Telehealth Ecosystems in Practice
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI230732

A Study on Optimization and Evaluation of the Visualization of Complex Algorithm Results in Remote Monitoring of COPD

Mathilde D. SANDER[a,1*], Rikke L. MADSEN[a*], Sisse H. LAURSEN[a,b] and Stine HANGAARD[a]
[a] *Aalborg University, Department of Health and Science, Aalborg, Denmark*
[b] *Research Unit for Clinical Nursing, Aalborg University Hospital, Aalborg, Denmark*
ORCiD ID: Mathilde D. Sander https://orcid.org/0000-0003-2532-7934,
Rikke L. Madsen https://orcid.org/0009-0006-2007-1385
*These authors contributed equally to the study and article

Abstract. Background: Artificial intelligence (AI) can potentially increase the quality of telemonitoring in chronic obstructive pulmonary disease (COPD). However, the output from AI is often difficult for clinicians to understand due to the complexity. This challenge may be accommodated by visualizing the AI results, however it hasn't been studied how this could be done specifically, i.e., considering which visual elements to include. Aim: To investigate how complex results from a predictive algorithm for patients with COPD can be translated into easily understandable data for the clinicians. Methods: Semi-structured interviews were conducted to explore clinicians' needs when visualizing the results of a predictive algorithm. This formed a basis for creating a prototype of an updated user interface. The user interface was evaluated using usability tests through the "Think aloud" method. Results: The clinicians pointed out the need for visualization of exacerbation alerts and the development in patients' data. Furthermore, they wanted the system to provide more information about what caused exacerbation alerts. Elements such as color and icons were described as particularly useful. The usability of the prototype was primarily assessed as easily understandable and advantageous in connection to the functions of the predictive algorithm. Conclusion: Predictive algorithm use in telemonitoring of COPD can be optimized by clearly visualizing the algorithm's alerts, clarifying the reasons for algorithm output, and by providing a clear overview of the development in the patient's data. This can contribute to clarity when the clinicians should act and why they should act on alerts from predictive algorithms.

Keywords. COPD, Telemonitoring, Artificial Intelligence, Usability, Predictive Algorithm, User Interface.

1. Introduction

Over 400 million people suffer from chronic obstructive pulmonary disease (COPD) worldwide [1]. The condition is often associated with exacerbations, i.e., a worsening of symptoms, which can have a major impact on individuals' quality of life [2].

[1] Corresponding Author: Mathilde D. Sander, E-mail: Mathilde.d.sander@gmail.com

Furthermore, the exacerbations are associated with increased mortality and high healthcare costs with exacerbations being responsible for about 50% of COPD expenses [3]. The use of telemonitoring in COPD has shown promising results in improving clinical outcomes and reducing financial costs [4]. The past few years, focus has increased on using artificial intelligence (AI) in telemonitoring [5] and has contributed positively to clinical decision support (CDS) for COPD patients, leading to a decrease in hospitalization and a reduction in financial costs [6]. However, the use of AI has also brought new technical, legal, and medical challenges into practice. Using AI in form of predictive algorithms for CDS has resulted in distrust among clinicians as the algorithms have been assessed as unreliable due to lack of explainability [7]. Hence, outputs from algorithms tend to be non-transparent when working towards more accurate results, while user-friendly interfaces are deprioritized [4, 8]. Prioritizing explainable algorithms, that are easily understood by the end-users, is essential as it has been shown to contribute to constructive CDS tools [9]. Therefore, this study aimed to investigate how complex outputs from a predictive algorithm for patients with COPD could be translated into easily understandable data for the clinicians. Furthermore, the study aimed to develop a user interface for an existing CDS system to achieve a higher level of explainability.

2. Methods

This study is a case study carried out in collaboration with the Danish *TeleCare North organization*. Throughout the study the ISO-standard ISO 9241-11:2018 was used as a theoretical framework.

The data collection included firstly semi-structured interviews to gain insight into the clinicians' needs in relation to making a predictive algorithm more useful as a CDS tool. The purpose of the algorithm was to predict exacerbations in patients with COPD who received telemonitoring as part of their treatment. The clinical implementation of the predictive algorithm is described in detail elsewhere [10]. Two groups of informants were recruited; COPD specialists who had worked with the predictive algorithm and clinicians with knowledge about COPD who did not know the predictive algorithm. The latter were included as representatives for newly hired personnel. The interview guide was designed with the elements from the ISO standard. The data from the interviews was analyzed using Virginia Braun and Victoria Clarke's Reflexive Thematic Analysis [11]. The *Technology Acceptance Model* (*TAM*) and *Aaron Antonovsky's* theory of *Sense of Coherence (SOC*) were used as part of the analytical framework. The findings from the semi-structured interviews were used to create a prototype for an updated user interface.

Furthermore usability tests were carried out as data collection to investigate the usability of the adapted user interface [12]. This was done through a "Think aloud" test that included ten tasks to ensure that the participants tested the entire system. The criteria for inclusion were a clinically relevant background and knowledge about health and technology preferably within telemedicine. The results were structured and processed through the three central concepts from the ISO 9241-11:2018: *Effectiveness, Efficiency* and *Satisfaction*.

3. Results

3.1. Interview findings

The informants included three nurses and one physiotherapist between 25-65 years of age. A total of three main themes and three sub themes were identified. The identified themes were *Lack of understanding of alerts (subtheme: Additional information)*, *Visualization of results (subtheme: Visualization of development)*, *Visual tools (subtheme: Prioritizing*).Overall, the participants found the alerts from the predictive algorithm difficult to identify or understand leading to a lack of *comprehensibility* and *meaningfulness* and thus affecting SOC. The clinicians stated that they interpreted the alerts as errors when they did not understand the basis for the alert. To increase the *perceived usefulness* (*PU*) of the algorithm and increase the *effectiveness,* the participants needed to know the foundation for the alert. Furthermore, they needed to be able to access the patients' data over a period of time as this would improve the *manageability* for the clinicians by providing the necessary resources. This would indirectly increase the clinicians' sense of *meaningfulness*. The participants found that the total usability of the predictive algorithm could be increased by focusing on both the alerts and the development in the patients' data over time. The accessibility of the necessary information makes the system easier to operate and thereby improves the *perceived ease of use* (*PEOU*). Participants described the possibility of a better overview that could ease the clinicians' work processes in both quality and safety by visualizing different data in different ways. Some suggestions included color-coded thresholds for the predictive algorithm and icons. According to the participants, a separate icon for the predictive algorithm would highlight the alerts and make it more accessible and could thus lead to both higher *PEOU* and *satisfaction*. Furthermore, visual tools such as colors and icons were assessed as helpful in prioritizing patients and tasks which would have the potential to increase the *efficiency* and *effectiveness* and provide a useful CDS tool. In addition, the participants stated that visualization of the development in the patients' data could make the clinicians more aware of changes prior to algorithm alerts and potentially reduce hospital stays. This could potentially increase the *effectiveness* and *efficiency* in the usage of the predictive algorithm.

3.2. Usability results

The usability tests were carried out in the developed user interface. The changes from the existing user interface focused primarily on the visibility and explainability of the predictive algorithm. Color changing icons were added to indicate and clarify the result of the algorithm as well as arrows indicating the development in the algorithm's results. Furthermore a graph was added to visualize the development in the underlying data from the algorithm. The participants included five participants, four women and one man, with both academic and clinical backgrounds. The participants found that the user interface was overall *effective* to use. However, they identified a few elements that could be improved to increase the overall *effectiveness* and ensure that tasks are carried out as intended. Three of the participants described the user interface as intuitive and meaningful making it easy to carry out the given tasks. Two of the participants, however, experienced issues with the user interface making it seem less intuitive. The participants found that colors and different symbols contributed to the overall perception of the user interface resulting in increased *efficiency.* Elements such as arrows didn't improve the

efficiency at first as some participants needed to navigate the interface for a while before realizing the purpose of these. Lack of understanding for the included interface icons caused the participants to lose track. Furthermore, according to the participants, *efficiency* is decreased when further information is not permanently visible in the overview of the patient, but instead has to be accessed by navigating through the system.

The participants' overall *satisfaction* with the user interface was impacted by some of the features not being clear to the participants. Furthermore, the participants stated that users must dare to explore the system without being afraid of making mistakes to discover how to navigate it. Regarding *satisfaction,* the participants described the importance of knowing what caused the predictive algorithm alert to be able to assess whether it provides reliable information.

4. Discussion

The aim of this study was to investigate how complex results from a predictive algorithm for patients with COPD could be translated into easily understandable data for the clinicians. The lack of explanation from predictive algorithms is problematic, as it does not provide a basis for CDS, which can lead to mistrust from the users [7]. The results from this study showed that the users' lack of trust in the algorithm could lead to them seeing the alerts as errors. The amount of trust in the predictive algorithm can be a difficult balance in the decision making. On one hand, a certain level of trust is necessary to exploit the possibilities of using AI [6]. On the other hand, too much trust can lead to addiction and misuse of AI [6]. Therefore, as evident from the findings of this study, it is still necessary to adhere to the clinicians' judgment since the predictive algorithms should not be favored over the clinicians' own assessments but should instead support their decisions. Hence, the clinicians expressed that they need additional information about the predictive algorithm. Some clinicians felt that the alert of the predictive algorithm was based on the immediate deviation rather than considering long-term trends. Similar results were found in a study by Ware, et al [13]. However, too many alerts and elements can make it difficult for clinicians to distinguish between the important and less important alerts, which might result in important alerts being ignored. Furthermore, it can disrupt and obstruct the clinicians' workflow [14]. Health professionals' time can be seen as inefficient and make it difficult for the clinicians to familiarize themselves with an algorithm [15]. This is despite the fact that studies show that predictive algorithms reduce time and improve the accuracy of the treatment [15]. Therefore, it is important that the amount of information is adequate as too much information can hamper the work processes [8]. In addition, conflicting experiences of the usefulness of the predictive algorithm can affect a successful implementation, as this requires sufficient usability [13].

5. Strengths and Limitations

One strength of the study is its focus on a minimally documented area considering the scientific literature. Another strength was the inclusion of both end-users and possible end-users with usability expertise to ensure specific and relevant insight to the subject. A third strength is that the case study was performed through an iterative process where the end-users' inputs were included in the development of the user interface and in the

evaluation of the adapted version which ensures relevance for the impacted clinicians. A limitation is the limited number of participants in the study, why theoretical data saturation was not reach [16]. Furthermore, two of four participants had worked with the predictive algorithm but had only experienced a few alerts. This could mean that some of the findings are more hypothetical even though examples of the existing user interface were included.

6. Conclusion

Visualization of complex results from a predictive algorithm for patients with COPD can lead to increased usefulness. However, this needs to apply to the underlying information on which the predictive algorithm relies on, and not only to the alerts generated by the predictive algorithm. The adjustments made to the prototype accommodated this by making elements more accessible and understandable, which helped to strengthen the general applicability of the predictive algorithm. These adjustments help to ensure that the predictive algorithm functions as a supplementary tool for the clinicians. However, further studies with a larger sample are needed to investigate further on the study subject.

References

[1] Disease GI for COL. *Digital Gold Report - Chapter 1*. Elsevier Inc.
[2] Christian Riis H, Jensen MH, Cichosz SL, et al. Prediction of exacerbation onset in chronic obstructive pulmonary disease patients. 40. Epub ahead of print 2016. DOI: 10.3109/03091902.2015.1105317.
[3] McKinstry B. The use of remote monitoring technologies in managing chronic obstructive pulmonary disease. 2013; 106: 883–885.
[4] Cruz J, Brooks D, Marques A. Home telemonitoring effectiveness in COPD: a systematic review. 2014; 68: 369–378.
[5] Barbosa MT, Sousa CS, Morais-Almeida M, et al. Telemedicine in COPD: An Overview by Topics. 17. Epub ahead of print 2020. DOI: 10.1080/15412555.2020.1815182.
[6] Branley-Bell D, Whitworth R, Coventry L. User Trust and Understanding of Explainable AI: Exploring Algorithm Visualisations and User Biases. Epub ahead of print 2020. DOI: 10.1007/978-3-030-49065-2_27.
[7] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. 2020; 58: 82–115.
[8] Blecker S, Pandya R, Stork S, et al. Interruptive Versus Noninterruptive Clinical Decision Support: Usability Study. 6. Epub ahead of print 2019. DOI: 10.2196/12469.
[9] Olczak J, Pavlopoulos J, Prijs J, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. 92. Epub ahead of print 2021. DOI: 10.1080/17453674.2021.1918389.
[10] Secher PH, Hangaard S, Kronborg T, et al. Clinical implementation of an algorithm for predicting exacerbations in patients with COPD in telemonitoring: a study protocol for a single-blinded randomized controlled trial. 23. Epub ahead of print 2022. DOI: 10.1186/s13063-022-06292-y.
[11] Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? 2020; 18: 328–352.
[12] Jordan PW. *An introduction to usability*. CRC Press, 2002.
[13] Ware P, Shah AA, Ross HJ, et al. Challenges of Telemonitoring Programs for Complex Chronic Conditions: Randomized Controlled Trial With an Embedded Qualitative Study. 24. Epub ahead of print 2022. DOI: 10.2196/31754.
[14] Payne TH, Hines LE, Chan RC, et al. Recommendations to improve the usability of drug-drug interaction clinical decision support alerts. 2015; 22: 1243–1250.
[15] Sangers TE, et al. Towards successful implementation of artificial intelligence in skin cancer care: a qualitative study exploring the views of dermatologists and general practitioners. 1. Epub ahead of print 2022.
[16] Kvale S, Brinkmann S. *Interview - Det kvalitative forskningsinterview som håndværk*. Hans Reitzels Forlag, 2015.