoL) and control the burden of COPD. In asthma, self-management involving prompt intervention in line with a personalized asthma action plan reduces acute exacerbations. Education, self-monitoring and regular review are recommended to improve outcomes and prevent exacerbations in both conditions. New information and communication technologies and telemonitoring can potentially support patient self-management.

Home telemonitoring encompasses the non-invasive exchange of information using electronic devices and telecommunication technologies and allows clinical data to be collected remotely on a routine and regular basis. Telehealth-enabled chronic care management services are promoted as being effective in supporting people with long-term conditions at home, and the telemonitoring of patients with chronic respiratory conditions has become a major research field of the respiratory community because of its potential to promote self-management, improve control, increase quality of life and prevent admissions.

Some authors have argued that telemonitoring is a promising alternative – or adjunct – to provision of traditional healthcare services both in asthma and COPD. However, although some studies have shown that telemonitoring may improve some clinical outcomes and reduce healthcare costs, the effects of telehealth interventions on emergency department attendance, hospital admissions, duration of admissions, HRQoL, costs and mortality remain less certain.

Major reported factors associated with the apparent lack of effectiveness of telemonitoring in COPD and asthma include the dearth of useful early predictors of deteriorations using a telemonitoring approach, poor patient compliance with telemonitoring tasks, and the poor performance of conventional algorithms (typically derived from studies of paper-based symptom diaries), for detecting exacerbations. Among other factors described in the telehealth literature are poor integration into routine healthcare and into the day-to-day lives of patient. Developing accurate predictive algorithms with demonstrable clinical reliability is a priority for the future consolidation of telemonitoring of chronic respiratory diseases.

Home telemonitoring enables the collection of large data sets at the individual level. The application of predictive analytics to derive effective preventive algorithms using these data sets and data available from existing electronic medical records (including data on antibiotic/steroid use and hospital admission) may improve outcomes, enhance disease management and patients’ experiences and increase effectiveness and cost-efficiency of telemonitoring. In this context, recent advances in data mining and machine learning methods offer the opportunity to combine prior knowledge of the clinical context with telemonitoring data sets to reveal predictive patterns.

Given the current lack of an existing systematic synthesis of data in this area, this manuscript aims to provide a thorough review of the algorithms used for detecting exacerbations and supporting clinical decision in patients with respiratory conditions, (more specifically in COPD and asthma) that have been reported in home telemonitoring studies. The study looks at significant service-related factors associated with the performance of home monitoring algorithms such as the technology deployed, data collection rate, system interoperability, the agreed exacerbation definition, trial factors such as sample size and study duration, and the machine learning techniques used in the reviewed studies.

**Theoretical background**

This section includes a brief overview of the fields of predictive analytics and machine learning and their application to the healthcare context.

**Predictive analytics in healthcare**

Healthcare system use, clinical trials and real-time telemonitoring systems are providing data sets larger than are manageable by conventional management tools. Data science is an emerging discipline that combines and draws connections between the fields of statistics and information and computer science, which can support the management of these complex and large data sets. In the prediction of health outcomes, conventional statistical analysis is usually applied by calculating scores for risk stratification. Generally, the underlying belief is that a small number of important variables exist on which the model can rely. However, these variables typically interact with each other in a concealed way which means that often they are not retained in the predictive model.

Predictive analytics, a subset of data science, is based on inductive inference rather than classical statistics and is well suited to the analysis of high-dimensional data sets and to the automated knowledge discovery process.
Predictive analytics refers to the systematic use of statistical or machine learning methods to make predictions and support decision-making. It uses computational techniques rooted in several domains including statistics, database management, artificial intelligence, machine learning, pattern recognition and data visualization. Currently, predictive analytics is being applied in many fields such as public safety and security, cybersecurity and social media. In healthcare, health analytics offers methods that can contribute to the potential improvement of patient care by supporting and enhancing medical decision-making.

Within predictive analytics, modelling takes data with a known variable of interest or target variable and a series of predictor values or features. A model that relates the target variable and the features is then developed. Therefore, the two basic components involved in predictive analytics are (a) the data that underlie the model and (b) the set of mathematical techniques that are applied to the data to draw inference. In telehealthcare, clinical prediction models usually manage data that include patient, disease and environmental characteristics as well as treatment information to predict a diagnostic or prognostic outcome.

**Machine learning.** Machine learning is the crucial methodology in predictive analytics. Conventional statistical analysis focuses on explaining data and relies on an expert (i.e. human) to formulate and discover cause–effect relationships, driven by a set of predefined assumptions. Machine learning is more data-focused and orientated to generating hypotheses and building predictive models using algorithms. Through machine learning, a computer is programmed to reveal hidden relationships that explain or predict a particular outcome. The resulting predictive model may be improved over time with increased number of cases.

Machine learning works by (a) defining goals, (b) exploring data and (c) training, refining and validating models. Supervised, unsupervised and complementary algorithms can be used in the process. Supervised learning includes two categories of algorithms. If the output consists of one or more continuous variables, regression algorithms are used. Classification algorithms are applied for categorical response values. In the case of telemonitoring and early detection of patients’ deterioration, supervised learning (classification algorithms) can be used to model a health-related outcome (e.g. alarm state because of exacerbation of respiratory symptoms) using a collection of a priori known inputs and outputs (training data) and exploring interactions to produce an inferred function (classifier) that can predict a response (output) given a valid set of predictor values. Common classification algorithms for supervised learning in the healthcare field include artificial neural networks, decision trees, random forests, Bayesian networks, k-nearest neighbors, support vector machines, linear discriminant analysis, k-means clustering and logistic regression.

There are three stages involved in machine learning: training, validation and testing of the algorithm. In practice, machine learning algorithms are trained by estimation of unknown internal parameters using a trial/study data set in which category labels are manually assigned to predictors (e.g. defined exacerbations). The model then needs to be validated and tested to quantify its performance. Experimental validation using an external data set is the best method of validating a model and ensuring generalizability. However, in telehealthcare interventions, to acquire further samples is costly and the amount of manually labelled data available is small. In this scenario, different strategies can be followed for internally evaluating the model. When large data sets are available, data may be randomly divided into three parts: training, testing and validation sets. The training set is used to build the model via a learning algorithm and to identify discriminating features of the predictor variables. Occasionally, different models can be combined to achieve results that outperform any of the individual models. This is known as an ensemble of classifiers. The validation set is used to assess how well the model perform against real data, to ensure stability and, in some cases, to fine-tune the model. The test set is used to assess the prediction error of the final model. The splitting of available data into two disjoint sets (training and test set) is also a widespread practice. In data-poor situations, when disjoint sets cannot be created, resampling methods such as cross-validation and bootstrapping may be used for internal validity.

**Methodology**

**Search strategy, eligibility criteria and study selection**

We undertook a systematic review, reported according to Preferred Reporting Items for Systematic
Reviews and Meta-Analyses standards with searches conducted in Medline, Scopus, Web of Science and Cochrane library in July 2015. Search terms were based on a combination of the following keywords: [telemonitoring OR telecare OR telehealth OR telehomecare OR home monitoring] AND [COPD OR chronic obstructive pulmonary disease OR asthma OR respiratory diseases]. Additional searches were performed in the references cited in the finally selected articles to find relevant additional studies.

The inclusion criteria were as follows: (1) the study involved home telemonitoring of COPD or asthma and described the algorithms used to detect episodes, (2) the study presented results on the performance of algorithms for the automatic prediction and/or detection of respiratory exacerbations, and (3) the article was written in English and published in peer-reviewed journals from 2005. This date was chosen according to the findings of a recent systematic review that reported the majority of the articles that described home telemonitoring interventions were published from 2010 to the present, suggesting that this is a relatively new field in respiratory research.

Some small pilot and feasibility studies were included because the expected outcomes were directly related to the design and implementation of algorithms for the early detection of respiratory exacerbations. Conference, posters abstracts and general reviews were not considered. Publications that focused on settings or conditions other than home telemonitoring in COPD and asthma were excluded.

Data extraction and quality assessment
Two authors (DS and MAF) extracted the data from the selected studies and a third author (AL) checked for inconsistencies. The following data were extracted and are summarized in three tables. Table 1 reports the first author’s last name, year of publication, country where the study was conducted, type of patients (e.g. disease) to which telemonitoring was applied, duration, study design, number of subjects that completed the telemonitoring study, admissions and/or exacerbations in the telemonitoring group and the exacerbation criteria applied. Table 2 summarizes the specification of clinical and physiological parameters collected in the telemonitoring (e.g. symptoms and physiological variables), the features and algorithms used for triggering automatic alerts and the personalization strategies. Table 3 describes the technology deployed during the study, provides details about data transmission, patients’ compliance with remote monitoring tasks and study outcomes.

The quality of included studies was independently evaluated by two reviewers using a modified version of the tool developed by Hailey et al. Articles were rated from A (high quality and high degree of confidence in the study findings) to E (poor quality and unacceptable uncertainty in the study findings). Disagreements were resolved by consensus.

Because of the diverse outcome measurements and the heterogeneous nature of data collected and clinical diversity of the studies, a meta-analysis was not performed; instead we used a narrative synthesis.

Results
Study selection
Totally, 407 articles were identified with the search strategy. After removing duplicates and review articles, 189 articles were retained. Two researchers screened title, abstract and keywords (DSM and MAF) to determine eligibility for this review. These potentially relevant articles were examined and 32 articles with potential significance for this study were selected and full text was retrieved. The bibliography of each manuscript was screened and four additional relevant studies were identified. Full-text articles were assessed for eligibility, and finally, 20 were included. The majority were from European countries, with two from United States and one each from Canada and Australia.

Study characteristics and quality assessment
Study characteristics are presented in Table 1. Sixteen studies focused on COPD and four studies involved patients with asthma. Nine of the studies were controlled trials, six were prospective cohort studies, four were pilot/feasibility studies and one was an uncontrolled before and after studies. The number of telemonitored patients that completed the studies ranged from 5 to 169. Two studies used the same data set.

With respect to the quality assessment, 11 studies were rated A (high quality), 7 were rated B (good quality) and 2 were rated D (poor to fair quality).

Home telemonitoring duration, exacerbation criteria and admissions
Duration of the telemonitoring period ranged from 3 to 15 months (Table 1). Ten studies had a duration...
Table 1. Description of the characteristics of the selected studies for telemonitoring COPD and asthma.

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Country</th>
<th>Disease</th>
<th>Length of follow-up</th>
<th>Study design</th>
<th>Telemonitored patients that completed the study</th>
<th>Admissions and exacerbations in telemonitoring group</th>
<th>Exacerbation criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deschildre</td>
<td>2012</td>
<td>France</td>
<td>Asthma</td>
<td>12 months</td>
<td>RCT</td>
<td>15</td>
<td>2 (median) exacerbations per patient/year</td>
<td>Defined by a systemic corticosteroid course during the follow-up</td>
</tr>
<tr>
<td>Finkelstein</td>
<td>2013</td>
<td>United States</td>
<td>Asthma</td>
<td>7000 records</td>
<td>Data from a previous prospective observational trial</td>
<td>26</td>
<td>148 exacerbations. Admissions not detailed.</td>
<td>Not specified</td>
</tr>
<tr>
<td>Kupczyk</td>
<td>2013</td>
<td>Sweden</td>
<td>Asthma</td>
<td>12 months</td>
<td>Cohort study</td>
<td>169</td>
<td>122 exacerbations, 14 admissions</td>
<td>ATS/ERS criteria</td>
</tr>
<tr>
<td>van Vliet</td>
<td>2015</td>
<td>The Netherlands</td>
<td>Asthma</td>
<td>12 months</td>
<td>Prospective, observational</td>
<td>94</td>
<td>77 exacerbations, ATS/ERS criteria.</td>
<td>Increase in symptoms or the patient initiated the use of antibiotics and/or prednisolone.</td>
</tr>
<tr>
<td>Trappenburg</td>
<td>2008</td>
<td>The Netherlands</td>
<td>COPD</td>
<td>6 months</td>
<td>Multicentre, RCT</td>
<td>59</td>
<td>59 exacerbations, 38 admissions 1 admission.</td>
<td>Anthonisen criteria</td>
</tr>
<tr>
<td>Koff</td>
<td>2009</td>
<td>United States</td>
<td>COPD</td>
<td>3 months</td>
<td>RCT</td>
<td>19</td>
<td>9 exacerbations.</td>
<td>Anthonisen criteria</td>
</tr>
<tr>
<td>Sund</td>
<td>2009</td>
<td>United Kingdom</td>
<td>COPD</td>
<td>6 months</td>
<td>Uncontrolled before-after study</td>
<td>18</td>
<td>75 exacerbations, 6 admissions</td>
<td>Increase in symptoms or the patient initiated the use of antibiotics and/or prednisolone.</td>
</tr>
<tr>
<td>Halpin</td>
<td>2011</td>
<td>United Kingdom</td>
<td>COPD</td>
<td>4 months</td>
<td>RCT</td>
<td>77</td>
<td>86 exacerbations, 3 hospitalization</td>
<td>Anthonisen criteria</td>
</tr>
<tr>
<td>Jensen</td>
<td>2012</td>
<td>Denmark</td>
<td>COPD</td>
<td>4 months</td>
<td>RCT (a study conducted for a different purpose)</td>
<td>57</td>
<td>Admissions not detailed. 10 exacerbations accounted.</td>
<td>Hospital admission, administration of antibiotics or corticosteroids in combination with worsening of specific symptoms.</td>
</tr>
<tr>
<td>Dinesen, Haesum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yañez</td>
<td>2012</td>
<td>Spain</td>
<td>COPD</td>
<td>3 months</td>
<td>Multicentre, prospective cohort study.</td>
<td>89</td>
<td>30 admissions</td>
<td>Not specified</td>
</tr>
<tr>
<td>Heijden</td>
<td>2013</td>
<td>The Netherlands</td>
<td>COPD</td>
<td>9 days</td>
<td>Pilot study</td>
<td>5</td>
<td>Not detailed</td>
<td>Explored the performance of the system with three different approaches: symptoms, drugs and contact.</td>
</tr>
<tr>
<td>Johnston</td>
<td>2013</td>
<td>Canada</td>
<td>COPD</td>
<td>15 months</td>
<td>Prospective, observational</td>
<td>49 (46 with an extended period)</td>
<td>191 exacerbations 148 admissions</td>
<td>Anthonisen criteria</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Country</th>
<th>Disease</th>
<th>Length of follow-up</th>
<th>Study design</th>
<th>Telemonitored patients that completed the study</th>
<th>Admissions and exacerbations in telemonitoring group</th>
<th>Exacerbation criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedone$^{58}$</td>
<td>2013</td>
<td>Italy</td>
<td>COPD</td>
<td>9 months</td>
<td>RCT</td>
<td>50</td>
<td>0.13 per patient/year</td>
<td>Increase in symptoms that necessitates a change in regular medication.</td>
</tr>
<tr>
<td>Pinnock$^{13}$</td>
<td>2013</td>
<td>United Kingdom</td>
<td>COPD</td>
<td>12 months</td>
<td>Multicentre, RCT</td>
<td>127</td>
<td>1.2 per patient/year with a diagnosis of AECOPD.</td>
<td>Increase in symptoms that necessitates a change in regular medication.</td>
</tr>
<tr>
<td>Granero$^{59}$</td>
<td>2014</td>
<td>Spain</td>
<td>COPD</td>
<td>6 months</td>
<td>Pilot study</td>
<td>15</td>
<td>41 exacerbations. Admissions not detailed.</td>
<td>Anthonisen criteria</td>
</tr>
<tr>
<td>Segrelles et al.$^{64}$</td>
<td>2014</td>
<td>Spain</td>
<td>COPD</td>
<td>7 months</td>
<td>Controlled, non-blind clinical trial</td>
<td>26</td>
<td>12 admissions</td>
<td>Adapted from the GOLD definition.</td>
</tr>
<tr>
<td>Burton$^{60}$</td>
<td>2015</td>
<td>United Kingdom</td>
<td>COPD</td>
<td>200 days</td>
<td>Pilot study</td>
<td>19</td>
<td>172 exacerbations. Admissions not detailed.</td>
<td>Anthonisen criteria or antibiotics intake.</td>
</tr>
<tr>
<td>Hardinge$^{61}$</td>
<td>2015</td>
<td>United Kingdom</td>
<td>COPD</td>
<td>6 months</td>
<td>Cohort study</td>
<td>18</td>
<td>7 admissions</td>
<td>The patient initiated the use of either antibiotics, oral steroids or both antibiotics and steroids.</td>
</tr>
<tr>
<td>Mohktar$^{62}$ et al.$^{81}$</td>
<td>2015</td>
<td>Australia</td>
<td>COPD</td>
<td>12 months</td>
<td>RCT. Data from a previously reported home telehealth study.$^{81}$</td>
<td>21</td>
<td>1.3 per patient/year</td>
<td>Adapted from the GOLD definition.</td>
</tr>
<tr>
<td>Morillo$^{63}$</td>
<td>2015</td>
<td>Spain</td>
<td>COPD</td>
<td>6 months</td>
<td>Pilot study</td>
<td>15</td>
<td>33 exacerbations. Admissions not detailed.</td>
<td>The patient initiated the use of either antibiotics, oral steroids or both antibiotics and steroids as well as visits to emergency unit or admissions.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Symptoms</th>
<th>SpO₂</th>
<th>LF</th>
<th>BP</th>
<th>HR</th>
<th>W</th>
<th>T</th>
<th>RR</th>
<th>LS</th>
<th>PA</th>
<th>Other</th>
<th>Features extraction methods</th>
<th>Personalization, training and validation</th>
<th>Automatic algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deschildre⁶⁶</td>
<td>2012</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>Intervention if FEV₁ ≥ 80%</td>
<td>Holdout validation. Personalization discussed as a future improvement.</td>
<td>Naive Bayesian Classifier and Support Vector Machines.</td>
<td></td>
</tr>
<tr>
<td>Finkelstein⁶⁷</td>
<td>2013</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>MDL algorithm used to assess the rank and select attributes.</td>
<td>Holdout validation. Personalization discussed as a future improvement.</td>
<td>k-NN algorithm using only FeNO and EBC.</td>
<td></td>
</tr>
<tr>
<td>Kupczyk⁶⁸</td>
<td>2013</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>Decrease in PEF on two consecutive days or a 20% increase in day symptoms on two consecutive days</td>
<td>Baseline defined as personal best during optimization phase of the study.</td>
<td>Decrease in PEF on two consecutive days or a 20% increase in day symptoms on two consecutive days</td>
<td></td>
</tr>
<tr>
<td>van Vliet⁶⁹</td>
<td>2015</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>FeNO, EBC measured during clinical visits every 2 months.</td>
<td>Holdout validation.</td>
<td>k-NN algorithm using only FeNO and EBC.</td>
<td></td>
</tr>
<tr>
<td>Trappenbur⁵¹</td>
<td>2008</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>Recognition of changes in defined patterns of major and minor respiratory symptoms.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koff⁶⁵</td>
<td>2009</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>Three flags based on symptoms and monitored parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sund⁵²</td>
<td>2009</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>Baseline defined as exacerbation-free period of 14 days.</td>
<td>Automatic threshold based. Time score plot for symptoms and FEV₁.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halpin⁵³</td>
<td>2011</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>Personalized baseline recalculated every 4 weeks unless an event occurred.</td>
<td>Automatic threshold based. Increase in total score ≥ 12 above baseline for 2 days.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Symptoms</th>
<th>SpO₂</th>
<th>LF</th>
<th>BP</th>
<th>HR</th>
<th>W</th>
<th>T</th>
<th>RR</th>
<th>LS</th>
<th>PA</th>
<th>Other</th>
<th>Features extraction methods</th>
<th>Personalization, training and validation</th>
<th>Automatic algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jensen⁴⁴</td>
<td>2012</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>273 features extracted. Mean, SD of features 5 to 30–days prior to exacerbations. SEPCOR algorithm for dimensionality reduction.</td>
<td>Leave-one-episode-out cross validation.</td>
<td>Treated as a pattern recognition problem. Linear discriminant functions to predict the risk of exacerbation in the next 30 days.</td>
</tr>
<tr>
<td>Dinesen⁷⁹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haesum⁸⁰</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yañez⁵⁵</td>
<td>2012</td>
<td>■</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Baseline of 5 days prior to exacerbation.</td>
<td>Predictions personalized by entering patient-specific data. Two data sets used for model validation with a total of 99 patients.</td>
<td>Automatic threshold based. Increase in breathing rate.</td>
</tr>
<tr>
<td>Heijden⁵⁶</td>
<td>2013</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Baseline for each symptom defined as the scoring on a given day.</td>
<td>Predictions personalized by entering patient-specific data. Two data sets used for model validation with a total of 99 patients.</td>
<td>Expert knowledge-based model. Bayesian network model. Status. Probabilistic model incorporated in the smartphone.</td>
</tr>
<tr>
<td>Johnston⁵⁷</td>
<td>2013</td>
<td>■</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alerts customized for patients by physicians on the basis of the patient’s clinical situation.</td>
<td>Alerts were triggered for symptom changes (one or more symptoms moved up two levels of severity on a given day or any symptom reached its worst level on a given day), missed diary transmissions or medical care.</td>
<td>Threshold based. Alerts were triggered for symptom changes (one or more symptoms moved up two levels of severity on a given day or any symptom reached its worst level on a given day), missed diary transmissions or medical care.</td>
</tr>
<tr>
<td>Pedone⁵⁸</td>
<td>2013</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Alerts customized for patients by physicians on the basis of the patient’s clinical situation.</td>
<td>Alerts were triggered for symptom changes (one or more symptoms moved up two levels of severity on a given day or any symptom reached its worst level on a given day), missed diary transmissions or medical care.</td>
<td>Automatic threshold based. Predefined range. No further details provided.</td>
</tr>
<tr>
<td>Pinnock¹³</td>
<td>2013</td>
<td>■</td>
<td>■</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Oxygen saturation alert was able to be defined on an individual patient basis.</td>
<td>Oxygen saturation alert was able to be defined on an individual patient basis.</td>
<td>Threshold based. Symptom scores.</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Symptoms</th>
<th>SpO₂</th>
<th>LF</th>
<th>BP</th>
<th>HR</th>
<th>W</th>
<th>T</th>
<th>RR</th>
<th>LS</th>
<th>PA</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granero⁵⁹</td>
<td>2014</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segrelles Calvo⁶⁴</td>
<td>2014</td>
<td>☐ ☐ ☐ ☐ ☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burton⁶⁰</td>
<td>2015</td>
<td>☐ ☐ ☐ ☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harding⁶¹</td>
<td>2015</td>
<td>☐ ☐ ☐ ☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mohktar⁶²</td>
<td>2015</td>
<td>☐ ☐ ☐ ☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antoniades⁸¹</td>
<td>2015</td>
<td>☐ ☐ ☐ ☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Features extraction methods**
- Average scores for total, major, minor and complementary symptoms.
- Ten-fold Cross validation.
- Probabilistic neural network.
- Thresholds defined by personal patient’s limits.
- Thresholds were defined with 'personal limits'.
- Multilevel logistic regression was used to examine the association between FEV₁, pulse and SpO₂ and total symptom score.
- PDF and CDF estimated for each data type.
- Automatic threshold. Patients-specific alert thresholds value derived from 95th centile of CDF.
- Leave one subject out cross validation.
- CART algorithm attempted to predict the patient’s condition one day in advance. Comparison of their current physiological measurements against the distribution of their measurements over the previous month.
- Ten-fold Cross validation.
- Treated as a pattern recognition problem. K-means clustering with symptom scores. Lung sounds not reported.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Equipment</th>
<th>Data collection rate</th>
<th>Compliance</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deschildre⁶⁶</td>
<td>2012</td>
<td>Symptoms recorded in a paper diary. Portable spirometer.</td>
<td>Daily</td>
<td>Between 30% and 96%</td>
<td>The hypothesis that daily home telemonitoring of FEV₁ would be superior to conventional management in terms of severe exacerbations and healthcare use was not supported.</td>
</tr>
<tr>
<td>Finkelstein⁶⁷</td>
<td>2013</td>
<td>Laptop computer with a peak exploratory flow meter.</td>
<td>Daily</td>
<td></td>
<td>Using a 7-day window, a support vector machine was able to predict asthma exacerbation in day 8 with Se of 84% and Sp of 77%. Naïve Bayesian classifier showed a Se of 80% and Sp of 77%.</td>
</tr>
<tr>
<td>Kupczyk⁶⁸</td>
<td>2013</td>
<td>Home monitor, portable spirometer and peak-flow meter</td>
<td>Daily</td>
<td>86.6% and 79.4% in patients with severe and mild asthma.</td>
<td>The algorithm was able to detect severe exacerbations with Se of 65% and Sp of 95%.</td>
</tr>
<tr>
<td>van Vliet⁶⁹</td>
<td>2015</td>
<td>Home monitor, a portable spirometer.</td>
<td>Daily, but transmitted twice a week.</td>
<td>87.3%</td>
<td>The predictive power of FeNO and inflammatory markers in EBC for prediction of an asthma exacerbation was low, even when combined with clinical characteristics and symptoms. k-NN provided 52% of accuracy. The AUC for the model was 59%.</td>
</tr>
<tr>
<td>Trappenburg⁵¹</td>
<td>2008</td>
<td>Home monitor (Health Buddy).</td>
<td>Daily</td>
<td></td>
<td>The telemonitored group showed a significant decrease in hospital admission rates and exacerbations. 9 red flags activated. One resulted in a hospitalization.</td>
</tr>
<tr>
<td>Koff⁶⁵</td>
<td>2009</td>
<td>Home monitor (Health Buddy), pulsoximeter, FEV₁ monitor and pedometer.</td>
<td>Daily</td>
<td></td>
<td>55 (73%) of exacerbations detected. 37 by symptom score, 6 by &gt;10% decline in FEV₁ only and 12 by symptoms and decline in FEV₁ &gt; 10%. EXACT-PRO did not appear able to detect exacerbations neither severity nor duration.</td>
</tr>
<tr>
<td>Sund⁵²</td>
<td>2009</td>
<td>PDA and a portable spirometer.</td>
<td>Daily</td>
<td>77%</td>
<td></td>
</tr>
<tr>
<td>Halpin⁵³</td>
<td>2011</td>
<td>Blackberry smartphone</td>
<td>Daily</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Equipment</td>
<td>Data collection rate</td>
<td>Compliance</td>
<td>Performance</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Jensen</td>
<td>2012</td>
<td>Home monitor, digital body weight scale, blood pressure meter, pulse oximeter, spirometer and step counter.</td>
<td>Variable</td>
<td></td>
<td>Prediction of risk of exacerbation in the next 30 days. Se of 70%, Sp of 95%. PPV of 75% and NPV of 94.7%. Not useful to help with self-administration of medication. Seven of ten exacerbations detected. AUC = 0.73.</td>
</tr>
<tr>
<td>Dinesen</td>
<td>2012</td>
<td>Respiratory monitor installed in the domiciliary oxygen supply.</td>
<td>Recorded daily. Stored in local memory.</td>
<td></td>
<td>24 h before admission, a mean increase of 4.4/min (30% from baseline) provided the best combination of Se of 66% and Sp of 93%. (AUC = 0.79). Two days before hospitalization, a mean increase of 2.3/min (15% change from baseline) was associated with a Se of 72% and a Sp of 77% (AUC = 0.76).</td>
</tr>
<tr>
<td>Haesum</td>
<td>2013</td>
<td>Smartphone, spirometer and pulse oximeter.</td>
<td>Variable</td>
<td></td>
<td>System evaluated on offline databases and for different definitions of exacerbations. In dataset B, true positive rate of 0.88, false positive rate of 0.20, AUC = 0.87.</td>
</tr>
<tr>
<td>Yañez</td>
<td>2012</td>
<td>Respiratory monitor installed in the domiciliary oxygen supply.</td>
<td>Recorded daily. Stored in local memory.</td>
<td></td>
<td>Respiratory symptom scores before exacerbations and after the participant reported a return to normal breathing showed no differences.</td>
</tr>
<tr>
<td>Heijden</td>
<td>2013</td>
<td>Blackberry smartphone.</td>
<td>Daily</td>
<td>100%</td>
<td>Only oxygen saturation could identify timely exacerbations. Reference to possible phenotypic variability in COPD. Telemonitoring could cut by 33% the risk for hospitalizations. However, the average length of stay was longer in the study than in the control group.</td>
</tr>
<tr>
<td>Johnston</td>
<td>2013</td>
<td>Mobile phone, pulse oximeter and a wrist band with sensors for heart rate, physical activity, near body temperature and galvanic skin response.</td>
<td>Wristband, transparent for user, 5 times/hour. Pulse-oximeter, daily.</td>
<td>The proportion of days with at least three measurements ranged from 0% to 97% (mean: 60%).</td>
<td>No benefits found in automatic generated alerts in term of clinical outcomes. The authors claim for the lack of early predictors of exacerbations and accurate algorithms.</td>
</tr>
<tr>
<td>Pedone</td>
<td>2013</td>
<td>Touch screen, pulse oximeter.</td>
<td>Daily</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
Table 3. (continued)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Equipment</th>
<th>Data collection rate</th>
<th>Compliance</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granero$^{59}$</td>
<td>2014</td>
<td>Ad-hoc tablet and electronic stethoscope.</td>
<td>Daily</td>
<td></td>
<td>Early prediction of COPD exacerbations with a margin, as average, of 4.8 days.</td>
</tr>
<tr>
<td>Segrelles Calvo$^{64}$</td>
<td>2014</td>
<td>Portable spirometer, pulse oximeter, heart rate monitor and blood pressure monitor</td>
<td>Daily</td>
<td></td>
<td>50 clinical alerts identified. 12 exacerbations not detected. False positives automatically triggered were not reported.</td>
</tr>
<tr>
<td>Burton$^{60}$</td>
<td>2015</td>
<td>Touch screen, pulse oximeter and a portable spirometer.</td>
<td>Daily</td>
<td></td>
<td>Three patterns of exacerbations detected. Compared to baseline, at the onset of exacerbations, the mean pulse rate increased from 87 to 94 /min and the mean SpO$_2$ fell from 93.6 to 92.4%. Physiological variables did not differentiate between exacerbations and isolated bad days.</td>
</tr>
<tr>
<td>Hardinge$^{61}$</td>
<td>2015</td>
<td>Tablet computer, pulse oximeter.</td>
<td>Daily</td>
<td></td>
<td>Within 3 months, 95% of sessions completed.</td>
</tr>
<tr>
<td>Mohktar$^{62}$ and Antoniades$^{81}$</td>
<td>2015, 2012</td>
<td>Laptop computer, blood pressure monitor, pulse oximeter, stethoscope, thermometer, scale, pneumotachograph, electrocardiogram touch plate and thermometer.</td>
<td>Daily</td>
<td></td>
<td>Pending on reporting results.</td>
</tr>
<tr>
<td>Morillo$^{63}$</td>
<td>2015</td>
<td>Ad-hoc tablet and electronic stethoscope.</td>
<td>Daily</td>
<td></td>
<td>Early prediction of COPD exacerbations with a margin, as average, of 4.5 days. Se and Sp were 74.6% and 89.7%, respectively, and AUC was 0.84. 31 of 33 AECOPD. Mention to possible phenotypic variability in COPD.</td>
</tr>
</tbody>
</table>

FEV$_1$: Forced expiratory volume in 1 second. Se: sensitivity, Sp: specificity, AUC: area under the received operating characteristic curve, PPV: positive predicted value, FeNO: fractional exhaled nitric oxide, EBC: exhaled breath condensate, k-NN: k-nearest neighbour, CART: classification and regression tree, NPV: negative predictive value, AECOPD: exacerbation of COPD.
greater than 6 months, five studies had a duration of 6 months and five studies had a duration shorter than 6 months.

An ‘event-based’ definition of an exacerbation,\(^{70}\) including self-administration of medication or unscheduled visits to emergency units and/or admissions, was used in nine studies (\(n = 6\) in COPD and \(n = 3\) in asthma studies). Symptom-based definitions of exacerbation (e.g. Anthonisen criteria\(^ {71}\)) were used in seven COPD studies. Two studies did not detail the criteria used and the remaining two studies explored different definitions of exacerbation.

The following symptom questionnaires were used in the studies: Clinical COPD Questionnaire (CCQ),\(^ {72}\) exacerbations of chronic pulmonary disease Tool (EXACT)- patient-reported outcome (PRO),\(^ {73}\) St Georges Respiratory Questionnaire (SGRQ),\(^ {74}\) Paediatric Asthma Quality of Life Questionnaire (PAQLQ),\(^ {75}\) Asthma Control Questionnaire (ACQ)\(^ {76}\) along with a variety of ad-hoc questionnaires for COPD and asthma.

The number of admissions in each study caused by worsening in respiratory symptoms ranged from 1 to 148. The maximum number of recorded exacerbations during a study was 191 for a telemonitored group of 49 subjects. The highest average number of exacerbations was 9 per patient.

Factors that influence algorithms’ performance

Clinical data, sensors and devices. Tables 2 and 3 present information about the telemonitoring system deployed during the study period. Clinical telemonitored data consisted of symptoms collected using electronic questionnaires (\(n = 3\)), physiological measurements (\(n = 4\)) or a combination of symptoms and physiological measurements (\(n = 13\)). The data acquisition process was similar in all studies. Typically, symptoms were reported using electronic diaries completed via the telemonitoring device. The physiological parameters were collected via various peripherals and transmitted by telemonitoring system. All of asthma studies recorded lung function (peak exploratory flow (PEF) in two cases and forced expiratory volume in one second (FEV\(_1\)) in three studies), as well as symptoms. A wider range of telemetered physiological variables were recorded in COPD: lung function using FEV\(_1\) (\(n = 6\)), peakflow (\(n = 1\)) and forced vital capacity (FVC) (\(n = 1\)), oximetry (\(n = 9\)), heart rate (\(n = 7\)), weight (\(n = 2\)), respiratory rate (\(n = 2\)), physical activity (\(n = 3\)), blood pressure (\(n = 3\)), body temperature (\(n = 2\)) and lung sounds (\(n = 2\)).

Data were recorded at least on a daily basis in all studies. More specifically, data were gathered more than once per day (\(n = 1\)), with adjustable frequency (\(n = 2\)), or recorded daily although transmitted twice a week (\(n = 1\)). Patient collaboration was always required during data acquisition excepting in one study in COPD where a wristband recorded heart rate, physical activity, temperature and galvanic skin response and transmitted the physiological information several times per hour without requiring patient intervention.

Feature extraction, personalization and model validation.

There was limited description of how the collected data were processed. Mean (\(n = 4\)), standard deviation, the percentage of change of the parameter from the distribution mean and the standard score or z-score of parameters were among the used features (\(n = 2\)). Two studies applied statistical methods to select predictive features. The heuristic ‘SEPCOR’ algorithm that is based on the variability measure of each feature,\(^ {77}\) and the minimum description length algorithm were each used in one study.

In the studies that reported predictive analytic models, personalization of the algorithm was generally discussed as a future improvement (\(n = 3\)). Personalized predictions by entering patient-specific data were implemented in two studies. Model validation was performed through different approaches: holdout validation (\(n = 2\)), 10-fold cross validation (\(n = 2\)), leave-one-episode-out cross validation (\(n = 1\)) and leave-one-subject-out cross validation (\(n = 1\)).

Algorithms for automatic detection of health deterioration

Different strategies were followed to detect early exacerbations of respiratory symptoms. In the simplest approach, the decision support consisted of a basic decision rule based on assigning threshold values to the collected parameters (\(n = 12\)). This resulted in an alert when scores crossed the predefined threshold indicating deterioration in the patient’s condition. In the remaining studies, the challenge of early detection of exacerbations was addressed using machine learning techniques (\(n = 8\)). More specifically, naïve Bayesian classifiers (\(n = 1\)), support vector machines (\(n = 1\)), probabilistic neural networks (\(n = 1\)), a
Bayesian network model \((n = 1)\) and \(k\)-NN \((n = 1)\) were among the techniques utilized. In addition, classification and regression trees \((n = 1)\), multilevel logistic regression \((n = 1)\), linear discriminant analysis \((n = 1)\) and clustering techniques like \(k\)-means clustering \((n = 1)\) were alternatively implemented.

In some studies, predictions were personalized by entering patient-specific data or by modelling individual characteristics. In the case of a decision rule based on score thresholds, an individual baseline was established according to different criteria \((n = 9)\). In asthma, a non-personalized approach that used a fixed common threshold of 80% of FEV\(_1\) for all patients was also reported \((n = 1)\). Recognition of changes in previously defined patterns of major and minor respiratory symptoms was also implemented \((n = 2)\).

When used, the baseline was defined in various ways: as an exacerbation-free period of 14 days \((n = 1)\), at the beginning of the study \((n = 1)\), recalculated every 4 weeks unless an exacerbation occurred \((n = 1)\), as the average of the last 10 days of the treatment optimization period \((n = 1)\), using 5 days prior to exacerbation \((n = 1)\), defined as the scoring of a given day \((n = 1)\) or calculated as 95th centile of the probability density function \((n = 1)\).

Results in terms of system performance were not always reported. In COPD, when reported, accuracy in detecting exacerbations ranged from 40% to 94%, sensitivity from 70% to 80% and specificity from 61% to 95%. In asthma, only one study reported accuracy (52%), whilst achieved sensitivity ranged from 65% to 84% and specificity from 77% to 95%.

**Discussion**

This systematic literature review provides a description of the algorithms used in home telemonitoring interventions in COPD and asthma for supporting clinical decision and summarizes the findings related to factors associated with the performance of home monitoring algorithms. Reviewed studies were heterogeneous in terms of their definition of exacerbation, telemonitoring duration, telemonitored measurements, type of technology deployed and algorithms used. In asthma, the home telemonitoring comprised symptoms and lung function (PEF and FEV\(_1\)): in COPD a much wider range of physiological measures were used.

**Home telemonitoring duration, exacerbation criteria and admissions**

A common limitation in the selected studies was the examination of a relatively small group of patients who presented a high rate of exacerbations. Winter periods were extensively selected for trials because exacerbations are more likely than in other seasons.\(^{51}\) Although this may have enabled recruitment of ‘at-risk’ populations it may have affected generalizability of results. Randomized controlled trials with a larger sample of patients over periods longer than 6 months might improve the robustness of algorithms and the generalizability of their findings.\(^{62,82}\)

The lack of consensus on a definition of an exacerbation\(^*\) and unreported episodes that are poorly labelled in data sets \(^4\) may adversely affect algorithms’ performance. Identifying an exacerbation episode is dependent on establishing a link between the characteristics being measured and exacerbations. Asthma attacks are relatively clearly defined; in COPD the absence of a gold standard definition makes identifying exacerbations difficult. Of the questionnaires CCQ, PAQLQ, ACQ and SQRG widely used in the reviewed studies, only EXACT-PRO was specifically designed and validated for detecting COPD exacerbations. Finally, in most of studies patients were given self-management advise encouraging them to intervene, usually at quite an early stage, if parameters strayed outside of expected. This may have influenced some outcomes (e.g. hospital admission and length of exacerbation) possibly blunting the relationship between the original predictive features and outcomes.

**Factors that influence algorithms’ performance**

**Clinical data, sensors and devices.** The combination of symptoms and physiological measurements was the most common approach in the selected works \((n = 13)\). All the asthma studies utilized daily readings, though one only transmitted data twice weekly. In COPD, physiological measurements have not proved to be able to predict deteriorations, either because they change late in the time course of exacerbation, they cannot be measured reliably or because therapeutic interventions during the experiment alter the outcomes hindering the accuracy of algorithms.\(^{17}\) For this reason all studies used a daily data collection (only one study collected data daily but stored data in an internal memory of the
respiratory monitor). Whilst the 7-day recall scores and the daily diary scores have been found to be equivalent in detecting changes over time of the impact of COPD symptoms, only daily data seem to be suitable if the outcome of interest is detecting the onset of exacerbations.\textsuperscript{85} One study used a compromise solution,\textsuperscript{56} whereby the patient with a low risk of exacerbation monitored on a weekly basis and when risk increased according to the predictive model, reporting tasks could be scheduled daily to ensure timely detection.

Since learning algorithms and deployed predictive models rely on timely reporting of accurate data, good patient compliance with reporting tasks is needed. Perceived ease of use of home telehealth devices is a significant predictor of compliance.\textsuperscript{86} Systems that offer a unified interface with contextual data on smartphones and provide interactive user feedback and self-adaptive sensor polling could be a solution to alleviate the patient burden of reporting tasks and optimize the quality of collected data.\textsuperscript{87} One study, which used a wearable sensor that automatically transmitted heart rate, temperature and activity levels five times an hour, achieved average reporting rates of 60% though the reason for 40% loss of data was due to patients not wearing the device or technical failure is not reported.\textsuperscript{58}

**Features extraction, personalization and model validation.** Although feature extraction can help to build derived values that enhance the subsequent learning and generalization steps of algorithms, these techniques were seldom implemented in the selected studies. Additionally, personalization of predictive models that used machine learning techniques was only implemented in a single study.\textsuperscript{56}

With respect to the procedure to estimate the prediction error in studies that applied machine learning methods, different approaches were followed in six studies. None applied external validity and a number of internal validity methods based on holdout and cross validation techniques were reported. Holdout approaches work well in large data sets, but can give inaccurate performance estimations because the results are highly dependent on how the training and test sets are chosen.\textsuperscript{88} Cross validation and bootstrapping can help to overcome this challenge, but the latter technique was not used in the reviewed works. Cross validation was used in different ways in four studies. It is advisable that none of the same person’s measures are in the training set and in the test set. Leave-one-subject-out cross-validation\textsuperscript{88,89} can allow for subject-to-subject variation,\textsuperscript{88} but was not carried out for any of the machine learning experiments.

**Algorithms for automatic detection of health deterioration and service implications**

The evidence that telemonitoring can enhance outcomes and reduce costs remains unclear,\textsuperscript{10} in part because of the limited performance of the predictive models used.\textsuperscript{13,14,20} Conventional threshold-based algorithms, adopted in the majority of reviewed studies (n = 12), show poor performance in early detecting respiratory exacerbations or identifying severity and duration in COPD.\textsuperscript{13,53} with the best reported accuracy being 73% of exacerbations detected.\textsuperscript{52} best sensitivity/specificity 66%/93% \textsuperscript{55} 24 h before hospitalization. COPD studies generally used symptoms scores occasionally in association with FEV\textsubscript{1}, arterial oxygen saturation (SpO\textsubscript{2}), heart and breathing rate, but there was no consistency between studies in terms of the threshold used or how the individual patient’s baseline was established/calculated. In asthma, only one study using a threshold-based algorithm reported performance as a sensitivity of 65% and a specificity of 95%.\textsuperscript{68}

If a sensitivity level over 70% is considered acceptable,\textsuperscript{54} the performance of current threshold-based algorithms may not be adequate for the real-time detection of events after experimental validation. As an alternative, the data-driven (n = 7) and knowledge-based (n = 1) approaches used in some of the reviewed studies provided encouraging results. With a purely data-driven approach, one of the studies used a probabilistic graphical model that incorporated clinical knowledge firstly defined by clinicians (based on their experience and knowledge) and later fine-tuned by experimental data. The transparency of knowledge-based models allows the system structure and inference to be understood by clinicians in contrast with black box approaches like data-driven models.\textsuperscript{90}

However, the use of semantic mapping between gathered remote telemonitoring data and existing patient-specific data and models to complement the predictive model using ontologies\textsuperscript{91} and standards\textsuperscript{92,93} was not found in the literature included in this review. Furthermore, the evaluation of strategies for connecting different predictive personalized models was not considered in the selected studies. Such
'ensemble modelling' could offer a compromise between clinical comprehension and statistical goodness of fit in order to improve the quality and robustness of the whole system.\textsuperscript{94}

Integration of telemonitoring services with the electronic healthcare systems used routinely by healthcare professionals was not described in any of the studies. Instead, an electronic health record was designed for the specific purpose of each trial. This lack of integration with the existing health record systems does not enable intelligent algorithms to interpret data from multiple sources.\textsuperscript{95} In this context, novel systems\textsuperscript{96} that automatically data-mine inpatient decision support from electronic medical records and that are predictive of real practice patterns and clinical outcomes are being evaluated with promising results.\textsuperscript{97}

Study limitations

The search was undertaken by a team with extensive experience of home telemonitoring studies and predictive analytics and using multiple databases. Despite this, the present work has limitations. Due to the heterogeneity of the data, a formal meta-analysis was not possible. The literature search was comprehensive but we may have missed home telemonitoring interventions including asthma or COPD but not specifically indexed as such, or because they were not written in English. A main limitation relates to the heterogeneous interventions, which makes comparisons hard and conclusions difficult to draw. Furthermore, the number of studies is insufficient to measure the potential of novel predictive models.

Conclusions

Our review of the algorithms used in home telemonitoring interventions in COPD and asthma suggests that the development of predictive models with clinically useful levels of accuracy, sensitivity and specificity has not yet been achieved.

A range of physiological measurements, consensus in the definition of exacerbations, achievement of high-compliance rates, the development and personalization of predictive models using machine learning techniques and the strategic importance of interoperability of telemonitoring systems with routine healthcare record systems have been discussed in this work as key factors that influence – or potentially could influence – the performance of systems for the remote monitoring of chronic respiratory patients.

New predictive models considering symptoms and physiological measures are being tested in pilot studies with positive outcomes. However, the algorithms that underpin these models have to be validated in large samples of patients, for longer periods of time and with well-established protocols. Further research is needed to identify new features associated with symptoms and physiological signals that enable the early detection of deteriorations, especially in COPD. Only then will telemonitoring achieve the aim of preventing hospital admissions, contributing to the reduction of health resource utilization and improving quality of life of patients.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

31. McCue C. Data mining and predictive analytics in public safety and security. *IT Prof* 2006; 8: 12–18.
34. Chowriappa P, Dua S and Todorov Y. Introduction to Machine Learning in Healthcare Informatics, in


