International Journal of Medical Informatics 111 (2018) 72–76

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Population exacerbation incidence contains predictive information of acute exacerbations in patients with chronic obstructive pulmonary disease in telecare

Thomas Kronborg*, Lasse Mark, Simon Lebech Cichosz, Pernille Heyckendorff Secher, Ole Hejlesen

Department of Health Science and Technology, Aalborg University, Denmark

ARTICLE INFO

Keywords:
Chronic obstructive pulmonary disease
COPD
Exacerbation
Hospitalization
Telehealth

ABSTRACT

Objective: Chronic obstructive pulmonary disease (COPD) is a major global burden largely resulting from acute exacerbations. We investigated whether the incidences of patient and population exacerbations contain predictive information for continuous prediction of exacerbations in COPD patients.

Methods: Data analysis was performed using home measurements from 1225 patients included in the large-scale telehomecare trial *TeleCare North*, where data supported 84 exacerbations occurring in 57 patients. Twenty-nine predictors were extracted and validated in two prediction models based on logistic regression. One model without and one model with inclusion of patient and population exacerbation incidences as potential predictors. The predictors were then evaluated by discriminative abilities between periods with and without exacerbation.

Results: The optimal predictor combinations provided an average area under the receiver operation characteristics curve of 0.63 with exclusion; inclusion of the population exacerbation incidence provided a curve of 0.74 ($p < 0.05$). These results were based on a two-fold patient dependent cross-validation.

Discussion: The present study has presented how the population exacerbation incidence contains predictive information in the continuous prediction of exacerbations in COPD patients. A system capable of predicting acute exacerbations could potentially prevent some cases of COPD-related complications and increase the health-related quality of life among COPD patients in telecare.

1. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is the fourth leading cause of death in the world [1] and is a major cause of chronic morbidity and mortality [2]. In 2012, COPD caused 3 million deaths, equal to 6% of all deaths globally that year [3]. Furthermore, COPD is expected to become the third leading cause of death in the world by 2030 [4]. COPD is an umbrella term that includes several chronic lung diseases and is characterized by an irreversible persistent airflow limitation that is usually progressive and associated with an increased inflammatory response in the airways [5]. Patients suffering from COPD are also affected by exacerbations, a condition characterized by a worsening of respiratory symptoms and an acceleration of the disease [6,7]. The socio-economic burden related to COPD and exacerbations of COPD are not to be underestimated, as respiratory disease-related expenses comprise 6% of the total healthcare budget in the European Union. COPD accounts for 56% of this expense, or 38.6 billion Euros, of which exacerbations account for the greatest proportion [5]. Exacerbations of COPD are responsible for approximately 110,000 deaths and more than 500,000 hospitalizations in the United States each year [7]. Furthermore, exacerbations dramatically affect a patient's quality of life [6]. Proper medication can significantly impair the rate of lung function decline, as well as the mortality related to COPD exacerbations [8]. Therefore, the prediction of approaching exacerbations could be essential for timely medical treatment and the prevention of exacerbation-related complications [5].

Previous studies have attempted to predict exacerbations several days prior to an occurrence, with the purpose of beginning proper treatment and thereby preventing hospitalizations [9–12]. These studies have been based on either a symptom score or physiological parameters. The first three studies have been limited by small or non-representative sample sizes but clearly indicate that the early detection of COPD exacerbations is possible. Additionally, one study showed great variability within the symptoms of COPD exacerbations.

* Corresponding author at: Fredrik bajers vej 7C, Room C1-219, 9220 Aalborg Ø, Denmark.
E-mail addresses: tkl@hst.aau.dk (T. Kronborg), lm@hst.aau.dk (L. Mark), simcich@hst.aau.dk (S.L. Cichosz), phl@hst.aau.dk (P.H. Secher), okh@hst.aau.dk (O. Hejlesen).

https://doi.org/10.1016/j.ijmedinf.2017.12.026
Received 12 September 2017; Received in revised form 19 December 2017; Accepted 28 December 2017

Symptoms are not consistently perceived by patients in the same manner, exhibiting not only seasonal variation but also changes in symptom perception over the course of a week or even within a single day [13]. To the best of our knowledge, no previous studies have attempted to use patient and population exacerbation incidence in the prediction of acute exacerbations in COPD patients in telecare. Therefore, the present paper will investigate the possibility of including the patient and population exacerbation incidence in combination with physiological parameters and questionnaires.

2. Methods

2.1. Data

The study data were derived from 1225 patients who were included in the *TeleCare North Trial,* a large-scale randomized controlled telehomecare trial that was conducted in the North Denmark Region between June 2012 and June 2015. The *TeleCare North Trial* is a collaboration between municipalities, general practitioners, hospitals, and Aalborg University and included patients diagnosed with COPD. Patients were equipped with a telehomecare solution, including a tablet connected via Bluetooth to a pulse oximeter (Nonin, Onyx II% SpO2, A & D Medical, Tokyo, Japan), a blood pressure monitor (Model UA-767, plus BT-C, Nonin Medical, Minnesota, USA), and a scale (Precision Health Scale, UC-321PBT-C, A & D Medical, Tokyo, Japan) [14]. Measured physiological parameters included oxygen saturation, pulse, weight, and blood pressure. The questionnaire consisted of six questions related to symptoms and worsening cough, mucus, and shortness of breath along with six questions related to inhaler use and antibiotic treatment. Patients had a two-week start-up phase that consisted of daily measurements; then, patients were subsequently instructed to perform one or two measurements on a weekly basis.

2.2. Definition of exacerbation

There has been a disagreement in the definition of exacerbations, especially in regard to the retrospective identification of exacerbations for database purposes [15–17]. Generally, there have been two different approaches in the development of a common definition of exacerbations. These approaches have been either symptom-based or event-based, where the event-based definitions are based on events such as hospitalizations or other health care contact [17]. In this study, exacerbations were defined as hospitalizations with a diagnosis related to the lungs, and this information was provided by the municipalities in the North Denmark Region and included the hospitalization date, hospitalization length, and admission diagnosis among other predictors.

2.3. Data structuring

The prediction of exacerbations was addressed as a pattern recognition problem. The principle behind the data structuring is a window moving with respect to time, inspired by Jensen et al. [9]. This method provided real-time simulation where each measurement was used for the prediction of an exacerbation. The data for each patient were organized into periods of 30 days because studies have shown that symptoms of exacerbations are affected in a period of 30 days before and 14 days after an exacerbation [13,18]. Each period was categorized as a control period, a prodromal period, an exacerbation period, or a recovery period; these were defined as:

- Control periods: Periods not influenced by a preceding or following exacerbation.
- Prodromal periods: Periods partly overlapping between control and exacerbation periods.
- Exacerbation periods: Periods with a subsequent occurrence of exacerbation.
- Recovery periods: Periods of two weeks post exacerbation.

Periods with less than three measurements were excluded because this was considered too infrequent for extracting mathematical predictors, as explained subsequently. It is preferable to differentiate between the control and exacerbation periods; thus, the overlapping prodromal and recovery periods were excluded in the prediction. Prodromal periods were considered inconsistent depending on how much the prodromal period overlapped the control period. Therefore, prodromal periods could be considered correctly classified as both a control and an exacerbation period. Recovery periods were used to ensure that a patient had returned to normal before a new period was initiated.

2.4. Predictor extraction

As with the majority of trials [19], the *TeleCare North Trial* suffered from missing data. Therefore, SPSS 23 (IBM Corp, Armonk, NY, USA) was used to analyse the existence of missing data as well as to perform the subsequent imputation as logistic regression is desired for the predictive model. The imputation method used by SPSS is referred to as a fully conditional specification that is an iterative Markov chain Monte Carlo method. For each iteration and each variable, the method fits a univariate model using all other available variables as predictors for the imputation of missing values [20]. Missing data was then imputed using the average of five imputations.

For each period of 30 days, several predictors were derived, including mathematical predictors of physiological parameters, as inspired by [9], questionnaires, and patient and population exacerbation incidence. Mathematical predictors included the difference between the last measurement and the mean, standard deviation, variance, and the slope from linear regression for physiological measurements of saturation, pulse, weight, and blood pressure (systolic, diastolic, and mean arterial) for the 30-day period. Questionnaires included self-reports of increased dyspnea, increased sputum, and sputum purulence if reported to be typically present during exacerbation. A combination of these provided a binary predictor whether these symptoms were all present as a proposed definition for exacerbation by Burge and Wedzicha [16]. Patient exacerbation incidence included the number of exacerbations for a given patient in the past three, six, and twelve months. Population exacerbation incidence included the frequency of exacerbations in other patients within the last month. This method resulted in 29 predictors extracted for every period.

2.5. Predictive model

As described, 29 predictors were extracted from each period. To avoid over-fitting, a maximum of four predictors was allowed in the model with inspiration from the criterion proposed in Jain et al. [21], using at least ten times as many training samples per class as the number of predictors. MATLAB Release 2016b (The MathWorks, Inc., Natick, Massachusetts, United States) was used to train and evaluate a logistic regression for all predictor combinations using a five-fold patient dependent cross-validation. Logistic regressions was chosen due to transparency of the model and robustness to imbalanced data [22]. Each predictor combination was evaluated by the area under the receiver operating characteristics curve (AUC), averaged from the five iterations of cross-validation, to identify the optimal predictor combination with both exclusion and inclusion of patient and population exacerbation incidence as potential predictors. The optimal predictor combination, both with exclusion and inclusion of patient and population exacerbation incidence, was then validated using a two-fold cross-validation with 50/50 partitioning of data using logistic regression. The difference in averaged AUC for the two models was tested using the method proposed by Hanley and McNeil [23].