



This electronic thesis or dissertation has been downloaded from the University of Bristol Research Portal, <http://research-information.bristol.ac.uk>

Author:

Glyde, Henry Mark M G

Title:

Predictive Modelling Acute Exacerbations of Chronic Obstructive Pulmonary Disease

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited on the University of Bristol Research Portal. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Predictive Modelling Acute Exacerbations of Chronic Obstructive Pulmonary Disease

By

HENRY M G GLYDE



Department of Engineering Mathematics and Technology
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in
accordance with the requirements of the degree of DOCTOR
OF PHILOSOPHY in the Faculty of Engineering.

SEPTEMBER 2023

Word count: 61,356

Abstract

Chronic obstructive pulmonary disease (COPD) is a highly prevalent chronic respiratory disease and a leading cause of disability and death globally. People living with COPD are at risk of periods of sustained worsening of respiratory symptoms beyond their normal stable state and are acute in onset. Current evidence suggests timely and accurate intervention in exacerbations can reduce recovery time, severity, and frequency of exacerbations with marked improvements in health-related quality of life. This PhD thesis aims to describe work on the development of a COPD exacerbation support tool which will be a service within the myCOPD app that can accurately predict acute exacerbations of COPD (AECOPD). A dual literature review of remote patient monitoring (RPM) in randomised controlled trials (RCT) and machine learning for AECOPD prediction highlighted machine learning has the potential to improve patient outcomes. Data analysis of myCOPD patient data led to the creation of labels for exacerbations and stable health. Patient users who matched the labels were more likely to be in a GOLD (Global Initiative for Chronic Obstructive Lung Disease) group greater than A and B, with more frequent use of rescue packs and more engagement with using myCOPD. Modelling of the myCOPD dataset demonstrated the potential exacerbation predictive value of patient-entered data in a real-world digital therapeutic. AdaBoost and EasyEnsemble Classifier models achieved a sensitivity of 67% and 35% and a specificity of 69.5% and 89%, respectively. The Sensing, Predictions, and Alerts in COPD Exacerbations (SPACE) Study I, using thematic analysis, identified that a limited number of patient users understood that the machine learning models were impractical for real-world application. There is a need to further improve model accuracy, and develop a framework to build patient trust and understanding in predictive models. SPACE Study II, a mixed methods study revealed that a group of myCOPD patient users were willing to engage in physiological and functional sensors regularly to support exacerbation prediction. However, the study suggested that maintaining long-term adherence to digital spirometry among these individuals may prove challenging. Analysis of myCOPD patient-entered data and sensor data shows the sensors' predictive capability warrants further study. Continued research following this PhD includes SPACE Study III, using sensor data to develop exacerbation prediction models with greater accuracy and SPACE Study IV an interventional study to identify the safety and efficacy of the finalised intervention.

Dedication and acknowledgements

I want to express my gratitude to the Engineering and Physical Sciences Research Council (GRANT NUMBER: 2275456), my mhealth, and the Bristol Biomedical Research Centre for their financial support throughout my doctoral journey. This support has been instrumental in enabling me to pursue my research goals and contribute to the field of healthcare.

A heartfelt appreciation goes out to the participants of SPACE Study I and II, without whom this research would not have been possible. Your willingness to devote your time and support to this endeavour is deeply valued and has significantly enriched the quality of my work.

I must extend my sincere thanks to two exceptional individuals, Alison and Tom, who generously dedicated their time and expertise to support my work, going above and beyond by offering their insights and guidance consistently on the last Friday of every month. Your contributions were never taken for granted, and I am immensely grateful for everything you did to assist me over the past four years. I also wish to acknowledge the remarkable team at the Academic Respiratory Unit, including Alice, George, and Caitlin. Your guidance has been invaluable, and I cherish the camaraderie we shared as colleagues. Your kindness and friendship made this academic journey all the more rewarding. A special mention is reserved for Ben and, especially, Beth, who played pivotal roles in shaping my approach to qualitative research. Your careful encouragement and thoughtful advice have been a beacon of light throughout my doctoral studies, and I cannot overstate how lost I would have been without your guidance.

I would like to extend my appreciation to my supervisors, James and Ian, for their invaluable mentorship, unwavering support, and kindness. Their guidance has been instrumental in shaping my research and fostering my academic growth. Beyond the academic realm, their genuine care and support have been a constant source of strength during the more challenging moments of my doctoral pursuit. Their willingness to invest not only in my intellectual development but also in my well-being has made my academic journey very rewarding. I am deeply grateful for their mentorship and the impact they have made on my academic career.

Finally, I want to thank my family, including my brother Adam and my sisters Katie and Emily for being there when I needed them the most. I also want to express my immense appreciation to Donna for helping me discover my full potential. Lastly, to my parents, Mark and Diana, for whom I am eternally grateful for their unwavering support in any endeavour I embark on.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific references in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: HGLYDE DATE: 22/09/2023

Table of Contents

	Page
List of Tables	xi
List of Figures	xiii
1 Introduction	1
1.1 Research questions	2
1.2 Contributions	3
1.2.1 myCOPD industry collaboration	3
1.2.2 Dual review and narrative synthesis	3
1.2.3 Data analytics and predictive modelling	4
1.2.4 SPACE Study I and II	5
1.3 Thesis Structure	6
2 Dual Literature Review and Narrative Synthesis	9
2.1 Introduction	9
2.2 Methodology	11
2.2.1 Search strategy	11
2.2.2 Study selection	11
2.3 Results	12
2.3.1 Remote patient monitoring for AECOPD intervention	29
2.3.2 Remote patient monitoring and machine learning for AECOPD prediction	31
2.3.3 Machine learning approaches	33
2.3.4 Study design	36
2.3.5 Patient factors	38
2.3.6 A review of indices in remote patient monitoring	39
2.4 Discussion	43
2.4.1 Limitations	44
2.4.2 Recommendations	45
2.4.3 Conclusion	45

TABLE OF CONTENTS

3 Data Preparation and Analytics	47
3.1 Introduction	47
3.2 Methods	48
3.2.1 The feature window	48
3.2.2 Label creation	50
3.2.3 DataFrame transformation	52
3.2.4 Demographic analysis	54
3.2.5 Feature analysis	54
3.3 Results	55
3.3.1 Label visualisations	55
3.3.2 Engagement analysis	58
3.3.3 Demographic visualisations	58
3.3.4 Feature visualisations	64
3.4 Discussion and conclusions	69
3.4.1 Wrangling	70
3.4.2 Labeling	70
3.4.3 Demographics	72
3.4.4 Features	74
3.4.5 Conclusions	74
4 Exacerbation Predictive Modelling	75
4.1 Introduction	75
4.2 Methodology	76
4.2.1 Data Extraction and Preparation	76
4.2.2 Study Design	77
4.2.3 Initial testing for predictive capacity	78
4.2.4 The application of AdaBoost models	81
4.2.5 Application of t-SNE	85
4.3 Results	86
4.3.1 Initial exacerbation modelling	86
4.3.2 Development of ensemble exacerbation prediction models	87
4.3.3 Misclassification analysis	91
4.3.4 t-SNE for dimensionality reduction and visualisation	92
4.4 Discussion and conclusions	93
5 Sensing, Predictions, and Alerts in COPD Exacerbations Study I	99
5.1 Introduction	100
5.2 Methods	101
5.2.1 Participants	101

TABLE OF CONTENTS

5.2.2	Focus groups	101
5.2.3	Thematic analysis	104
5.3	Results	105
5.3.1	Theme 1: With further improvement the exacerbation prediction models will be acceptable	105
5.3.2	Theme 2: The exacerbation prediction models aren't useful in the real world	110
5.3.3	Theme 3: Seeking understanding and trust in exacerbation prediction models	111
5.3.4	Theme 4: Insights and desires the context of an exacerbation warning .	113
5.3.5	Theme 5: An exacerbation prediction would trigger preparation and caution	118
5.4	Discussion and conclusions	120
5.4.1	Limitations	125
5.4.2	Conclusions	126
6	Sensing, Predictions, and Alerts in COPD Exacerbations Study II	127
6.1	Introduction	128
6.2	Methods	129
6.2.1	Participants	129
6.2.2	Procedure	129
6.2.3	Data analyses	131
6.3	Results	131
6.3.1	Theme 1: Seamless incorporation of daily sensing	133
6.3.2	Theme 2: Accuracy is important to adherence	138
6.3.3	Theme 3: Perceived inaccuracy and usability issues may hinder digital spirometry adherence	139
6.3.4	Theme 4: Adoption of increasing physical activity and calorie tracking .	141
6.3.5	Quantitative use rates of the myCOPD app and sensing technologies .	142
6.3.6	Sensor data quality and exacerbation prediction potential	145
6.4	Discussion and conclusions	157
6.4.1	Limitations	162
6.4.2	Conclusions	162
7	Discussion	165
7.1	Remote patient monitoring and machine learning in acute exacerbations of COPD	167
7.2	Modelling data from a widely deployed digital therapeutic	171
7.3	Patient perspectives on exacerbation prediction models, warnings and sensing technologies	175
7.4	Directions for future research	182

TABLE OF CONTENTS

7.5 Conclusions	185
A Search methods and strategy	187
B Narrative syntheses	189
C SPACE Study I interview schedules	217
C.1 Focus group - exacerbation model accuracy	217
C.2 Focus group - exacerbation warning	217
D SPACE Study II interview schedules	219
D.1 Focus group - sensors	219
D.2 Daily interviews	219
D.3 Midpoint and final interviews	220
Bibliography	223

List of Tables

Table	Page
1 Acronyms and their meanings in order of appearance.	xviii
2.1 Summary descriptions of narrative synthesis of randomised controlled trials of remote patient monitoring in COPD	13
2.2 Summary descriptions of narrative synthesis of machine learning and remote patient monitoring studies in COPD	20
2.3 Summary of common machine learning algorithms in AECOPD prediction from the narrative synthesis	26
3.1 Symptom score DataFrame formed by reading in symptom score CSV file from myCOPD data extract through pandas.	49
3.2 CAT and symptom scores after processing each data point to include any entry 14 days before. nan means there was no entry found for that day.	50
3.3 All of the dynamic DataFrames merged in one DataFrame.	53
3.4 The baseline characteristics merged onto the end of the dynamic DataFrame. . . .	53
3.5 myCOPD app usage comparing the labelled group to the non-labelled group. . . .	58
3.6 myCOPD comparing engagement using the Cooper <i>et al.</i> definition of very high engagement (>20 times every 100 days) between all users who activated the mycopd app, the labelled user group and non-labelled users.	58
4.1 Summary statistics of study participants.	77
4.2 Measures of performance of KNN exacerbation prediction models.	87
4.3 Measures of performance of AdaBoost exacerbation prediction models.	90
7.1 The patient self-report options linked with the respective symptom score and self-report description	172
7.2 A summary of the findings and implications from qualitative research in this thesis	175
7.3 Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist	177
B.1 Narrative synthesis of randomised controlled trials of remote patient monitoring in COPD	189

LIST OF TABLES

B.2 Narrative synthesis of remote patient monitoring and machine learning to predict AECOPD.	207
--	-----

List of Figures

Figure	Page
2.1 Search procedure for randomised controlled trials on remote patient monitoring interventional studies to improve exacerbation-related outcomes in COPD.	13
2.2 Search procedure for empirical studies on remote patient monitoring and machine learning to predict AECOPD.	20
3.1 The opening screen of the myCOPD app when a patient logs on for a new day. The patient user will select one of four symptom scoring options.	50
3.2 Ongoing exacerbation within a single patient-user's exacerbation label. The exacerbation label shows the patient-user reported a symptom score of 3 at -14 days, -8 days, and 4 days in relation to the exacerbation label on day 0. Because the patient-user reported a symptom score of 3 at -14 days and the exacerbation label is between two other symptom scores of 3 the exacerbation label has captured an ongoing exacerbation rather than the initial point the exacerbation started.	55
3.3 Ongoing exacerbation within a single patient-user's stable label. The label shows the patient-user has reported multiple symptom scores of 3 around the stable label at day 0. There are 3 reports of symptom scores 3 before the label and 4 after. This is all within a one-month window. The stable label has captured a period of better health but during an ongoing exacerbation rather than a period of stable health. .	56
3.4 Successful identification of an exacerbation label in a single patient-user. The label shows the patient-user reported a symptom score of 1 at -10 days then a symptom score of 2 at -4 and -3 before reporting a symptom score of 3 at day 0 the exacerbation label. The steady worsening of symptoms over time suggests this label correctly captures the start of an exacerbation.	57
3.5 Successful identification of a stable label in a single patient-user. Day 0 is less than 3 and the surrounding symptom scores are less than 3 suggesting this is a period of stable health and accordingly, day 0 is a stable label.	57
3.6 Comparison of symptom score entry frequency from labelled patient users against non-labelled patient users, using all available patient data.	59

LIST OF FIGURES

3.7 The proportional age distribution comparison between labelled and non-labelled patient users, using all available patient data	60
3.8 The proportional gender distribution comparison between labelled and non-labelled patient users, using all available patient data.	60
3.9 The proportional GOLD group distribution comparison between labelled and non-labelled patient users, using all available data patient data.	61
3.10 the proportional mMRC dyspnea scores distribution comparison between labelled and non-labelled patient users, using all available patient data.	62
3.11 The proportional pack year distribution comparison between labelled and non-labelled patient users, using all available patient data.	62
3.12 The proportional antibiotic and steroid use distribution in the last year between labelled patient users and non-labelled patient users, using all available patient data.	63
3.13 The proportional stable CAT score distribution between labelled patient users and non-labelled patient users, using all available patient data.	63
3.14 The mean CAT scores from all patients, recorded 14 days before an exacerbation or stable label. Error bars signify 95% CI intervals.	64
3.15 The mean tile clicks from all patients, recorded 14 days before and after an exacerbation or stable label. Error bars signify 95% CI intervals.	65
3.16 The mean regular medication use from all patients, recorded 14 days before and after an exacerbation or stable label. Error bars signify 95% CI intervals.	65
3.17 The mean reliever medication use from all patients, recorded 14 days before and after an exacerbation or stable label. Error bars signify 95% CI intervals.	66
3.18 The mean symptom scores from all patients, recorded 14 days before and after an exacerbation or stable label. Error bars signify 95% CI intervals.	66
3.19 The figure illustrates the mean of the change in CAT scores relative to the mean stable CAT score for each patient, calculated 14 days before and following an exacerbation or stable label. Error bars represent 95% confidence intervals.	67
3.20 The figure illustrates the mean of the change in tile clicks relative to the mean stable tile clicks for each patient, calculated 14 days before and following an exacerbation or stable label. Error bars represent 95% confidence intervals.	67
3.21 The figure illustrates the mean of the change in the regular medication relative to the mean stable regular medication use for each patient, calculated 14 days before and following an exacerbation or stable label. Error bars represent 95% confidence intervals.	68
3.22 The figure illustrates the mean of the change in the reliever medication relative to the mean stable reliever medication use for each patient, calculated 14 days before and following an exacerbation or stable label. Error bars represent 95% confidence intervals.	68

LIST OF FIGURES

3.23 The relative importance of dynamic features for AECOPD prediction calculated from Gini impurity.	69
4.1 Schematic of participant selection and study design. A total of 506 participants with 56,329 records in myCOPD had their data randomly split into training (records = 39,430) and testing (records = 16,899) sets, leaving data from all participants spread across both sets. The machine learning algorithms were applied to the training dataset. Both models were tested in the testing set of the myCOPD sample.	78
4.2 Confusion Matrix for the KNN model. TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative.	87
4.3 Confusion Matrix for the KNN model with NearMiss. TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative.	87
4.4 ROC curve comparing the EasyEnsemble model against the AdaBoost with a CST model. Models were trained on 70% and tested on 30% of the myCOPD sample.	88
4.5 ROC curve comparing the EasyEnsemble model against the AdaBoost with a CST model. Models were trained on data from 354 participants and tested on data from 152 participants of the myCOPD sample.	88
4.6 Confusion Matrix for the EasyEnsemble Classifier model. TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative.	89
4.7 Confusion Matrix for the AdaBoost with a CST model. TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative.	89
4.8 Feature importance computed from the AdaBoost with CST on the myCOPD dataset	90
4.9 ROC curve for the AdaBoost with a CST model. The model was trained on the day of an exacerbation and three days prior. The training data contained 426 participants and was tested on data from 107 participants from the myCOPD sample.	91
4.10 Two rows of data extracted from a pandas DataFrame representing a single patient-user. The data was part of the testing set for the AdaBoost with CST model. The first row is data 8 days before an exacerbation and the second is 2 days before an exacerbation. However, the first row is misclassified as stable, while the second is correctly identified as an exacerbation by the model.	92
4.11 Data visualisations in 2D by t-SNE for exacerbation and stable labels. Colour: blue for exacerbation labelled data and red for stable labelled data.	92
4.12 Data visualisations in 2D by t-SNE for confusion matrix classes. Colour: red for true negative, green for false positive, blue for false negative, and purple for true positive.	93
5.1 Statistical representation of model 1.	102
5.2 Statistical representation of model 2.	102
5.3 Design of a passive alert.	103
5.4 Design of an active alert.	103

LIST OF FIGURES

5.5 Design of a support tool.	104
5.6 Thematic map from the focus groups.	105
6.1 Thematic map from the focus group and interviews.	132
6.2 myCOPD use rates. Each dot represents an engagement with the technology for that day. Blue is participant 1, orange is participant 2, green is participant 3, red is participant 4, purple is participant 5	143
6.3 Pulse oximeter use rates. Each dot represents an engagement with the technology for that day. Blue is participant 1, orange is participant 2, green is participant 3, red is participant 4, purple is participant 5	143
6.4 Spirometer use rates. Each dot represents an engagement with the technology for that day. Blue is participant 1, orange is participant 2, green is participant 3, red is participant 4, purple is participant 5	144
6.5 Wrist-worn wearable use rates. Each dot represents an engagement with the technology for that day. Blue is participant 1, orange is participant 2, green is participant 3, red is participant 4, purple is participant 5	145
6.6 Participant 2 pulse oximeter mean pulse rate during one month.	146
6.7 Participant 3 pulse oximeter mean pulse rate during one month.	146
6.8 Participant 2 pulse oximeter mean SpO ₂ during one month.	147
6.9 Participant 3 pulse oximeter mean SpO ₂ during one month.	147
6.10 Participant 2 spirometer peak flow during one month.	148
6.11 Participant 3 spirometer peak flow during one month.	149
6.12 Participant 2 spirometer FEV ₁ during one month.	149
6.13 Participant 3 spirometer FEV ₁ during one month.	150
6.14 Participant 2 wrist-worn wearable resting heart rate during one month.	150
6.15 Participant 5 wrist-worn wearable resting heart rate during the two weeks.	151
6.16 Participant 2 wrist-worn wearable heart rate variability as measured by the root mean square of successive differences (RMSSD) between normal heartbeats during one month.	151
6.17 Participant 5 wrist-worn wearable heart rate variability as measured by the root mean square of successive differences between normal heartbeats during two weeks.	152
6.18 Participant 2 wrist-worn wearable sleep SpO ₂ during one month.	153
6.19 Participant 2 wrist-worn wearable sleep respiratory rate during one month.	153
6.20 Participant 5 wrist-worn wearable sleep SpO ₂ during two weeks.	154
6.21 Participant 5 wrist-worn wearable sleep respiratory rate during two weeks.	154
6.22 Participant 2 wrist-worn wearable steps during one month.	155
6.23 Participant 2 wrist-worn wearable sedentary minutes during one month.	155
6.24 Participant 5 wrist-worn wearable steps during two weeks.	156
6.25 Participant 5 wrist-worn wearable sedentary minutes for two weeks.	156

List of Acronyms

Acronym	Meaning
COPD	Chronic obstructive pulmonary disease
AECOPD	Acute exacerbations of COPD
RPM	Remote patient monitoring
RCT	Randomised controlled trial
SPACE	Sensing, Predictions, and Alerts in COPD Exacerbations
SGRQ	St George's Respiratory Questionnaire
AI	Artificial intelligence
PPI	Patient and Public Involvement
EPM	Exacerbation prediction model
EWS	Exacerbation warning system
HCPs	Healthcare practitioners
PRO	Patient-reported outcomes
GOLD	Global Initiative for Chronic Obstructive Lung Disease
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ECG	Electrocardiogram
SpO ₂	Oxygen saturation
BP	Blood pressure
HR	Heart rate
HRQL	Health-related quality of life
FEV	Forced expiratory volume
6MWT	6-minute walking test
GP	General Practitioner
PR	Pulse rate
CAT	COPD Assessment Test
IRR	Incidence risk ratio
PEF	Peak Expiratory Flow
HADS	Hospital Anxiety and Depression Scale
SF12	Short form health survey

LIST OF FIGURES

BODE	body mass index-airway obstruction-dyspnea-exercise tolerance
RR	Respiratory rate
AUROC	area under the receiver operating characteristic curve
LDA	Linear discriminant analysis
RF	Random forest
SVM	Support Vector Machine
AQCE	Automated Questionnaire for the early detection of COPD Exacerbations
PNN	Probabilistic neural network
PPG	photoplethysmography
PPV	positive-predictive value
NPV	negative-predictive value
PCA	Principal component Analysis
CART	Classification and regression tree
KNN	K-nearest neighbours
LR	Logistic regression
FFNN	Feed-forward neural network
LSTM	Long short-term memory
MTNN	multitask neural network
FVC	Forced vital capacity
SVC	Slow vital capacity
TLC	Total lung capacity
RV	Residual volume
FRC	Functional residual capacity
ERV	Expiratory reserve volume
DLCO	diffusing capacity of the lungs for carbon monoxide
VA	Alveolar volume
DNN	Deep neural network

Table 1: Acronyms and their meanings in order of appearance.

Chapter 1

Introduction

This PhD provides initial evidence for the design of an exacerbation prediction tool that can support patients with chronic obstructive pulmonary disease (COPD) who suffer from acute exacerbations of COPD (AECOPD). COPD is a chronic respiratory disease characterised by airway obstruction, airway inflammation, airway thickening, and in some cases parenchymal destruction (emphysema). COPD accounts for 55% of all chronic respiratory diseases globally [1]. COPD is characterised by intermittent periods of significantly worsening symptoms known as exacerbations [2]. AECOPD in the UK is associated with 14% mortality at three months (184/1342) [3] with the 5-year survival rate <50% for patients hospitalised [4–6]. The rate of hospitalisation related to COPD is notably high at 8.3% (1157 out of 13,939) as reported by Whittaker *et al.* [7]. This high hospitalisation rate gains added significance when considered in conjunction with the comparatively low 5-year survival rate, pointing towards a substantial impact on mortality. Exacerbations often cause an increase in airway and systemic inflammation, increased disease progression and a reduction in quality of life [8–12]. It is estimated that exacerbations in COPD account for 50–75% of COPD-related costs [13]. Patients who suffer frequent AECOPD have more primary care interactions, increased emergency room visitation, increased hospitalisations, and increased admissions to intensive care [14]. A recent research priority-setting partnership in COPD found the highest-rated issue by patients or carers as ‘identify better ways to prevent exacerbations’ [15]. The researchers highlighted the importance of predicting and preventing exacerbations.

Research suggests that reducing delays in treatment (usually a rescue pack; oral corticosteroids and/or antibiotics) and correct identification of exacerbations can lessen the severity of exacerbations, improve health-related quality of life, and reduce recovery time after an exacerbation. Wilkinson et al found that a longer time to treatment in AECOPD was associated with an increase in recovery time of exacerbation symptoms [16]. They demonstrated that a greater number of correctly identified exacerbations treated by a physician resulted in a better health-related quality of life as seen in a reduced total St George’s Respiratory Questionnaire (SGRQ) scores. SGRQ scores are a measure used to assess the impact of respiratory diseases on a person’s quality of life, incorporating various aspects of symptoms, activities, and overall

well-being.

Evidence suggests that people with COPD are in need of AECOPD support and this support can likely improve patient outcomes [15, 16]. However, it is unclear how this intervention may be delivered. There is currently a large body of research that has grown over the last two decades investigating the use of remote patient monitoring (RPM) in COPD [17, 18]. RPM has emerged as a pivotal solution in modern healthcare, enabling continuous monitoring of patients outside conventional clinical settings [19]. The remotely monitored data can then be used by healthcare practitioners to inform decision-making for patients regarding treatment and intervention. The advent of artificial intelligence (AI) in healthcare has seen staggering growth in research using RPM in conjunction with machine learning and deep learning [20]. Combining machine learning with RPM presents a transformative approach to further enhance healthcare delivery. Machine learning techniques, adept at processing and analysing complex and voluminous patient data streams generated by RPM devices, offer the potential to extract valuable insights, predict health trends, and personalise patient care. By harnessing predictive modelling, machine learning algorithms can identify early signs of deteriorating health conditions, enabling timely interventions. Additionally, the integration of machine learning with RPM facilitates the development of personalised treatment plans, optimising interventions based on individual patient profiles and historical data. This combination of RPM and machine learning can support patients to make informed decisions, delivering proactive and tailored care that improves patient outcomes, enhances patient engagement, and ultimately transforms the landscape of healthcare by fostering a data-driven and patient-centric paradigm.

This thesis presents a detailed and thorough overview of the literature using RPM and machine learning in COPD for predicting or improving outcomes in AECOPD. Subsequently, I detail my work developing an exacerbation prediction tool through data analytics, machine learning, patient and public involvement (PPI), and sensor technology.

1.1 Research questions

1. What methods and outcomes exist for digital interventions in AECOPD?
2. What are the gaps in the literature for digital intervention in AECOPD?
3. Can myCOPD data be transformed into a usable and workable format for machine learning?
4. What is the nature of the population used for modelling AECOPD?
5. Can AECOPD prediction models be generated from the myCOPD data?

6. Is the accuracy of current prediction models acceptable to myCOPD users?
7. What does an exacerbation warning look like to myCOPD users?
8. Which sensors be used to improve the performance of AECOPD predictive models?

1.2 Contributions

In the section below I will explain the contributions to research during this doctoral study. I will also explain how these contributions answer the research questions. Contributions answering a specific research question are labelled with RQ_x, x denoting the specific number of the research question being answered.

1.2.1 myCOPD industry collaboration

This PhD involved collaboration with an industrial partner; my mhealth Ltd and its app myCOPD. myCOPD was launched in 2016 by my mhealth Ltd, a digital therapeutic company that developed and supports the app. It is an interactive cloud-based digital self-management app designed to support individuals with COPD. It is available for use in the NHS, with more than 70% of Integrated Care Boards having distributed the app to local healthcare services for use in clinical COPD management protocols. For the patient-user, it provides a platform to enter and store data to track health progression including health status and daily symptom monitoring, a self-management plan with a checklist and target option, and access to up-to-date health and lifestyle information through educational videos including content to encourage exercise with an approved pulmonary rehabilitation course [21]. Patients are also able to record their medication in the medication diary and set reminders.

Collaboration with myCOPD has paved the way for initiating the design and development of an exacerbation prediction tool. The overarching goal is to build and integrate this tool seamlessly into the self-management practices of myCOPD patients. Through research from this PhD, myCOPD users may one day have access to an AECOPD prediction tool, enabling timely and precise interventions. This collaborative effort holds the potential to enhance overall outcomes, offering a proactive approach that may lead to improved health results for individuals managing COPD.

1.2.2 Dual review and narrative synthesis

We have submitted a dual review and narrative synthesis to the Journal of Medical Internet Research (JMIR). The results from this work identified issues with current methods and gaps

in the literature for AECOPD intervention with RPM and machine learning. The numbered statements below are a summary of key findings and recommendations:

1. Introducing machine learning algorithms to predict exacerbations could enable significant improvements in the accuracy of identifying AECOPD at the earliest point, potentially predicting before their occurrence (RQ1).
2. There is a lack of consistency in digital health interventions for exacerbation prevention which limits interpretation and comparisons and may explain variability in the success. There needs to be standardisation of exacerbation labels, study design, and data collection to demonstrate the capability of these systems and allow for further improvements (RQ1).
3. The clinical utility of machine learning approaches will need to be demonstrated with RCT and real-world studies which are currently lacking (RQ2).
4. There is a lack of research into patient factors and there is a need to study the adoption of RPM in the long term and to develop an understanding of the burden of daily/weekly RPM and potential solutions to overcome this burden (RQ2).
5. There is evidence to show suitable methods of RPM that should be incorporated in future research to enable early identification of exacerbations (RQ1).
6. Importantly, these changes and systems will be of little use without engaging patient communities to determine their needs and how such systems could elicit the optimal response and the best outcomes (RQ2).

1.2.3 Data analytics and predictive modelling

Data wrangling of the myCOPD data set resulted in the transformation of raw data into a format suitable for the application of machine learning (RQ3). This involves structuring the data with dynamic features (variables that change frequently, e.g., daily or weekly, such as symptoms and medication use) and baseline characteristics (variables that change annually or remain constant, such as age and gender). Exacerbations and stable states were labelled and applied to this data connected to rows of data if the event occurred 1-8 days after the data entry (RQ3). The population that matched the labels had demographics and characteristics visualised and dynamic features were visualised (RQ4). This data can be used in future work and the findings were poster presented at the European Respiratory Congress - Milan 2023 and will be submitted as a letter to the American Journal of Respiratory and Critical Care Medicine. Using the dataset I demonstrated the potential exacerbation predictive value of patient-entered data in a real-world digital therapeutic (RQ5). The findings from this work were presented at the European Respiratory Congress - Barcelona 2022 and are under peer-review in *Heliyon*

a peer-reviewed open-access journal. The numbered statements below are a summary of key findings:

1. The accuracy of the models is still relatively low.
2. The negative predictive value is high indicating potential utility in supporting patients using rescue packs unnecessarily.
3. Further work is required to improve positive predictive performance.

1.2.4 SPACE Study I and II

The Sensing, Prediction and Alerts in COPD Exacerbations (SPACE) Studies were initiated following the results of the modelling. SPACE Study I and II were completed during this PhD. However, there are plans for the completion of SPACE Study III and IV in postdoctoral research which is discussed in more depth in Chapter 7.

1.2.4.1 SPACE Study I

This study aimed to gather information on myCOPD patient-user perceptions and attitudes towards the exacerbation prediction models and their deployment in the future. This study used focus groups which were analysed thematically to identify the acceptability of the accuracy of current exacerbation prediction models (EPM) and the outline of the design of potential exacerbation warning systems (EWS) (RQ6, RQ7). The numbered statements below are a summary of key findings:

1. Participants were very supportive of the work for the SPACE studies.
2. They expressed dissatisfaction with the current performance of EPM (RQ6).
3. They felt the models didn't need to be much better until they would consider using/trusting them (RQ6).
4. There is a strong possibility trust could be quickly lost with inaccurate predictions (RQ6).
5. Participants also communicated their satisfaction with some preliminary EWS designs requesting the ability to have access to several options with the language revolving around indication or suspicion of risk (RQ7).
6. Including a percentage risk of certainty of prediction could enable self-management decision-making (RQ7).

1.2.4.2 SPACE Study II

In this study, I worked to understand myCOPD patient-user perspectives on sensors and how they might be used to enhance the prediction of exacerbation models. Initially, there was a focus group that was analysed thematically using a grounded theory approach to discern attitudes regarding different sensing modalities. Following on from this focus group I gathered data from sensors, the myCOPD app, 5 days of daily interviews at initiation, a midpoint interview after two weeks, and a final interview at study completion (4 weeks). The numbered statements below are a summary of key findings:

1. Seamless incorporation of daily sensor use in myCOPD users is possible (RQ8)
2. The data from the sensors is usable in conjunction with myCOPD data (RQ8).
3. Further study is warranted to identify the predictive capability of sensors for AECOPD (RQ8).

1.3 Thesis Structure

In Chapter 1, I provide an introduction to the significant impact of AECOPD, the contributions I have made during my PhD, and an overview of the thesis structure.

I provide a narrative synthesis of the currently available evidence through a dual literature review in Chapter 2. This chapter identified the issues surrounding RPM for AECOPD intervention and the need for new studies that use machine learning to improve patient outcomes with the potential for widespread deployment.

Chapter 3 includes data wrangling and analysis of the myCOPD data set. In this chapter, I discuss the transformation of the myCOPD data into a usable format to conduct analytics and enable the application of machine learning techniques. This chapter also contains a description of the process for labelling exacerbations in the myCOPD data set and visualisations of the population demographics extracted from this labelled data and features generated from their app use data.

Following on, I discuss the work I did on a preliminary predictive analysis to identify the potential the transformed myCOPD data and features had for predicting AECOPD using machine learning (Chapter 4). This chapter demonstrates there is a predictive capacity for the models generated from myCOPD data.

--- 1.3. THESIS STRUCTURE

In Chapter 5, I describe the work completed in SPACE Study I, a qualitative study that used focus groups and thematic analysis to identify patient acceptability to the accuracy of prediction models and the design of an exacerbation warning.

SPACE Study II is the focus of Chapter 6 where I report on the findings of a mixed methods study using thematic analysis of a focus group and interviews discussing daily sensor use and quantitative analysis of sensor data and myCOPD data used in conjunction.

In the final chapter of this thesis (Chapter 7), I provide a summary and critical analysis of the findings from the research during my PhD and discuss how these results will inform future research which could lead to potential deployment of an intervention for AECOPD.

Chapter 2

Dual Literature Review and Narrative Synthesis

AECOPD is associated with high mortality, morbidity and poor quality of life and constitutes a substantial burden to patients and healthcare systems [3, 9–12]. New approaches to prevent or reduce the severity of AECOPD are urgently needed. Internationally, this has prompted increased interest in the potential of remote patient monitoring (RPM) and digital medicine. RPM refers to the direct transmission of patient-reported outcomes, physiological, and functional data including heart rate, weight, blood pressure, oxygen saturation, physical activity, and lung function (spirometry) directly to healthcare professionals through automation, web-based data entry or phone-based data entry. Machine learning has the potential to enhance RPM in COPD by increasing the accuracy and precision of AECOPD prediction systems. Here we conduct a dual review of RPM randomised controlled trials (RCT) in AECOPD and machine learning studies combined with RPM to predict AECOPD. We review the evidence and concepts behind RPM and machine learning and go on to discuss the strengths, limitations, and clinical utility of available systems. We have generated a list of recommendations needed to deliver patient and healthcare system benefits. RPM, and in particular the incorporation of machine learning appears to have the potential to improve the predictive capabilities of RPM for AECOPD significantly. Advances in RPM and machine learning require a greater focus on patient co-design, identification and clinical validation of the optimal physiological, behavioural and environmental sensors. This focus should ultimately result in RCT set against usual care to provide an evidence base for their safety, efficacy and cost-effectiveness which could, once present, transform outcomes through the widespread implementation of new approaches to care.

2.1 Introduction

The ultimate objective of the PhD was to work towards developing an exacerbation prediction tool. However, the landscape of the literature was unclear and warranted exploration through a

review. I believe that the available literature should be viewed from two perspectives. The first perspective involves RPM of patients with COPD and requires intensive oversight by healthcare practitioners (HCPs) to promptly address early indicators of deteriorating health in AECOPD cases. This intervention provides real-time insight to HCPs, facilitating the identification of prodromal AECOPD health fluctuations and enabling timely intervention. This in itself justified a review to identify strengths and limitations and to determine if the field has been successful. The second aspect, with comparatively less research, involves machine learning applied to RPM. This reduces the need for regular clinical oversight and holds the promise of enhanced AECOPD prediction accuracy.

Because the first perspective has more available work the gold standard was selected for review, that being RCT. RPM with machine learning has very few RCTS or studies that look into patient-reported outcomes (PRO). Because of this, any study that uses machine learning with RPM was deemed suitable for the second review search.

RPM is a method of healthcare delivery that uses wearable devices and sensors to gather patient data outside of traditional healthcare settings. Using RPM, data becomes available that provides a much clearer and more detailed picture of the patient's health. This data can be reviewed by the patient and a team of healthcare professionals to identify changes in patient health status at the earliest point that may signal an impending exacerbation allowing for timely intervention. Machine learning approaches applied to this data can potentially enhance RPM's predictive capability. It's important to state that the role of RPM is uncertain for AECOPD management. Some studies suggest benefits and others have no effect hence the importance of a review and evidence synthesis to understand the current state of the art and to define the next steps in developing and testing technology to support this global health challenge.

The Global Initiative for Chronic Obstructive Lung Disease (GOLD) has recently adopted a new definition of AECOPD: "an event characterised by dyspnea and/or cough and sputum that worsen over 14 days, which may be accompanied by tachypnea and/or tachycardia and is often associated with increased local and systemic inflammation caused by airway infection, pollution, or other insult to the airways" [22]. Despite efforts to best define AECOPD, there is no gold standard diagnostic test and clinicians will often make a diagnosis based on a patient's symptoms and vital signs. The absence of such a test means there is a lack of standardisation across the literature for the determination or labelling of exacerbation events. The authors of each study generate their own criteria based on their interpretation of the clinical definition and provide a justification for their rationale. Part of this review discusses the approaches to defining AECOPD, providing guidance on the direction for future research to reach a consensus enabling standardisation.

In this dual review, we seek to identify the varying approaches to remote monitoring for exacerbation intervention and prediction, the advancements that machine learning can provide, how exacerbations are labelled, the design of the studies, types of RPM, and a range of approaches to the task of applying machine learning to RPM.

2.2 Methodology

We conducted two literature searches following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [23].

2.2.1 Search strategy

The searches were conducted between April and May 2023 in 2 electronic databases (Scopus and Web of Science) covering publications since the databases began. The first search strategy included search strings in the three main areas: COPD, RPM (telemedicine, telemonitoring, remote patient monitoring, real-time monitoring, telehealth, mhealth, digital health), study design (intervention, trial) and outcome (exacerbation frequency, exacerbation duration, emergency room visits, hospital admissions, hospital readmissions, primary care interaction, healthcare costs, quality of life, days in hospital). The second search strategy also included COPD and RPM but instead of study design and outcome, the search term was machine learning modelling (machine learning, deep learning, prediction models, algorithms). The full search strategies for both searches and for each database are presented in Appendix File A. We conducted a review by selecting articles that met the criteria of being published in peer-reviewed journals or conference proceedings while excluding abstracts, dissertations, systematic reviews, and case studies from our analysis.

2.2.2 Study selection

To be included in the first search, studies were required to (1) specifically examine the use of RPM in COPD; (2) be an RCT; (3) have an exacerbation-related outcome variable i.e. hospital admissions, exacerbation frequency; health-related quality of life (4) be published between the start date of each electronic database and May 2023; (5) be full freely available articles; (6) be published in English.

For the second search, fewer studies were available, therefore, studies were not required to (2) be an RCT and instead of including (3) an exacerbation-related outcome variable, studies were required to incorporate (7) a form of AI modelling, usually machine learning algorithms for exacerbation prediction.

Papers were excluded from the study for any one of the following reasons: (a) the study is a systematic literature review, (b) the study did not include one outcome related to either (3) (first search) or (7) (second search), (c) for the first search the focus of the intervention was behaviour change (physical activity (PA), medication adherence, inhaler technique) or remote rehabilitation (usually pulmonary rehabilitation) rather than remote monitoring, (d) the main study outcome was cost and does not include patient-related outcomes or machine learning predictions. Studies incorporating machine learning or variations were not included in the first search but were referred to in the second search even if they were a randomised controlled trial with RPM.

Two authors independently assessed the results obtained from the first literature search. Articles were screened in four steps: first, duplicates were removed, and then the title, abstract, and keywords were screened. Articles were screened on the inclusion and exclusion criteria outlined above. If authors could not determine suitability during the screening, full-text articles were assessed for inclusion criteria and exclusion criteria. Full-text articles were excluded for not reporting outcomes for COPD patients or the accuracy of COPD exacerbation prediction separately (in the case of studies with multiple diseases).

Remote monitoring studies were not included in the review if they reported on remote monitoring as an alternative to hospitalisation for exacerbation treatment, did not include the specifics of the RPM use, or if they focused on the diagnosis of AECOPD rather than prediction. The lead author then extracted the information from each study that met the inclusion criteria and used it to create a narrative synthesis: participant characteristics, study design, RPM used, and results. The narrative synthesis of the dual reviews is shown in Tables 2.1 and 2.2 which the remaining authors checked.

2.3 Results

We screened and analysed data from April to May 2023. Through the first search, we identified 216 studies. Of these, 29 were included in the review (Figure 2.1).

2.3. RESULTS

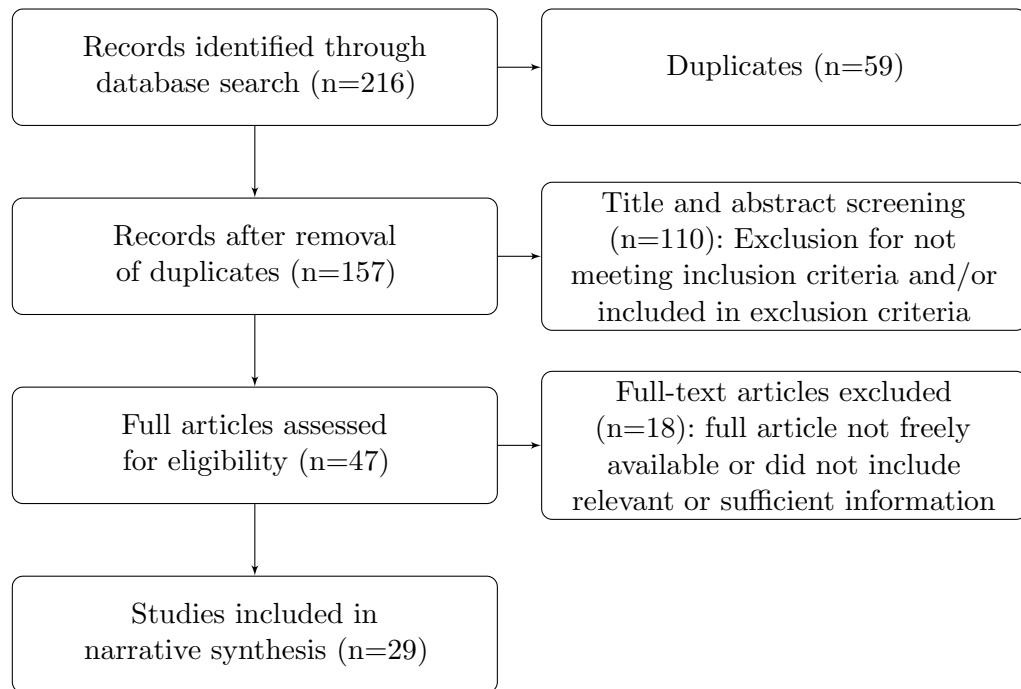


Figure 2.1: Search procedure for randomised controlled trials on remote patient monitoring interventional studies to improve exacerbation-related outcomes in COPD.

Table 2.1 provides a summary of a narrative synthesis from the studies identified in the first search. The full narrative synthesis can be found in table B.1.

Table 2.1: Summary descriptions of narrative synthesis of randomised controlled trials of remote patient monitoring in COPD

Authors	Summary description
Toledo <i>et al.</i> [24]	Using One-lead ECG, lung function, oxygen saturation (SpO ₂), blood pressure (BP), heart rate (HR) and PRO for RPM, there was no significant difference between the intervention (n=67) and control group (n=90) in days to hospital admission, the number of admissions and health-related quality of life (HRQL). Exacerbations were not defined; patients were monitored for abnormal readings.
Koff <i>et al.</i> [25]	Using PRO (dyspnea, cough, sputum, fever) and self-report (activity, depression, forced expiratory volume in 1 second (FEV ₁)), 6-minute walking test (6MWT), and SpO ₂) for RPM, there was a significant improvement in HRQL in the intervention group (n=20) compared to the control group (n=20). The patients were monitored for changes in indices; exacerbations were determined retrospectively.

Vitacca <i>et al.</i> [26]	Using PRO (dyspnea, cough, sputum, sputum colour, weight/ankle oedema, temperature, neurological status, ventilator interaction (if used as part of care), and walk) and self-report (SpO ₂ and HR) for RPM, there was a significant reduction in hospitalisations, exacerbations, urgent contact to GP and admission to the emergency room in the intervention group (n=57) compared to the control group (n=40). Exacerbations were not defined; clinical variation from an established baseline resulted in contact.
Halpin <i>et al.</i> [27]	Using PRO (breathlessness, cough, congestion and fatigue) for RPM, there was no significant difference between the intervention (n=40) and control group (n=39) in exacerbation frequency, duration, and severity, and HRQL. Exacerbations were defined using the EXACT PRO (a 14-item daily diary); responses to a subset of diary items were used as a trigger to contact the patient.
Dinesen <i>et al.</i> [28]	UsingBP, pulse rate (PR), weight, SpO ₂ , and lung function for RPM, there was a significant reduction in hospitalisations in the intervention group (n=57) compared to the control group (n=48) but no significant difference in length of stay or cost. Exacerbations were not defined; Healthcare professionals at a hospital could assess the patient's data and provide advice to the patient.
De San Miguel <i>et al.</i> [29]	Using BP, weight, temperature, PR, SpO ₂ and general state of health for RPM, there was no significant difference between the intervention (n=36) and control group (n=35) in hospitalisation, admission to the emergency room, and length of stay but there was a significant cost reduction. Exacerbations were not defined; patients measured their physiological parameters and answered health-related questions which were monitored daily by a nurse. Any deviations outside of the patient's normal parameters triggered an alert.
Jehn <i>et al.</i> [30]	Using Daily COPD Assessment Test (CAT), daily lung function, and weekly 6MWT (accelerometry) for RPM, there were significantly fewer exacerbations, shorter hospital stays and significant improvement in CAT in the intervention group (n=32) compared to the control group (n=30) but no significant difference in primary care interactions, lung function or 6MWT. Exacerbations were defined as a significant worsening of respiratory symptoms requiring a change in medication (rescue pack) and the presence of at least one of increased dyspnea, increased sputum production, and change in sputum purulence.
Pedone <i>et al.</i> [31]	Using Wristband monitor (HR, physical activity (PA), temperature, and galvanic skin response), and SpO ₂ for RPM, the intervention group (n=50) experienced fewer single events and a reduction in the incidence of multiple events compared to the control group (n=49). The physician defined exacerbations via abnormal readings.

2.3. RESULTS

Pinnock <i>et al.</i> [32]	Using PRO (dyspnea, sputum purulence, sputum volume, cough, wheeze) medication self-report, SpO ₂ , and upper respiratory infection and fever identification questionnaires for RPM, there was no significant difference between the intervention (n=128) and control group (n=128) in the number of days to admission, the mean number of admissions, and the duration of admissions, or HRQL. exacerbations were defined as a sustained worsening of patient symptoms from their usual stable state, beyond normal day-to-day variations, acute in onset, necessitating a change in treatment. Algorithms, based on the symptom score, alerted the clinical monitoring team for an exacerbation.
Sorknaes <i>et al.</i> [33]	Using PR, SpO ₂ , and lung function for RPM, there was no significant difference between the intervention (n=132) and control group (n=134) in mortality, time to readmission, the mean number of hospital readmissions, and the mean number of readmission days. exacerbations were not defined; a nurse collected the patient measurements remotely and could organise rapid treatment.
Calvo <i>et al.</i> [34]	Using BP, SpO ₂ , HR and peak expiratory flow (PEF) for RPM, the intervention group (n=29) experienced a significant reduction in emergency department visits, hospitalisations, duration of hospital stay, need for non-invasive mechanical ventilation, and an increase in time to the first severe exacerbation compared to the control group (n=30). A clinical alert was generated if indices exceeded pre-established limits, the pulmonologist would then classify the exacerbation as moderate, severe or very severe.
McDowell <i>et al.</i> [35]	Using BP, HR, SpO ₂ and PRO (dyspnea, cough, sputum, tiredness) for RPM, the intervention group (n=55) experienced a significant improvement in HRQL compared to the control group (n=55) but not in emergency department visits, hospital admissions or exacerbations. Exacerbations were not defined; an alert signalled abnormal clinical observations or a positive response to symptom-related questions from the patient. The community respiratory team determined home visit or hospital admission.
Ringbæk <i>et al.</i> [36]	Using lung function, SpO ₂ , weight, and PRO (dyspnea, sputum colour, volume, and purulence for RPM, there was no significant difference between the intervention (n=141) and control group (n=140) in hospital admissions, time to first admission, or all-cause hospital admissions. However, there was a significant reduction in the severity of exacerbations in the intervention group compared to controls. Exacerbations were not defined; Observations were automatically categorized (green/yellow/red) and prioritised. Respiratory nurses could connect patients to respiratory medicine specialists for alarming values (single red or two yellows).

Cordova <i>et al.</i> [37]	Using PEF and PRO (dyspnea, and sputum quantity, colour, and consistency, cough, wheeze, sore throat, nasal congestion, and high temperature) for RPM, there was no significant difference between the intervention (n=34) and control group (n=33) in hospitalisation rates, hospital duration, or mortality. However, the intervention group's lung function (PEF) and HRQL of life significantly improved and were sustained for up to 24 months but were unchanged in the control. Exacerbations lacked a specific definition; instead, changes from the patient's initial symptom values were identified by an algorithm based on a symptom severity index that assigned weights to individual symptoms.
Ho <i>et al.</i> [38]	Using self-report (SpO ₂ , temperature, BP) and PRO (disease-related symptoms for RPM, the intervention group (n=53) experienced a significant reduction in COPD-related readmission, the number of all-cause readmission, emergency room admission, and an increase in time to readmission compared to the control group (n=53). Patients reported via an electronic diary; submitted data underwent processing using a predefined algorithm established by the study team. When a warning was triggered, the pulmonologist assessed the patient's data.
Vianello <i>et al.</i> [39]	Using self-report every other day (HR and SpO ₂) and a single measurement of SpO ₂ once a day during periods of clinical worsening for RPM, there was no significant difference between the intervention (n=230) and control group (n=104) in HRQL and hospital admission rate. However, hospital readmission was significantly lower in the intervention group. Operators reviewed the online data of each patient and if the HR and/or SpO ₂ values were outside of the patient's "normal" range, patients were contacted and then clinical staff were alerted after abnormal secondary measure.
Farmer <i>et al.</i> [40]	Using PRO, SpO ₂ , and HR for RPM, there was no significant difference between the intervention (n=110) and control group (n=56) in HRQL and risk of hospital admission rate. However, generic health status was significantly better in the intervention group and the median number of visits to general practitioners and practice nurses was lower. Respiratory clinicians reviewed the data; alerts were generated when PR, SpO ₂ , or the symptom score went above the safety threshold.
Lilholt <i>et al.</i> [41]	Using PRO (questions related to COPD exacerbations and symptoms), BP, HR, weight and SpO ₂ for RPM, there was no significant difference between the intervention (n=258) and control group (n=316) in HRQL. No exacerbation definition; Healthcare personnel monitored and contacted patients if there were adverse changes in their values and responses.

2.3. RESULTS

Rixon <i>et al.</i> [42]	Using SpO ₂ , BP, weight, and questions about health for RPM, there was no significant difference between the intervention (n=275) and control group (n=172) in HRQL. lacking exacerbation definition; Data was transmitted to a monitoring centre, where it was reviewed by health-care professionals, triggering an appropriate response.
Kessler <i>et al.</i> [43]	Using lung function, SpO ₂ , and HR, with participants on long-term oxygen therapy also having daily oxygen use and respiration rate for RPM, there was no significant difference between the intervention (n=157) and control group (n=162) in exacerbation frequencies, time to first exacerbation, and all-cause hospitalisation days per year. However, the intervention group had significantly fewer acute care hospitalisation days per year, a lower body mass index-airway obstruction-dyspnea-exercise tolerance (BODE) index and a lower mortality rate. The data was transmitted to a hospital physician who made all decisions regarding exacerbation intervention.
Soriano <i>et al.</i> [44]	Using SpO ₂ , BP, lung function, and respiratory rate (RR) and oxygen therapy compliance for RPM, there was no significant difference between the intervention (n=115) and control group (n=114) in the proportion of participants who had an AECOPD, the number of all-cause deaths, total resource utilization cost. Clinical alerts are generated based on deviations in various health indices. These include a drop of 5% or more in SpO ₂ , an HR exceeding 100 bpm, an increase of SBP by 15mmHg above the baseline value, a 5% or greater increase in RR, and specific criteria related to spirometry.
Tupper <i>et al.</i> [45]	Using PRO (dyspnea, sputum colour, volume, and purulence), lung function, SpO ₂ , and weight for RPM, there was a significant difference in improvement between the intervention (n=141) and control group (n=140) in the 15D instrument but there was no significant difference in CAT score. The observations were transferred to a call centre at each participant's local hospital and automatically categorised (green/yellow/red). Abnormal measures were responded to by the specialist nurse and conferred with a respiratory specialist.

Walker <i>et al.</i> [46]	Using Within-breath respiratory mechanical impedance using forced oscillation technique for RPM, there was no significant difference between the intervention (n=135) and control group (n=158) in time-to-first-hospitalisation, HRQL, antibiotic prescriptions, or hospitalisation rate. An intention-to-treat analysis indicated fewer repeat hospitalisations. The algorithm produced respiratory alerts when it identified a deteriorating trend in at least one of the following Forced Oscillation Technique (FOT) parameters measured at 5 Hz: inspiratory resistance, inspiratory reactance, or the difference between inspiratory and expiratory reactance—an indicator of tidal expiratory flow limitation.
Mínguez Clemente <i>et al.</i> [47]	Using ECG (leads I, II and III), SpO2, HR, BP, temperature, and RR for RPM, there was no significant difference between the intervention (n=58) and control group (n=58) in baseline characteristics, time until first exacerbation, the number of exacerbations or costs. However, There was a significant decrease in the number of healthcare visits observed in the intervention compared to the control. Normality thresholds were set for the indices; readings outside thresholds generated an alert in the form of a text message sent to the physician.
Sink <i>et al.</i> [48]	Using PRO - breathing better, same, worse for RPM, the intervention group (n=83) experienced a significantly longer time to first hospitalisation and fewer hospitalisations compared to the control group (n=85). If a participant responded “worse” to breathlessness, the system promptly activated an alert to notify the subject’s designated healthcare provider.
Koff <i>et al.</i> [49]	Using PRO, SpO2, lung function, 6MWT (pedometer), and post-exertion SpO2 for RPM, the intervention group (n=352) experienced significant improvements in HRQL and BODE index and reductions in urgent office visits compared to the control group (n=159). However, mortality, hospitalisation, ICU visits, emergency department visits, and length of stay were not significantly different between groups. data was analysed by predetermined algorithms and categorised (green/yellow/red). Red indicates a possible decline in health status (breathless at rest, sputum purulence, fever, no walk, SpO2 <87, FEV ₁ >25% below baseline).

2.3. RESULTS

Rassouli <i>et al.</i> [50]	Using PRO (dyspnea, sputum volume, sputum colour, cough, fever, emergency medication) for RPM, the intervention group (n=168) experienced significant improvements in HRQL. However, there was no significant difference in the intervention vs. control in the emergency department visit rate, hospitalisation rate, number of exacerbations, days in hospital due to exacerbations, and cost. A yellow box served as the initial warning (when two or more questions were answered “yes”), but no immediate action was taken. If the yellow warning persisted and turned red the next day (indicating two consecutive “yes” answers), it signalled a potential Acute Exacerbation of Chronic Obstructive Pulmonary Disease (AECOPD).
Anderson <i>et al.</i> [51]	Using SpO ₂ , HR, PEF, weight, PRO (dyspnea, cough, sputum volume, and colour) for RPM, the intervention group (n=110) experienced a reduction in hospitalisation compared to the control group (112). However, there was no significant difference in the intervention vs. control in the time to first hospitalisation. An algorithm triggered respiratory alerts in the event of a deteriorating trend. Three categories were defined (green/yellow/red); a red alert was activated for abnormal measurements. In instances of Yellow or Red alerts, the respiratory nurse contacted the patient on the same day for exacerbation intervention.
Køpfli <i>et al.</i> [52]	Using SpO ₂ , HR, lung function, weight, and PRO (dyspnea, cough, sputum volume and sputum colour) for RPM, there was no significant difference in the intervention (n=101) vs control (n=97) in HRQL. Using the patient’s reported measurements and symptoms, three categories were established (green/yellow/red); yellow was a change in symptoms while red signalled a change in physiological readings. The incoming data were assessed by hospital respiratory nurses, who, in collaboration with a respiratory physician, determined if exacerbation intervention was necessary.

The second literature review search identified 350 articles. Of these, 23 were included in the review (Figure 2.2).

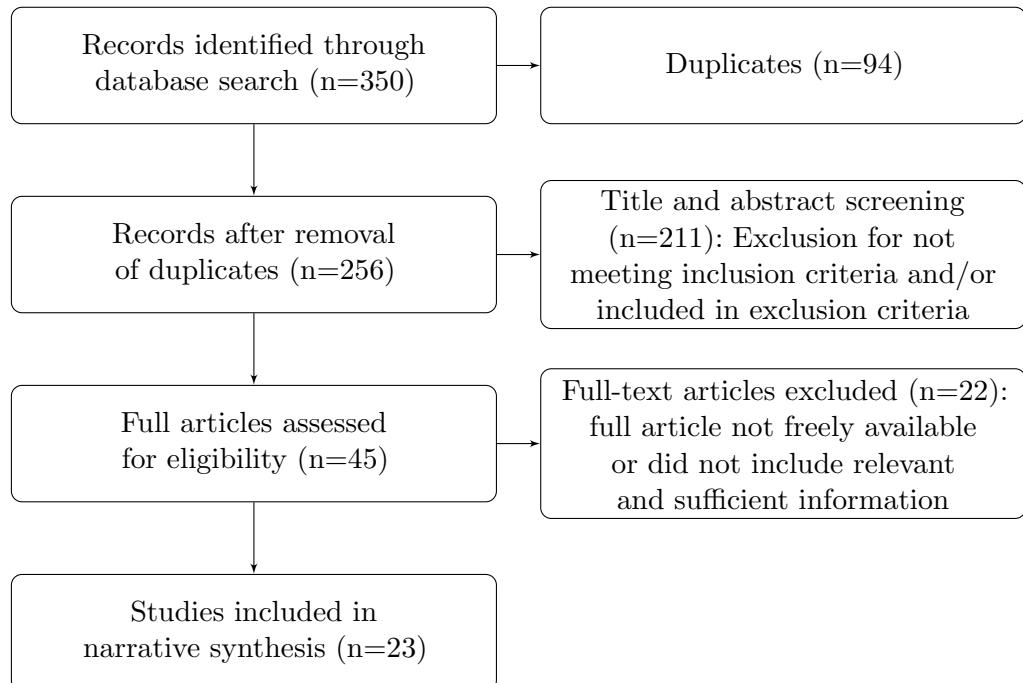


Figure 2.2: Search procedure for empirical studies on remote patient monitoring and machine learning to predict AECOPD.

Table 2.2 provides a summary of a narrative synthesis from the studies identified in the second search. The full narrative synthesis can be found in table B.2. Table 2.3 summarises and describes the different algorithms applied to RPM for AECOPD prediction.

Table 2.2: Summary descriptions of narrative synthesis of machine learning and remote patient monitoring studies in COPD

Authors	Summary description
Jensen <i>et al.</i> [53]	Using weight, BP, SpO ₂ and, lung function for RPM and linear discriminant analysis (LDA) for modelling. The model has a sensitivity of 70%, a specificity of 95%, and an AUROC of 0.73 for predicting AECOPD. Exacerbations were hospital admissions and/or patient records indicating the administration of a rescue pack in combination with one or more exacerbation-specific symptoms: increasing cough, purulent expectorate or fever. There were a total of 10 exacerbations in nine patients. Test episodes are defined as the 30-day interval before the exacerbation, followed by recovery periods of 14 days after the exacerbations. Control episodes are defined as the 30-day intervals after the recovery period.

2.3. RESULTS

Heijden <i>et al.</i> [54]	Using FEV ₁ , SpO ₂ , PRO (dyspnea, sputum production, sputum colour, cough, wheezing, activities, malaise, fever) for RPM and a Bayesian network for modelling. The model performance on a different data set includes an AUROC of 0.87, a sensitivity of 88%, a false-positive rate of 0.2 and an accuracy of 81% for predicting AECOPD. Exacerbations were defined as an acute event of worsening of COPD health status.
Bellos <i>et al.</i> [55]	Using ECG, RR, steps, standing and lying time, cough, snoring, temperature, BP, SpO ₂ , lung function, blood glucose, weight, PRO (dyspnea, cough, sputum), and demographics (lifestyle, mental status) for RPM and a hybrid classification system combining random forest (RF), support vector machine (SVM), and a rule-based system for modelling. The model achieved an accuracy of 94% for current disease severity estimation, which can be used to recognise an abnormal health episode or an AECOPD. the hybrid classification system was used to define exacerbations.
Fernández-Granero <i>et al.</i> [56]	Using PRO through the Automated Questionnaire for the early detection of COPD Exacerbations (AQCE) (general health status, cough, phlegm, dyspnea, sleep, cold symptoms, lung sounds and coordination test) for RPM and a probabilistic neural network (PNN) for modelling. The model has an accuracy of 88.3%, a sensitivity of 80.5%, a specificity of 94.34%, a positive predictive value (PPV) of 91.67%, and a negative predictive value (NPV) of 86.21% for predicting AECOPD. Exacerbations were defined according to Anthonisen criteria [57]
Shah <i>et al.</i> [58]	Using PRO (general health status, breathlessness, wheeze, cough, sputum colour, presence of cold, sleep breathing), SpO ₂ , HR, RR, and photoplethysmography (PPG) waveform for RPM and multivariate novelty detection using Parzen windows for modelling. The model achieved an AUROC of 0.9 for predicting AECOPD. Self-reported use of medications was an indicator of exacerbations. For every medication event, a premonitory period of three days was defined.
Fernández-Granero <i>et al.</i> [59]	Using PRO (breathlessness, cough, sputum, chest symptoms, difficulty bringing up sputum, fatigue, sleep disturbance, and health anxiety) for RPM and a PNN for modelling. The model has an accuracy of 88.3%, a sensitivity of 80.5% and a specificity of 94.3% for predicting AECOPD. The exacerbations of COPD Tool (EXACT) was used to define exacerbations.

Fernández-Granero <i>et al.</i> [60]	Using respiratory sounds through a respiratory sensor embedded in a self-tailored housing for RPM and principal component analysis (PCA) and an SVM for modelling. The model achieved a sensitivity of 73.76%, a specificity of 97.67%, a PPV of 84.66% and an NPV of 95.53% for predicting AECOPD. Exacerbation onsets (accounted for self-administration of medication, unscheduled visits to emergency units and/or admissions) and the previous seven days were labelled with “1” and the rest of the days were labelled with “0”. Periods of two weeks after the AECOPD, corresponding to recovery periods, were discarded.
Mohktar <i>et al.</i> [61]	Using lung function, SpO ₂ , RR, HR, temperature and weight for RPM and classification and regression tree (CART) for modelling. The model has an accuracy of 71.8%, specificity of 80.4%, and 61.1% sensitivity for predicting AECOPD. They had any two of the three (worsening sputum volume, sputum purulence, and breathlessness), and had started oral corticosteroids and/or antibiotics on the same day, and this day was not within two weeks of the previous exacerbation or they had any one of the three symptoms and had increased their rescue inhaler use on the same day, and this day was not within two weeks of the previous high risk of exacerbation.
Sanchez-Morillo <i>et al.</i> [62]	Using PRO via AQCE: general health status, cough, phlegm, dyspnea, sleep, cold symptoms, lung sounds, and coordination test for RPM and K-means clustering for modelling. The model has a sensitivity of 73%, a specificity of 74%, a PPV of 69% and an NPV of 78% for predicting AECOPD. Emergency department admission, primary care access, or self-administration of antibiotics and/or corticosteroids was considered in the definition of exacerbation. Each day in the study period was assigned an output label: “1” for the exacerbation onset and the previous 7 days and “0” otherwise.
Christian Riis <i>et al.</i> [63]	Using PRO, BP, SpO ₂ , and PR for RPM and k-nearest neighbours (KNN) for modelling. The model has a sensitivity of 73%, a specificity of 74%, a PPV of 69% and an NPV of 78% for predicting AECOPD. Hospitalisation or a 10-day period of self-treatment involving antibiotics and/or steroids is characterized as an exacerbation. The exacerbation dates served as reference points to divide measurements into 30-day episodes. Segments occurring before an exacerbation date were labelled as an exacerbation. Episodes falling between a recovery period of 14 days and a subsequent exacerbation were stable.

2.3. RESULTS

Shah <i>et al.</i> [64]	Using PR, RR, and SpO ₂ for RPM and logistic regression (LR) for modelling. The model has an AUROC of 0.682% (95% CI 0.681-0.682), an 80% sensitivity and 36% specificity, or 60% sensitivity and 68% specificity for predicting AECOPD. An exacerbation is defined as increased use of a reliever inhaler for at least 48 hours, starting oral steroids and/or antibiotics, presence of at least two symptoms (one of which should be a major symptom), or an HCP contact. They established a 7-day timeframe preceding an exacerbation event as the prodromal period. Likewise, stable periods are characterised as 7-day intervals where a patient remains in a stable state. These periods include a 7-day buffer at both the beginning and end of the defined timeframe.
Fernández-Granero <i>et al.</i> [65]	Using respiratory sounds were recorded daily with an electronic sensor for RPM and RF for modelling. The sensitivity of the model is 78.1%, the specificity is 95.9%, and the PPV is 94.1% for predicting AECOPD. The symptom-based definition of AECOPD was used in this study according to the Anthonisen criteria [57].
Kronborg <i>et al.</i> [66]	Using PRO (symptoms, cough, mucus, shortness of breath, inhaler use, antibiotic use), SpO ₂ , PR, weight, and BP for RPM and LR for modelling. The model achieved an AUROC of 0.74 for predicting AECOPD. Exacerbations were defined as hospitalisations. Data was organised into periods of 30 days. Each period was categorised as either control (not influenced by a preceding or following exacerbation), prodromal (partly overlapping between control and exacerbation), exacerbation (subsequent occurrence of exacerbation), and recovery (two weeks post exacerbation).
Nunavath <i>et al.</i> [67]	Using daily measurements of a symptom-specific questionnaire, SpO ₂ , and results of automatically generated health status overview for RPM and Deep Artificial Neural Networks; Feed-Forward Neural Networks (FFNN) for the classification of COPD patient's health category, and Long Short-Term Memory (LSTM) for modelling. The LSTM model was able to predict COPD patients' health conditions one day with an accuracy of 84.12%. The data underwent labelling, where instances were categorized as follows: stable patient condition (1), significant deterioration (2), and an urgent need for follow-up (5). Subsequently, classification was carried out using a feed-forward neural network (FFNN). This FFNN was constructed with an input layer comprising 5 neurons, a hidden layer containing 8 neurons, and an output layer consisting of 3 neurons.

Orchard <i>et al.</i> [68]	Using daily symptoms, physiological measures, and medication data, with baseline demography, COPD severity, quality of life, and hospital admissions linked with meteorological data for RPM and multitask neural networks (MTNN) for modelling. The model has an AUROC of 0.74 (95% CI 0.67-0.80) for predicting AECOPD. Five definitions were used including 2 major symptoms (Breathlessness, sputum colour, and sputum amount), 5 symptoms (including minor symptoms; cold, wheezing, sore throat, cough, and fever), 2 major or 1 major and minor followed by two bad days, and 5 symptoms or 4 symptoms on two consecutive days.
Boer <i>et al.</i> [69, 70]	Using 12 yes-or-no PRO, SpO ₂ , FEV ₁ , and temperature (forehead thermometer) for RPM and Bayesian network for modelling. The model has a sensitivity of 97.4%, a specificity of 65.6%, and a PPV of 13.4%. In an RCT there was not a statistically significant difference between the intervention group and the control group in exacerbation-free weeks, HRQL, or healthcare utilization. symptom and event-based definitions; a change for at least 2 consecutive days in two or more major symptoms or a change in any one major symptom plus one or more minor symptoms and contact with HCP that led to a new prescription of steroid and/or antibiotics or hospitalisation. A new exacerbation episode was preceded by at least 2 days without exacerbation.
Jin <i>et al.</i> [71]	Using non-invasive ventilator: airflow, pressure and SpO ₂ for RPM and SVM, RF and LDA for modelling. The best-performing model is the LDA model with an accuracy of 74.5%, a sensitivity of 77.6%, and a specificity of 42.9% for predicting AECOPD. The defined exacerbations as “an acute event characterised by a worsening of the patient’s respiratory symptoms from the stable state and beyond day-to-day variation, leading to a change in medical treatment and/or hospitalisation”.
Iadanza <i>et al.</i> [72]	Using lung function for RPM and C5.0 for modelling. The model has a sensitivity of 98.9%, a specificity of 96.2%, and an accuracy of 97.4% for predicting AECOPD. In the course of medical evaluations, patients were queried about any instances of exacerbations, hospital admissions, or the need for emergency care. Using this information, the risk of exacerbation was evaluated and categorised as either low or high.

2.3. RESULTS

Kronborg <i>et al.</i> [73]	Using SpO ₂ , PR, and BP for RPM and SVM for modelling. The model has an AUROC of 0.95 and a sensitivity of 94% for predicting AECOPD. Continuous measurements were organised retrospectively into 14-day periods using a moving window approach. Each period was categorised as either a control period if no subsequent exacerbation followed or a prodromal exacerbation period if there was a subsequent exacerbation. Exacerbations were identified based on events such as hospitalisation or self-initiated 10-day antibiotic and/or steroid treatment.
Patel <i>et al.</i> [74]	Using PRO, lung function, C-reactive protein (CRP) (finger-prick testing) for RPM and decision tree for modelling. The model has a sensitivity of 97.9%, a specificity of 84.0%, and a PPV of 38.4% for predicting AECOPD. In a prospective interventional study, compared to the 6 months pre-study, hospitalisations were reduced by 98% (90 vs 2, $p<0.001$). A mild/moderate exacerbation was defined as a sustained increase in respiratory symptoms for two consecutive days or more, involving at least two major symptoms or one major symptom and one minor symptom that required the initiation of steroids and/or antibiotics. A severe exacerbation was defined as an episode requiring hospitalisation. The prodrome of an exacerbation was identified as the 14-day period preceding the onset of a clinician-defined exacerbation.
Wu <i>et al.</i> [75]	Using Environmental, physiological, PRO, and quality of life data for RPM and deep neural network (DNN) for modelling. The model has an f1 of 0.923, an accuracy of 92.1%, a sensitivity of 90.4%, a specificity of 94% and an AUROC 0.964 for predicting AECOPD. COPD exacerbations were defined as the acute worsening of respiratory symptoms, resulting in additional therapy.
Chmiel <i>et al.</i> [76]	Using PRO (CAT score, symptom score) demographics (age, gender, smoking years, smoking status) for RPM and RF with Youden's J statistic as the threshold for modelling. The model achieved an AUROC of 0.727, a sensitivity of 76%, and a specificity of 63% for predicting AECOPD. A symptom-based criterion was employed to define exacerbation events, encompassing increased breathlessness, altered sputum characteristics, and the need for self-medication using steroids or antibiotics. Alternatively, exacerbation events were identified when a patient experienced significantly worsened breathing despite treatment, accompanied by chest pain or a high fever, leading to the necessity of seeking emergency care or hospital admission. The definition utilised a binary variable to signify whether a reported symptom episode was succeeded by an exacerbation event within the subsequent 3 days.
Wu <i>et al.</i> [77]	Using air pollution, PA, PRO and PEF for RPM and a cost-effective RF for modelling. The model has an accuracy of 88.6%, a sensitivity of 77.8%, a specificity of 94.9% and an f1 of 0.833 for predicting AECOPD. Lacking exacerbation definition.

Table 2.3: Summary of common machine learning algorithms in AECOPD prediction from the narrative synthesis

Type of algorithm	Description
Linear Discriminant Analysis (LDA) [53, 71]	LDA, is a dimensionality reduction and classification technique. It projects high-dimensional data onto a lower-dimensional space while preserving the discriminative information between classes. LDA minimises the within-class scatter (variance) and maximises the between-class scatter, leading to a lower-dimensional representation where data points from different classes are well-separated and tightly clustered within their respective classes.
K-Nearest Neighbour (KNN) [63]	KNN is a non-parametric machine learning algorithm used for both classification and regression tasks. It makes predictions based on the similarity of input samples to the training data. KNN finds the K nearest neighbours in the training dataset to a given input sample. Nearest is determined by a distance metric, commonly Euclidean distance. The K nearest neighbours are identified by calculating the distances between the input sample and all training samples and selecting the K samples with the shortest distances.
Random Forest (RF) [55, 65, 76, 77]	RF is an ensemble learning method used in machine learning for both classification and regression tasks. It combines the predictions of multiple decision trees to produce more accurate and robust results. RF builds a collection of decision trees, each tree is trained on a random subset of the training data and considers only a subset of the available features. By introducing randomness and diversity into the model, RF reduces overfitting and can provide more reliable predictions.
Support Vector Machine (SVM) [55, 60, 73]	SVM finds an optimal hyperplane that best separates different classes in the feature space. The hyperplane is chosen such that it maximises the margin, which is the distance between the hyperplane and the nearest data points from each class, known as support vectors. SVM maximisation of the margin makes it more robust to noise and outliers. In cases where the data is not linearly separable, SVMs employ a kernel function to map the input data into a higher-dimensional feature space, where it becomes separable.

Logistic Regression (LR) [64, 66]	LR is a supervised learning algorithm that predicts the probability of an instance belonging to a particular class based on input features. Using the logistic function, LR models the relationship between the predictors and the binary response variable. The logistic function, also known as the sigmoid function, maps the linear combination of predictor variables onto a probability scale between 0 and 1. This allows LR to estimate the probability of the binary outcome, providing insights into the likelihood of an event occurring. The model parameters are estimated through maximum likelihood estimation, where the objective is to maximise the likelihood of observing the given data.
Classification and Regression Tree (CART) [61, 74]	CART is a tree-like model used for classification and regression tasks. It starts with a root node and recursively splits the data based on features to create internal nodes. Each internal node represents a decision or condition, partitioning the data into subsets. The process continues until it reaches leaf nodes, which provide final predictions. The construction involves selecting the most informative feature at each node using measures like Gini impurity or information gain. Gini impurity measures misclassification probability, while information gain quantifies uncertainty reduction.
Bayesian Network [54, 69, 70]	A Bayesian network is a graphical model representing probabilistic relationships between variables. It uses a directed acyclic graph where nodes represent variables and edges indicate dependencies. The model incorporates prior knowledge as prior probabilities and combines them with observed data using Bayesian inference to estimate posterior probabilities.
Probabilistic Neural Network (PNN) [56, 59]	PNN uses a probabilistic approach for decision-making and provides probabilistic outputs instead of discrete class labels. It consists of layers: input, pattern, summation, and output. The pattern layer stores training patterns, while the summation layer measures similarity using a kernel function like the Gaussian kernel. The output layer generates a probability distribution for classes. Each output neuron represents a class, with its value indicating the probability of the input pattern belonging to that class. The probabilities are calculated based on weighted votes from training patterns, determined by similarity to stored patterns.

Parzen window for novelty detection [58]	Parzen window is a technique in novelty detection that estimates the probability density function of a dataset. It identifies instances differing significantly from the training data by estimating their density in the feature space. Using a window, such as a hypersphere, around each training point, the density is calculated by counting nearby points and weighing them based on distance. For classifying new instances, the density estimate is computed at their location using the same window size. If the density falls below a threshold, the instance is deemed novel. Parzen windows offer a flexible approach but require careful selection of the window size.
K-means clustering [62]	K-means clustering is an unsupervised learning algorithm that groups data into distinct clusters. It aims to minimise the squared distance between each data point and its assigned cluster's centroid. The algorithm begins by randomly selecting K centroids and assigning each point to the nearest centroid. The centroids are then updated by computing the mean of the points in each cluster. This process iterates until convergence, resulting in K clusters where each point belongs to the nearest centroid's cluster.
Long Short-Term Memory (LSTM) [67]	LSTM is an advanced recurrent neural network architecture that overcomes the limitations of capturing long-term dependencies in sequential data. It selectively remembers or forgets information using memory cells, which have input, forget, and output gates. These gates control the flow of information into, out of, and within the memory cell. The input gate decides how much new information to store, the forget gate determines what to discard, and the output gate controls the output to the next step in the sequence. This enables LSTMs to effectively handle long sequences and retain important information.
Multi-Task Neural Network (MTNN) [68]	MTNN performs multiple related tasks simultaneously by sharing information across them. It uses shared layers to learn common representations and task-specific layers for task-specific information. This joint learning captures patterns and relationships among tasks, improving overall performance and generalisation.

C5.0[72]	C5.0 is a machine learning algorithm for classification tasks. It is an extension of the C4.5 algorithm, designed to handle both categorical and continuous input variables. C5.0 constructs a decision tree model based on the provided training data, where each internal node represents a decision based on an input feature, and each leaf node corresponds to a predicted class or outcome. The construction of the decision tree in C5.0 involves a recursive partitioning process. At each step, the algorithm selects the best feature to split the data based on measures such as information gain or gain ratio.
Deep Neural Network (DNN) [77]	DNN is a type of artificial neural network that consists of multiple layers of interconnected nodes, or neurons. These networks are designed to learn and represent complex patterns and relationships in data by hierarchically processing information through these layers. The depth of a neural network refers to the number of hidden layers it contains. Each layer performs a transformation on the input data, gradually learning more abstract and higher-level representations as information flows through the network. The initial layers capture low-level features while the deeper layers learn more complex patterns and concepts.

2.3.1 Remote patient monitoring for AECOPD intervention

The current best clinical practice is for exacerbations to be self-managed by a COPD action plan which includes initiating a medication rescue pack made up of antibiotics and/or oral corticosteroids [78]. Patients may seek further medical intervention if their exacerbation is unmanageable at home by contacting their healthcare professional or calling urgent/emergency medical services. Recent progress and development of RPM services have enabled significant progress with regard to supporting people with AECOPD.

Dinesen *et al.* used remotely collected data including blood pressure, pulse rate, weight, oxygen saturation and lung function which were monitored by healthcare professionals (GP, nurse, or doctor) to advise the patient at the earliest possible point of onset of an AECOPD [28]. This system managed to significantly reduce the mean hospitalisation rate by greater than 50% in the intervention group compared to the control group. Furthermore, Calvo *et al.* showed the RPM group had remotely monitored patient data collected (BP, SpO₂, HR, and PEF) which was received, assessed and followed up by a clinical monitoring centre to detect and intervene at the earliest stages of exacerbations [34]. This resulted in a statistically significant reduction in emergency department visits (20 in the intervention vs 57 in conventional care), hospitalisations

(12 vs 33), duration of hospital stay (105 vs 276 days) and need for non-invasive mechanical ventilation (0 vs 8), and time to the first severe (requiring hospitalisation) exacerbation (141 vs 77 days)). Conversely, Pinnock *et al.* in a researcher-blind, multicentre, RCT known as the telescot trial had contrasting outcomes in comparison [32]. They randomised 128 patients to an RPM system and 128 to routine care. The number of days to admission did not differ significantly between groups and over one year, the mean number of COPD admissions was similar in both groups. Additionally, the intervention had no significant effect on health status (SGRQ) between groups. Crucially, the telescot trial was researcher blinded whereas the trial by Calvo *et al.* was open-label which had the potential to influence results. Furthermore, Pinnock *et al.* highlights the significance of enhanced clinical services and community healthcare resource utilization (HCRU). In the telescot trial, both the control and intervention groups received this enhanced community care, potentially contributing to the observed lack of effect. This becomes particularly noteworthy when comparing the results to studies where enhanced care is selectively provided to the RPM/intervention arm. Overall the RCT studies using RPM for AECOPD intervention tended towards non-significance.

The burden of frequent monitoring may also influence the impact of RPM since approaches used are resource-intensive and burdensome to both patient and clinician limiting widespread uptake. A range of clinical variables are monitored daily, this can include SpO₂, symptom self-reporting, or PA. Patients in these studies need to be supplied with equipment to take these measurements and must undergo training. In addition, the correct use of the data generated requires either interpretation from specialist HCPs that monitor patient data or patient's actions once notified by alerts generated from basic algorithms.

In summary, there is conflicting evidence for the role of RPM in AECOPD in the narrative synthesis of RPM papers (Tables 2.1 and 2.2). When analysing the findings of other reviews in this area, similar conclusions have been drawn. A systematic review and meta-analysis by Jang, Kim and Cho found RPM interventions did not reduce mortality or cost, were unlikely to prevent hospitalisations and did not improve quality of life [17]. A systematic review by Kruse *et al.* identified that 45% of the articles included in the review had overall improved patient outcomes whereas 38% had shown no improvement [18]. There is no real commonality in approach in either improved outcomes or no improvement. This is highlighted by Kruse *et al.* who state “high variability between the articles and the ways they provided telemonitoring services created conflicting results from the literature review”. There is considerable variation in the design of the studies, the type of data collected, participant recruitment, length of study, monitoring method, and outcome measures. This variability makes comparison difficult and likely has affected the varying success seen in the literature.

It is crucial to emphasise that the effectiveness of RPM in the context of AECOPD intervention, specifically in predicting exacerbations before their occurrence, remains uncertain. Unlike machine learning studies, there is frequently a deficiency in analysing the accuracy and timeliness of intervention. Consequently, we cannot confidently assert whether this approach can be used to predict exacerbations or instead, if it is only suitable to identify when a person is experiencing an exacerbation.

2.3.2 Remote patient monitoring and machine learning for AECOPD prediction

Machine learning is a subset of artificial intelligence that focuses on the use of data and algorithms, using mathematical models for learning without direct instruction. Most studies incorporating machine learning for AECOPD prediction focus on evaluating the accuracy and related performance metrics. This is in contrast to previous work that utilises RPM solely. This makes the comparison of the incorporation of machine learning challenging especially with regard to the indication of improving health outcomes for patients. Orchard *et al.* concludes that machine learning approaches “are superior to existing predictive algorithms” in their study comparing the two approaches [68]. There has been increased interest in the possibilities of automated systems offered through data generated via digital platforms and AI. The machine learning approach has the potential to improve upon traditional algorithms by considering all available variables and training models on historical data where all the variables led to either an exacerbation or a stable state. This generates a machine learning model that can be used to predict if a patient will exacerbate in real-time, usually days before an exacerbation will occur, creating a timely alert.

Orchard *et al.* considered machine learning in RPM for exacerbation management in COPD [68]. On a dataset with 363 days of RPM data from 135 patients, two basic symptom-counting algorithms were found to have an AUROC of 0.60 (95% confidence interval (CI) 0.51-0.69) and 0.58 (95% CI 0.50-0.67). Importantly, the authors also tested these algorithms in a real-world scenario allowing for missing data, with greater numbers of patient daily data and hospitalisations and found the performance of all the algorithms fell. The best-performing machine learning models resulted in an aggregated AUROC of 0.74 (95% CI 0.67-0.80) on the same data. Orchard *et al.* emphasised that traditional algorithms have very low accuracy and frequent false positives, with one of the most frequently used traditional algorithms performing no better than chance [68]. This paper highlights the potential machine learning has to enhance the field of RPM for COPD exacerbations by improving the accuracy of predictions.

A recent study attempting to predict AECOPD with machine learning uses the digital app

COPDPredict™ which uses patient PRO, FEV₁, and CRP and indicates a sensitivity of 97.9% and a specificity of 84.0% [74]. CRP was measured by the researchers using finger-prick blood samples and point-of-care testing with Eurolyser Diagnostica®. A key aspect of the exacerbation prediction model lies in its distinctive 2-week learning phase during the initial fortnight. In this period, daily Wellbeing Scores, FEV1 measurements every 3rd day, and blood CRP levels on days 1 and 14 contribute to baseline data entries. This distinctive learning phase suggests that COPDPredict™ relies on personalised models for exacerbation prediction, which can result in more accurate forecasts. However, the unique nature of personalised models introduces potential regulatory challenges. Interpretation complexities and the difficulty in obtaining essential validation may hinder the demonstration of the model's effectiveness and safety.

62/80 exacerbating patients managed their AECOPD at home using a provided rescue pack (oral prednisolone and antibiotics) as directed by the clinicians and an action plan. 14 participants did not improve on standard treatment and escalated home treatment to use a nebuliser (salbutamol and/or ipratropium bromide). Risk notification alerts are displayed against the patients on the COPDPredict™ decision support dashboard. The 98% patient compliance with completing the daily PRO suggests a low burden or the effectiveness of timed automatic prompt notifications via the App. The researchers found that the total number of hospitalisations for all patients was reduced by 98% during the 6-month use of COPDPredict™ compared to the previous 6-month period without COPDPredict™. However, due to the lower level of specificity and the prevailing occurrence of negative cases, with 2860 instances of stable health and 291 occurrences of AECOPD, the 90 study participants generated a total of 458 false positives throughout the six-month period. This high false-positive rate may be a barrier to the implementation of the intervention. Moreover, during the study blood CRP levels were measured by the research team to confirm the diagnosis of AECOPD. Clinical blood testing in the real world could be unfeasible or result in potential delays in intervention. This could be addressed by at-home blood diagnostic sensor devices.

Similarly, a paper by Boer *et al.* described the validity of a comparable tool named ACCESS [69]. ACCESS shared a similarly high sensitivity of 97.4% but a lower specificity of 65.6%. ACCESS was later tested in an RCT to identify if its early predictions can prevent AECOPD [70]. However, the researchers found no difference in the number of weeks without exacerbations, the number of symptom-based exacerbations, or exacerbation-related hospital admissions between the intervention and control groups. The lack of statistically significant findings may be due to Boer *et al.* not predicting exacerbations sufficiently far in advance. Patel *et al.* state that COPDPredict™ identified exacerbations of COPD at 7, 3 days (median, interquartile range (IQR)) before clinician-defined episodes, resulting in the sending of alerts to patients and clinicians. Boer *et al.* state they aimed to identify exacerbations early, however, there is no

evidence to suggest that ACCESS did predict exacerbations early. The ACCESS app needs a major symptom present for two consecutive days or two days of increasing dyspnea which could further delay necessary timely intervention. Boer *et al.* list several limitations that may have negatively affected ACCESS, including the small sample size, improvement in exacerbation management education of the control group, and the possible introduction of selection bias. It is also worth noting that ACCESS has a positive predictive value (PPV) of just 13.4% so a very small proportion of patients flagged as AECOPD are actually experiencing AECOPD. In the validation of ACCESS, this meant that only 112 out of 837 times ACCESS suggested a participant contact a healthcare professional was the participant actually having an exacerbation. This could have affected patient response or confidence in the system. Interestingly, there were no differences between the intervention and control groups in timely action (contacting a health care professional, starting a course of antibiotics and/or cortical steroids, or increasing bronchodilator use). This further supports the notion that the lack of effect may be due to the patient's loss of trust in the app which highlights the necessity to have models with strong predictive capabilities. Future interventional studies incorporating machine learning for exacerbation prediction must ensure that the predictive horizon is long enough, and the accuracy of the algorithm is sufficiently high for both sensitivity and specificity.

2.3.3 Machine learning approaches

5 papers were found to use neural networks for AECOPD prediction [56, 59, 67, 68, 77]. Neural networks excel at classification and prediction tasks due to their ability to model complex nonlinear relationships and automatically learn relevant features from data. Furthermore, neural networks can utilise transfer learning for efficient knowledge reuse, exhibit robustness to noisy data, and generalise well to new, unseen examples.

Nunavath *et al.* use a recurrent neural network (RNN) for exacerbation prediction [67]. RNN is a type of artificial neural network. These deep learning algorithms are frequently used in ordinal or temporal problems. Deep learning typically requires a large amount of data to achieve good performance. The data used in this study consisted of 96 patients and around 7300 records collected over two years. This is a relatively small sample and so the researchers built and applied a data augmentation technique to boost the performance of their deep-learning model. The data-augmentation technique artificially increases the size of their training set. This approach generated a large enough data set to apply RNN which may have resulted in a more powerful, robust model that can make more accurate predictions than a machine learning algorithm. Data augmentation can also enhance inter-patient generalisation which has value not only in the application of deep learning algorithms but all approaches to AECOPD prediction. The researchers in this paper applied Long short-term memory (LSTM), to their data (Daily

measurements of PRO and SpO₂). Interestingly, LSTM performs significantly better than SVM, with an accuracy of 84.1% compared to 77.1%, respectively. LSTM may be a better model as it can more efficiently remember or forget data than SVM. LSTM and SVM are deemed somewhat dated in certain machine learning applications. Modern transformer models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have gained widespread popularity, outperforming their predecessors. Renowned for their capacity to capture long-range dependencies, transformers excel in handling sequential and contextual information. Consequently, their integration into future approaches is evident.

The use of Bayesian network algorithms for predicting COPD exacerbations can be seen in the work by Heijden *et al.* [54]. Bayesian network models are probabilistic graphical models that represent and reason about uncertain relationships between variables. They use Bayesian inference to capture dependencies among variables, allowing for efficient probabilistic reasoning and decision-making in domains with uncertainty.

These algorithms have several advantages that may apply to exacerbation prediction such as providing a principled framework for incorporating prior knowledge and expert opinions, enabling the integration of domain expertise into the modelling process. Bayesian networks can also handle missing or incomplete data by leveraging probabilistic inference, allowing for robust predictions even in the presence of uncertainty which may be useful when it comes to the sparsity of patient-entered data. Bayesian networks provide interpretability by explicitly representing the dependencies among variables, which could help with clinician understanding and acceptability when explaining the predictions made by the model. Lastly, Bayesian networks can efficiently update predictions as new evidence is obtained, making them suitable for a dynamic environment such as RPM where patient data frequently changes over time. Heijden *et al.* note that the simplicity and interpretability of Bayesian models enable clinicians to trust in the model's decision-making and highlight the use of bootstrapping methods to improve model performance. Bootstrapping is a resampling technique that estimates statistics by sampling a dataset with replacement which reduces overfitting and improves the stability of the algorithm. Overfitting occurs when a model becomes overly complex and starts to learn the specifics of the training data instead of learning the underlying patterns and relationships. Reducing overfitting increases the likelihood that the learned model will generalise well to unseen data.

Fernandez-Granero, Sanchez-Morillo, and Leon-Jimenez conducted two studies in AECOPD in 2015 and 2017 [60, 65]. The two analyses use different machine learning methodologies on the same data enabling an interesting comparison. The 2017 analysis used an RF classifier which outperformed the SVM model from 2015, the models achieving a PPV of 94.1% vs

84.66%, respectively. An RF is an ensemble learning method using multiple DT/CART models. The RF overcomes the issues of DT/CART models by creating an ensemble that is a more stable model and is less prone to overfitting the data, resulting in a better performance than any single contributing model. SVM works very differently, it aims to identify an optimal hyperplane that maximally separates different classes in the feature space. Kernel functions are employed to handle linearly non-separable data and map it to a higher-dimensional space. A key advantage of SVM is its versatility; kernel functions can be specified for the decision function. The difference in model performance in these papers is not solely due to the algorithms. Notably, the 2017 analysis used the Markov chain Monte Carlo (MCMC) method to impute missing data. This would have improved data quality, especially if the dataset was sparse, perhaps improving model performance. Furthermore, this analysis also incorporated the feature subset selection (FSS) method into their approach whereas the 2015 analysis used principal component analysis (PCA). FSS and PCA are both used for feature reduction, FSS chooses the best features from a set determined by the developer while PCA projects the data linearly to a lower-dimensional subspace. The FSS method may have been more effective at retaining information. SVM performs well with high-dimensionality data bringing into question the necessity of PCA in the pipeline.

Two separate research groups have used LR to predict AECOPD [64, 66]. Shah *et al.* developed classifiers that can predict COPD exacerbation episodes with 60%-80% sensitivity which resulted in 68%-36% specificity. These are wide ranges, the large interval is likely due to the data quality causing a large trade-off in sensitivity and specificity. Comparatively, Kronborg *et al.* produced a model which achieved an AUROC of 0.74. Both incorporated similar pipelines and similar measures. Importantly, these are some of the lowest performances for predictive modelling of AECOPD which indicates the lack of suitability of LR for this problem. LR assumes a linear relationship between the independent variables and the log odds of the dependent variable. This linearity assumption may limit its ability to capture complex non-linear relationships in the data. In cases where the relationship is highly non-linear or involves interactions between variables, more flexible models like decision trees or neural networks are preferable.

Researchers need to consider the strengths and limitations of all the different approaches to this classification task. There will never be a one-size fits all approach to classifying patients as having an exacerbation or being stable. However, incorporating ensemble classification methods may improve the robustness and performance of models. Conversely, Naive Bayes models may improve interpretability for clinician understanding. Also, employing techniques to enhance the available data such as bootstrapping or data augmentation should be considered. When data augmentation is used, the application of deep learning may provide better performance than that of other machine learning algorithms. Whilst it is necessary to consider the machine

learning approach, how these studies are designed and how researchers choose to label the data is critical, as it is this data that the models are trained on. Importantly, there is significant variation in study design, exacerbation labelling and types of data collection in the literature that exists both in RPM and machine learning papers.

2.3.4 Study design

The studies identified in both searches are heterogeneous. Study duration, sample size, and outcomes all vary to some degree and all have a significant influence on the outcomes. In this section, we review how variations in these aspects can affect both RPM interventions and the development and testing of machine learning models.

The duration of published studies varies from 3 months to 1 year, which is of importance as seasonality can affect the likelihood of AECOPD [79–81]. This could be problematic, especially for interventions that utilise machine learning because such interventions may not be generalisable to other months of the year. Furthermore, parameters are set for model training based on how exacerbations are labelled which may further limit generalisation. In a review by Seemungal, Hurst, and Wedzicha they identified the annual exacerbation rate per person to be 0.5 up to 3.5, stating that 3 exacerbations per year may represent an upper limit to the mean rate in all populations [82]. Limiting the study periods limits the number of AECOPD captured reducing generalisability, and impacting the efficacy of an intervention deployed in the real world.

Small studies of only 10-30 participants have often been used [25, 29, 30, 34, 37, 54–56, 58–62, 65, 66, 71, 83, 84]. This may not capture many exacerbations so models will produce results that are unlikely to be generalisable to a larger population. This problem can be mitigated in part by recruiting frequent exacerbators which allow for more exacerbations to be captured. However, there could be more false positives when a model is trained on a dataset with highly exacerbation-sensitive individuals for use in a more general population of COPD patients that exacerbate less frequently. In addition, frequent exacerbators and patients who have never had an exacerbation may have different signals or predictors before AECOPD. The lack of this distinction in the literature makes it difficult to effectively compare studies and select the best possible criteria for future research or intervention. Going forward, researchers could aim to target interventions by looking at the performance of the interventions for both low-risk and high-risk AECOPD patients. There might be value in the development of a large standard dataset that could be used for benchmarking studies.

When considering model generalisability it is important to consider the diversity of study

2.3. RESULTS

participants. Smaller sample sizes or undiversified participant recruitment will reduce the variety of study participants in terms of age, sex, ethnicity, socioeconomic status, disease severity, frequency of exacerbations, and time since diagnosis. Including this diversity in RPM, studies are needed to validate that the intervention can be effectively deployed to any individual. Regarding machine learning studies, models need to be both trained and tested on a diversity of samples to ensure the generalisability of the final prediction model when used in a real-world setting.

Many studies use machine learning in combination with data available from an interventional study or RCT to generate predictive models. The main metric for the success of these studies is the accuracy/predictive capability of the models the study generated. Whilst this may be a necessary first step to identify the potential of exacerbation prediction models, machine learning from the strict conditions of an RCT may result in less applicability to a real-world setting and perhaps even lower accuracy of the models. RCTs are often conducted under controlled conditions with specific inclusion criteria and well-defined protocols. The data collected in such settings may not fully represent real-world variability and complexity. As a result, a model trained solely on RCT data may not generalise well to diverse populations, different settings, or uncontrolled environments encountered in real-world applications. However, RCT can provide a significant advantage despite these issues. A key problem in COPD is the lack of a gold standard diagnostic test for AECOPD. Diagnosis is based on self-recognition and self-treatment usually in conjunction with clinician diagnosis. Data from RCT often incorporate more rigorous testing for validation of the diagnosis of AECOPD creating a more objective label for exacerbation events improving certainty that you are modelling true AECOPD.

There is a lack of consensus on how AECOPD is defined which has resulted in significant variability in how AECOPD is predicted. Some research groups opt to use an “event-based” definition in which an exacerbation is recorded upon a visit to primary care or hospitalisation. Other research groups use a “symptom-based” definition where the patient’s self-reported symptoms were used to determine exacerbation. The latter symptom-based approach ties in more closely to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) recent update to the definition [85]. There are benefits and drawbacks to both approaches. Although an event-based definition may produce a better model performance, they are likely to be less effective in detecting mild or moderate exacerbations. Conversely, symptom-based events can be defined differently by different research groups, reducing the comparability of the models generated. Notably, over 90% of papers surveyed in this review include a symptom-based definition, some of which utilise both event and symptom-based definitions which is likely the optimal decision to cover all potential exacerbation severities. We recommend a consensus for defining symptom-based/event-based exacerbations on which to base a standard for exacerbation

prediction models. Moreover, part of this consensus should include an understanding of the length of the predictive horizon of models which first requires study to determine the optimal length of time to note a prediction and intervene in exacerbations to improve outcomes.

2.3.5 Patient factors

Possibly the most important aspect of RPM is patient engagement and the burden of monitoring. Most of the studies in this review rely on the daily use of multiple sensors and PRO. Where reported, these studies indicate there is very high engagement with monitoring. However, this may be due to the research setting which often provides a controlled environment where participants may receive additional attention, monitoring, and support from researchers. This heightened level of involvement and oversight can positively influence patient engagement and adherence to the intervention. Participants may feel motivated and enthusiastic about being part of a study that is investigating new treatments or interventions. These studies often have rigorous monitoring and accountability measures in place to ensure adherence to the intervention. Patients may be more diligent in their daily use of RPM because they know they are being monitored and assessed. Importantly, many of the RPM RCTs incorporate a warning system that triggers an alert and patient contact if the study participant has not conducted their monitoring for that day. To fully explore the possibility that research setting bias is influencing patient engagement, there need to be data-gathering studies exploring the engagement of an RPM intervention in the real-world setting over the long term to identify if patient engagement is consistent.

There is often no discussion when it comes to the burden of daily monitoring in these interventional and model development studies. Whilst the adherence is high it is unclear if study participants feel burdened by the daily use of multiple sensors. There are many causes for concern surrounding patient burden, especially regarding time and effort. Self-monitoring requires consistent tracking of symptoms, vital signs, medication adherence, and lifestyle factors, which can result in significant time and effort on self-monitoring activities, impacting their daily routines and quality of life. Such monitoring has the potential to lead to increased anxiety, fear, and stress due to the constant vigilance required to manage the condition. Given the high rates of co-morbidities, cognitive impairment, and often very minimal access to the internet and digital technologies capable of connecting to it, some patients may encounter difficulties in operating self-monitoring devices or health apps. This could lead to frustration and potential non-adherence. There is a need to undertake Patient and Public Involvement (PPI) research to better understand the burden to the patient-user and to develop potential strategies or identify sensor types that minimise the burden to increase the likelihood of long-term adoption interventions that use RPM.

2.3.6 A review of indices in remote patient monitoring

The main types of data collected are physiological measures (BP, HR/PR, RR, weight, SpO₂, and temperature), functional measures (lung function, PA), PRO (dyspnea, sputum, sleep quality, depression, anxiety, health-related quality of life), self-report (physiological measures, medication-usage, exacerbation history, demographics, and medical history), meteorological data, and biomarkers (blood CRP).

2.3.6.1 Physiological measures

A systematic review by Buekers *et al.* identified 71 papers that measured SpO₂ in COPD patients before exacerbations but firm conclusions on their predictive ability for exacerbations were not possible due to the lack of available information on implementation and performance [86]. However, Burton, Pinnock and McKinstry showed that compared to baseline, at the onset of exacerbations mean SpO₂ fell from 93.6% to 92.4% [87]. Shah *et al.* in their work identifying and predicting exacerbations of COPD found mean SpO₂ dropped from 94% in the stable period to 93% in the prodromal period [64]. With an increase in the AUROC of their AECOPD predictive models from 0.612 (0.612-0.612) to 0.682 (0.681-0.682) when SpO₂ was added as a feature to PR and RR.

Interestingly, Milkowska-Dymanowska *et al.* identified that there was a statistically significant decrease in SpO₂ preceding an exacerbation. However, they did not find the same for systolic blood pressure (SBP), diastolic blood pressure (DBP), or HR in beats per minute (BPM) [88]. DBP had a significantly lower bootstrap-boost *p* value than SBP, 0.15 and 0.79, respectively, suggesting SBP is more predictive for AECOPD but is less than SpO₂. Interestingly, Shah *et al.* found that mean PR increased by 3 BPM in the prodromal period [64]. However, when comparing the classifier performance using PR vs. SpO₂ as features, they found the AUROC for PR to be 0.578 (0.578 to 0.578, 95% CI) and SpO₂ to be 0.658 (0.657 to 0.658, 95% CI). This research suggests SpO₂ is a superior measure in RPM for predicting exacerbations, however, SBP and HR/PR may offer some predictive capability to enhance exacerbation prediction models.

Xie *et al.*; weight and SpO₂ were the two most selected features from their best models [89]. Mohktar *et al.* showed weight distribution mean and weight standard deviation (after frequent weight measurements) for a single individual had a higher feature in importance in a DT/CART model over several features including temperature percentage change [61]. Dinesen *et al.* used changes in weight as one of their patient indices in their RCT which resulted in a significant reduction in hospitalisation suggesting that it contributed to early AECOPD detection [28].

This information combined with the work by Shin *et al.* which demonstrates lower weight is associated with a greater risk of exacerbation suggests monitoring patient weight may improve prediction models [90].

Heijden, Velicova and Lucas found through a naive Bayes model that there is a dependency between exacerbation and body temperature [91]. They state that fever is a strong indicator of infection which explains this dependency. Lewis *et al.*, Ho *et al.*, and Pedone *et al.* all include body temperature as a measure in their interventions and all had positive outcomes such as fewer primary care interactions, fewer events, and longer time to readmission [31, 38, 83]. Only, Shah *et al.* and Mohktar *et al.* opt to use RR (breaths per minute) for exacerbation prediction with machine learning [61, 64]. This is of importance as Shah *et al.* found mean RR increased by 2 in the prodromal period to an exacerbation. Conversely, a feasibility study by Chau *et al.* used RR as one of three measures monitored by community nurses to enable prompt intervention at the earliest signs of exacerbation and did not result in an improvement in quality of life or a reduction of exacerbation [92]. However, the lack of improvement may be because it was a feasibility study and therefore did not include a power calculation.

Importantly, the work of Burton, Pinnock and Mckinstry found only weak associations between physiological variables and episodes of exacerbation [87]. They also determined that the physiological measurements were unable to differentiate between exacerbations and isolated bad days. Isolated bad days differ from exacerbations in that they entail distinct worsening of symptoms, potentially affecting health metrics such as SpO₂ and FEV₁. However, they lack the two-week decline characteristic of a full exacerbation event, which typically necessitates intervention. Changes in the physiological variables were monitored at the onset of exacerbations rather than the two-week prodromal period. Desaturation and other declines in physiological indices will occur late in an exacerbation event and as seen by this evidence are not predictive but at best diagnostic. Therefore, the collection of other features which may signal the onset of an event early on may be key or as the authors suggest, the use of improved algorithms.

2.3.6.2 Functional measures

FEV₁ is forced expiratory volume exhaled in 1 second, which is captured by a spirometer and is used in the diagnosis and monitoring of obstructive lung diseases. FEV₁ is frequently included in RPM interventions and is used with machine learning to predict AECOPD. In these studies, a digital spirometer is most frequently used at home by the patient which automatically uploads data to a server; rarely a user uploads a FEV₁ reading to a digital diary. There is limited data on lung function change preceding an exacerbation. However, in the Watz *et al.* post hoc analysis of the WISDOM clinical trial observed that mean FEV₁ began to decline two weeks

before the onset of an exacerbation [93]. This evidence underscores the utility of spirometry as a valuable tool for identifying and predicting AECOPD. Patel *et al.* used FEV₁ as a measure alongside CRP and PRO and a machine learning algorithm which showed a sensitivity of 97.9% and a specificity of 84.0% for AECOPD, further demonstrating the predictive capability of spirometry.

PA is rarely used as an input feature. However, utilisation of this measure could provide a burdenless approach to monitoring. Pedone *et al.* use PA as a measure in their RCT which showed a significant reduction in exacerbation events [31]. Chawla *et al.* demonstrated that those with lower PA the 1st week after discharge had an increased likelihood to have 30-day all-cause readmissions than those with higher activity (OR=6.7, $p=0.02$) [94]. Wrist-worn wearables that measure PA are becoming extremely prevalent in society through smartwatches. Unlike spirometry or pulse oximetry, these do not require engagement from the user and can monitor in the background seemingly making such devices a useful, burdenless tool in exacerbation prediction. Nevertheless, they are not an exact alternative to spirometry as they do provide measures directly for lung function.

2.3.6.3 Patient-reported outcomes

PRO is usually collected by patient input to a digital diary. This digital diary contains a subset set of questions that a patient is supposed to answer once a day or one set of questions in the morning and one in the evening. These questions are usually yes or no questions, such as “did you awake during the night due to breathlessness?”. However, questions can also be graded such as “how tight your chest feels today with 0 being not tight at all and 5 being very tight”. The rationale for using PRO for RPM is strong. The CAT is a validated health status tool that uses 8 graded questions to identify the impact COPD has on a person’s daily living [95]. Baseline CAT score has been shown to have a positive relationship with COPD exacerbation [96]. One unit increase in CAT score is associated with an 8% increased risk of an exacerbation [97]. The link between PRO and exacerbation risk isn’t restricted to the CAT. Hurst *et al.* showed an association between the mMRC dyspnea scale and exacerbations within a year [12]. However, many of the RPM studies using PRO do not demonstrate significant differences in intervention and control when trying to prevent hospital admission with COPD exacerbation [27, 32, 35–37, 83, 98]. It is difficult to determine the effectiveness of PRO in these studies due to the use of various other remotely monitored data. However, the questionnaires utilised in some studies differ from validated PRO, often designed by the researchers without evidence to support their effectiveness. With evidence to support the use of CAT scores, future research should encourage using this as the assessment.

2.3.6.4 Biological measures

There is evidence that exacerbations are associated with several biomarkers; most commonly acute-phase proteins and hormokines [99]. The most studied of these biomarkers is CRP [100, 101].

Patel *et al.* was the only study identified in this review to use CRP in the highly sensitive and specific COPDPredict™. The lack of widespread use of CRP in exacerbation prediction modelling relates to the feasibility of a widely deployed system that incorporates daily or even weekly point-of-care testing. This is because patients are unlikely to want to travel for weekly blood testing or to have weekly HCP home visits as it would be a highly resource-intensive project. There are potential systems in place that involve at-home finger prick blood sampling and then mailing blood samples for rapid (24-48h) testing. It is now more feasible that patients would be willing to consider administering and submitting results of point-of-care testing they undertake regularly themselves such as regular lateral flow tests deployed during the SARS-CoV-2 pandemic. Inflammatory cytokines in the sputum of COPD patients are also a potential target for lateral flow tests in monitoring for prediction of AECOPD [102]. Volatile Organic Compounds (VOC) in exhaled breath could also be a potential target if a suitable at-home monitoring device can be developed in the near future [103, 104]. Such devices could potentially differentiate between bacterial or viral origins of exacerbations enabling targeted therapies.

Despite the potential benefits of implementing remote biological monitoring, there are significant challenges stemming from cost and scalability. The practical execution is hindered by the substantial cost required for developing and deploying sophisticated and expensive home-based detection equipment. Ensuring data accuracy and managing the expenses associated with sensor development, sensor deployment, data transmission, and storage is a significant barrier. Widespread deployment is unlikely as achieving the scalability of costly digital biological monitoring technology will need significant investment. While advanced sensors offer the potential for high predictive capability, their development and implementation can be intricate, costly and uncertain. On the other hand, machine learning with simple sensors provides a more accessible and cost-effective solution, potentially reaching a wider audience. A balanced approach might involve integrating the strengths of both, combining simple sensors with sophisticated machine learning for a pragmatic and scalable solution.

2.4 Discussion

We have undertaken a literature review and narrative synthesis of RPM and machine learning for COPD exacerbations to address the problem of identifying better ways to intervene early and or prevent exacerbations to improve outcomes in COPD.

Evidence from this narrative synthesis demonstrates that the majority of the literature has inconclusive results. This is likely to be due to a combination of variations in study populations, study period, and RPM. In addition, there is limited insight into the accuracy or timeliness of AECOPD detection and intervention. Orchard *et al.* identifies limitations in the basic algorithms responsible for generating alerts from RPM, indicating their performance is no better than random chance [68]. This study underscores the substantial potential of machine learning to significantly augment the predictive capabilities of RPM.

Whilst early research for machine learning in the prediction of AECOPD is promising, there is much need for further development in the field. Much of the literature demonstrates the accuracy, sensitivity, specificity, PPV and NPV of machine learning models, however, there is less evidence on how these models translate into improving patient outcomes. Notable exceptions include Patel *et al.* whose COPDPredict™ may reduce hospitalisation [74] but required frequent blood testing and the use of expensive home-based detection equipment. Due to the lack of research in this field, it is too early to determine if exacerbation prediction and intervention can be deployed with machine learning to improve AECOPD outcomes, especially without the incorporation of biological indices. Nonetheless, it seems that the gold standard for exacerbation prediction should be machine learning, as this method is the only means to ascertain whether a prediction is being made as opposed to merely identifying an exacerbation.

Some adjustments can be made in future studies that may lead to more robust conclusions. For example, exacerbation definition and labelling are heterogeneous and would benefit from standardisation, it is critical to ensure that there is sufficient data capture by increasing the number of participants and keeping the study length to a minimum of 1 year should result in better generalisation of predictive models. In addition, utilising data augmentation, resampling, and feature selection techniques can further enhance training data for machine learning. Furthermore, incorporating ensemble techniques or neural networks into a modelling approach may greatly enhance the predictive power of AECOPD models.

In summary, identifying the best form of RPM for AECOPD prediction is not clear-cut. However, there are some insights from the literature that would benefit future studies. Clinically validated PRO like that of CAT, has the potential to greatly enhance predictive models with simple

implementation. Physiological and functional measures that appear to frequently perform well when incorporated with machine learning algorithms include pulse oximetry and spirometry but may be burdensome and resource-intensive to both patient and clinician. PPI research needs to be conducted with these sensors to explore the feasibility and likelihood of long-term adherence. The incorporation of sensing technologies that monitor in the background and are seemingly burdenless (wrist-worn wearables/smartwatches) should be considered. The use of biological measures is a rarely used method of RPM but from work, by Patel *et al.* it is clear that the addition to RPM could result in enough accuracy to make effective interventions.

2.4.1 Limitations

The number of studies we found for our analysis was limited, affecting how broadly we can apply our findings to the larger COPD population. The initial search was constrained because we specifically focused on RCTs, which are considered the most robust method for assessing intervention effects. While there are more feasibility and pilot studies available, they often lack statistical evidence of improved patient outcomes because they are not adequately powered. Similarly, when it comes to machine learning papers, our review included only a small number of them. This is primarily due to the relatively narrow scope of research in this area, as we found only 23 relevant papers during our literature search. The strict inclusion and exclusion criteria we used also contributed to the limited number of studies in our analysis. However, these criteria are essential to maintain objectivity and rigour in our search, ensuring the validity of our review.

Our rigorous criteria in designing our search strategy may have inadvertently overlooked certain RPM modalities. Specifically, we may have omitted studies that directly investigate the potential of specific variables, such as cough monitoring or respiratory rate monitoring, in predicting AECOPD.

Bias is a potential concern that may affect our conclusions in this review. We didn't conduct a bias assessment for either of our searches. In research to develop prediction models, bias is relatively prevalent, as noted by Navarro *et al.* in their systematic review of prediction models using supervised machine learning [105]. They point out potential risks like small study size, handling missing data poorly, and overfitting. However, the main focus of our narrative synthesis was to identify different machine learning approaches for predicting AECOPD, rather than extensively assessing bias. It's important to remember that the accuracies of the models we identified may vary when used by real-world patients. We shouldn't rely too heavily on reported accuracies until they are tested in RCTs to demonstrate their clinical value. In the case of RCTs for RPM in COPD, while they are less susceptible to bias compared to other

study designs, they are not entirely bias-free. Various potential bias sources include issues with randomisation, deviations from intended interventions, missing data, outcome measurement, and result reporting. Additionally, Lewis and Warlow highlight seven potential biases in RCTs, such as poor allocation concealment, baseline imbalances, unblinding, missing data, lack of intention-to-treat analysis, counting death as a positive outcome, and conflicts of interest [106]. It's essential to note these potential biases, as their occurrence and impact in the literature are unclear. Despite not conducting a bias assessment, our conclusion in this chapter still holds that RPM alone may not significantly improve patient outcomes, as the evidence consistently leans toward non-significant results.

2.4.2 Recommendations

1. Introducing machine learning algorithms to predict exacerbations could enable significant improvements in the accuracy of identifying AECOPD at the earliest point, enabling prediction before their occurrence.
2. There is a lack of consistency in digital health interventions for exacerbation prevention which limits interpretation and comparisons and may explain variability in the success. There needs to be standardisation of exacerbation labels, study design and data collection to demonstrate the capability of these systems and allow for further improvements.
3. The clinical utility of machine learning approaches will need to be demonstrated with RCT and real-world studies which are currently lacking.
4. There is a lack of research into patient factors, there is a need to study the adoption of RPM in the long term and to develop an understanding of the burden of daily/weekly RPM and potential solutions to overcome this burden.
5. There is preliminary evidence indicating promising methods for RPM. Nevertheless, the question of which sensor technologies offer the least burden to patients while maximising predictive capability remains unclear, warranting further study.
6. Importantly, these changes and systems will be of little use without engaging patient communities to determine their needs and how such systems could elicit the optimal response and the best outcomes.

2.4.3 Conclusion

COPD is a heterogeneous disease, one size probably won't fit all of the different models and differing approaches may be required for different phenotypes and severities of disease. Ultimately an individualised approach trained on a patient's own data may be the goal.

Chapter 3

Data Preparation and Analytics

There are preparatory steps that are necessary to complete for machine learning and analytic purposes. The first step involves data wrangling, the process of mapping and transforming data from its raw, unstructured form into a more usable format. This process involved capturing every entry available in the prior and following 14 days of each dynamic data point (dynamic data comprises features that undergo frequent changes, such as CAT and symptom scores). Subsequently, through co-creation between clinicians and engineers, we developed labels to capture the initiation of exacerbations and stable health (designed to exclude temporary improvements in health during ongoing exacerbation events). This enabled the labelling of each row of data based on the occurrence of a stable or exacerbation event in 1-8 days. Baseline characteristics such as demographics and disease severity were then attached to this labelled data and features were engineered. A more limited portion of the myCOPD population possessed data conducive to applying machine learning techniques; this population had a greater proportion of high engagement in use with the myCOPD app and in general had more severe disease. Visualisation of features extracted from the myCOPD app suggests the change in CAT score from a mean baseline is likely a very powerful predictor of AECOPD.

3.1 Introduction

This chapter delves into the process of data wrangling, processing, and analysis of a vast dataset acquired from a cohort of COPD patients who actively engaged with a specially designed mobile application. The diverse nature of the data necessitates the application of methodologies to extract meaningful information, uncover hidden patterns, and derive valuable clinical knowledge.

The myCOPD data extract created on December 7 2021, contained 13,682 activated patient users (users who have activated their accounts and logged into the app). This was received as raw data in multiple CSV files for different features and characteristics. The files included: symptom scores, CAT scores, medication diaries, tile clicks, flu vaccination history, COPD GOLD groups, mMRC dyspnea scores, smoking history (pack year), exacerbation history (rescue pack times, hospitalisation times, severity of last exacerbation), and demographics (age and gender). After

reviewing the literature in Chapter 2, this project aimed to identify if the myCOPD data set has any predictive capability for AECOPD. However, the data first needs to be transformed into a usable format where machine learning can be applied. Secondly, features of the dataset need to be engineered and visualised to identify suitability for modelling.

In this Chapter, I show my initial work wrangling, exploring, and visualising the myCOPD data through:

1. Generating pandas DataFrames of the dynamic data where each data point has all available data points from the previous 14 days and the next 14 days.
2. Using patient-entered symptom scores to design exacerbation and stable labels through co-creation between clinicians and engineers.
3. Combine the dynamic features with the baseline characteristics entered by myCOPD patient users.
4. Identify if there is a significant difference in characteristics between the population of users who match the labels and those who don't.
5. Visualise the available features for insight into their potential for AECOPD prediction.

3.2 Methods

3.2.1 The feature window

CAT scores, symptom scores, tile clicks and medication use were extracted in CSV files with the value attached to a DateTime variable. The CAT is a routinely used questionnaire designed to gauge the impact of COPD on a patient's overall well-being and daily functioning. It consists of eight questions that cover various aspects of COPD-related symptoms, such as breathlessness, coughing, and activity limitations. Each question is scored on a scale from 0 to 5, with higher scores indicating a greater impact of COPD on the patient's life. The total score provides healthcare professionals with valuable insights into the patient's condition, helping them tailor treatment and interventions to improve the patient's quality of life. Symptom scores are unique to the myCOPD app and are described in more detail in 3.2.2. These raw CSV files were transformed into a DataFrame using pandas - Python Data Analysis Library. An example of one of these DataFrames can be seen in Table 3.1. The patient key column is the identifier assigned to each patient-user upon activation of their myCOPD account. The date column is the date at which the symptom score was generated. The symptom score column comes from

3.2. METHODS

the number attributed to a patient's self-assessment in myCOPD which is explained in more detail in 3.2.2.

Table 3.1: Symptom score DataFrame formed by reading in symptom score CSV file from myCOPD data extract through pandas.

patient key	date	symptom score
2013	10/09/2018	1
2013	11/09/2018	1
2013	12/09/2018	1
2014	04/08/2018	1
2014	08/08/2018	2

The DataFrames were processed through a feature window which was programmed to retrieve the available entries that were within either 14 days before or 14 days after each data point. A 14-day timeframe was chosen based on clinical guidance from the Bristol Academic Respiratory Unit, suggesting that respiratory symptoms start to decline approximately 14 days before an exacerbation occurs. This structured the data as cross-sectional, with many observations of many subjects at one point in time. This decision was chiefly made due to the sparsity of data. Many of the myCOPD app users log on only once a week and they only receive a prompt to complete the CAT monthly. As a result, there is a lack of continuity that is necessary for data to be organised as a time series for modelling. Importantly, the symptom score variable was combined into each DataFrame, one with each dynamic variable matching the entry date of the symptom scores before generating the feature window. This is because symptom score was used for labelling exacerbations and so had to be added into each DataFrame. Without this step, data is lost when merging the columns if a dynamic variable does not occur on the same day as a symptom score variable. A section of the CAT score DataFrame consisting of the 14 days before each data point can be seen in Table 3.2. The patient key column is the identifier assigned to each patient-user upon activation of their myCOPD account. The date column is the date at which the CAT score was generated by the user logging into the myCOPD app. The CAT score column is the patient user's answers to the CAT on how their COPD symptoms affect their life. The columns after that are named x days before are where the system has searched for entries of CAT or symptoms up to 14 days before the date column. The entries named nan are where there was no data for that day and so are left empty.

Table 3.2: CAT and symptom scores after processing each data point to include any entry 14 days before. nan means there was no entry found for that day.

patient key	date	symptom score	CAT score	symptom 1 days before	CAT 1 days before	symptom 2 days before	CAT 2 days before
36	19/05/2020	1	nan	1	nan	nan	nan
36	29/05/2020	1	nan	1	nan	nan	nan
38	13/12/2019	1	20	1	nan	1	nan
38	15/12/2019	2	23	1	nan	1	20
38	01/10/2020	2	nan	1	nan	1	nan

3.2.2 Label creation

Patient users of the myCOPD app must complete a self-assessment of their symptoms for that day before they can enter the myCOPD app on each new day. This screen is shown in Figure 3.1. The patient has four options: “normal for me” (no change in normal symptoms) represented by a symptom score of 1, “mild deterioration” (have used their reliever inhaler) represented by a symptom score of 2, “moderate deterioration” (worsening of symptoms they have taken antibiotics and/or corticosteroids) represented by a symptom score of 3, and “severe deterioration” (they have contacted primary care provider, called 999, or have been hospitalised) represented by a symptom score of 4.

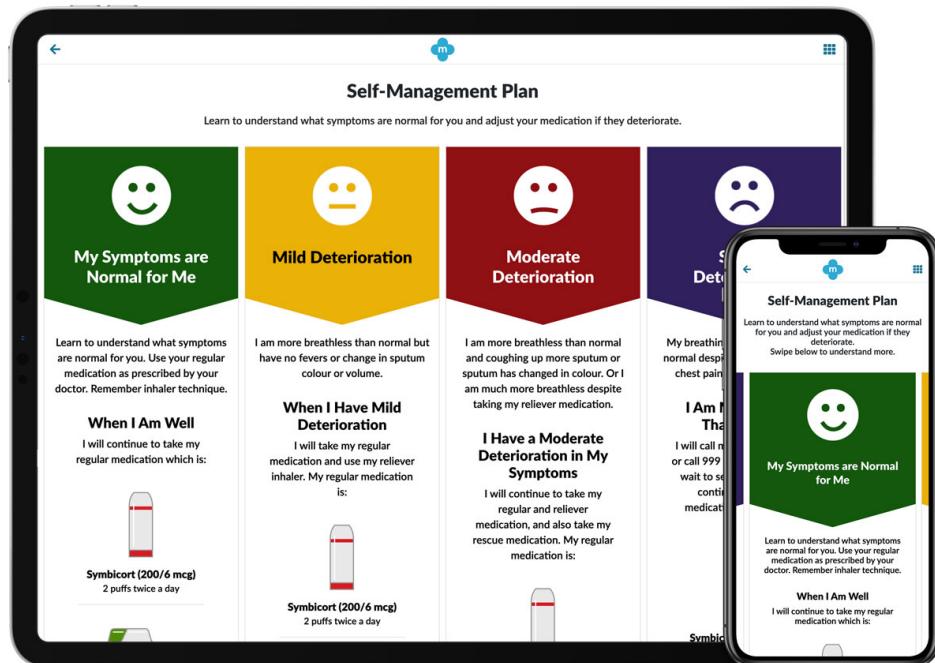


Figure 3.1: The opening screen of the myCOPD app when a patient logs on for a new day. The patient user will select one of four symptom scoring options.

3.2. METHODS

Using these symptom scores and the DataFrames for 14 days before and after, through co-creation between engineers and clinicians we developed the labels. Co-creation is a collaborative process where individuals actively cooperate to develop and refine products, services, or solutions. Co-creation fosters mutual engagement, open dialogue, and iterative feedback, yielding outcomes that are more relevant, innovative, and aligned with the audience.

The co-creation meetings used for label development were conducted in iterative steps. The first meeting focused on the definition of exacerbations and the suitability of symptom scores for defining exacerbations within the myCOPD data. In this first meeting, clinicians from my mhealth and the Bristol Academic Respiratory Unit classified symptom scores 1 and 2 as stable health and symptom scores 3 and 4 as an AECOPD. This decision was based on the previous GOLD definition for exacerbation which was defined as “an acute worsening of respiratory symptoms that results in additional therapy” [22]. GOLD has recently adopted a new definition “an event characterised by dyspnea and/or cough and sputum that worsen over 14 days, which may be accompanied by tachypnea and/or tachycardia and is often associated with increased local and systemic inflammation caused by airway infection, pollution, or other insult to the airways” [22]. Our clinical consensus aligns with both of these definitions: when a patient user selects a symptom score of 3 or 4, it signifies a significant worsening of respiratory symptoms and that they will soon be using additional therapy. To validate the appropriateness of these labels, engineers plotted symptom scores for 14 days both before and after a labelled day. This allowed for the visualisation of the patient’s self-reported health status over one month, facilitating discussions in subsequent co-creation meetings to assess the labels’ validity.

In the second meeting, the visualisations were presented to the group and discussion between clinicians and engineers determined that these initial labels captured exacerbations that were part of an ongoing exacerbation, not an initial event. The label of stable health also sometimes captures momentary improvement in an ongoing exacerbation rather than a distinct period of stable health. Through dialogue, it was reasoned that creating a window of time free from a symptom score of 3 or 4 would create the separation necessary to isolate exacerbations and stable health. So, the label was further ideated to include no reports of symptom scores of 3 or 4 entered in the last 7 days. Visualisations of these events were produced showing the symptom scores 14 days before and 14 days after a label.

In the third co-creation meeting these visualisations were discussed and it was found that the labels were still not correctly isolating the starting point of exacerbations and periods of stable health. It was then gathered that the exacerbation-free window (symptom score of 1 and 2) should be extended to 14 days before a label to create greater separation. It was also put forward by the clinicians that part of the label should include that an entry had to occur 1-8

days prior. This was selected as the prediction horizon as the clinicians determined prediction at this time could enable early meaningful intervention and would likely be early enough to detect the beginnings of change in health status prodromal to an AECOPD. Moreover, engineers indicated that this prediction horizon would likely have enough signal to enable AECOPD prediction.

In the fourth and final co-creation meeting these visualisations were shared and interpreted. It was confirmed by both clinical and engineering researchers that the visualisations showed suitable labels for exacerbations and stable health. The visualisations were found to only capture the initial point of onset for an exacerbation and stable health points were only captured in periods of ongoing stable health rather than momentary improvements in health during an exacerbation.

3.2.3 DataFrame transformation

To transform the data into a usable format for demographic analysis, feature analysis, and ultimately to apply machine learning algorithms, I began by mapping the dynamic features into one dynamic DataFrame. To do this I started by labelling each date in each DataFrame with a column named exacerbation score. The exacerbation score was similar to the symptom score still using 1-4, however, it was created using the stable or exacerbation labels. Patient data that did not match either the exacerbation or stable label were removed from further analysis. Data points that were removed either were part of an ongoing exacerbation or data entry was too sparse to have an entry 1-8 days before a stable or exacerbation-labelled event. The patient users were then split into two groups, the first being labelled patient users ($n=1,758$) which are patient users who had at least one data point that matched the stable label and one data point that matched the exacerbation label. The second group is patients who did not have both labels and were so-called non-labelled patient users ($n=11,273$). A representation of a section of the combined dynamic data is shown in Table 3.3. The patient key column is the identifier assigned to each patient-user upon activation of their myCOPD account. The date column is the date on which the label was generated. The exacerbation score column is the symptom score report for a generated label. The CAT score column is the patient user's answers to the CAT 1-8 days before the label. The symptom score column is the patient user's self-assessment 1-8 days before the label. Tile clicks are how many tiles the patient-user entered in the myCOPD app 1-8 days before the label. Regular is how much regular medication was used 1-8 days before the label. Reliever is how much reliever medication was used 1-8 days before the label.

3.2. METHODS

Table 3.3: All of the dynamic DataFrames merged in one DataFrame.

patient key	date	exacerbation score	CAT score	symptom score	tile clicks	regular	reliever
59050	21/10/2021	3	34	2	5	5	8
60053	20/09/2021	3	25	2	15	2	0
87	02/08/2020	1	26	1	2	1	0
124	17/07/2020	1	17	1	1	3	0
124	01/10/2020	2	31	2	1	2	1

After the formation of the dynamic DataFrame, baseline characteristics (demographics, exacerbation history, disease severity) were matched to patient keys to generate the final DataFrame, as seen in Table 3.4. The gender column is the patient user's self-description as either male (1), female (2), or not specified (0). The age column is the age of the patient user. The GOLD group column is a system basis for the stage of COPD (1/A = mild, 2/B = moderate, 3/C = severe 4/D = very severe). The mMRC dyspnea column is based on the mMRC dyspnea scale which is disability due to breathlessness (0 = only gets breathless with strenuous exercise, 1 = gets short of breath when hurrying on level ground or walking up a slight hill, 2 = on level ground, walks slower than people of their age because of breathlessness, or they have to stop for breath when walking at their own pace on the level, stop for breath after walking about 100 yards or after a few minutes on level ground, 3 = stops for breath after walking about 100 yards or after a few minutes on level ground, and 4 = too breathless to leave the house or breathless when dressing/undressing). The smoking status column measures if the patient-user is a current smoker (2), ex-smoker (1), or non-smoker (0). The pack-year column is the patient user's number of packs of cigarettes smoked per day times the number of years the person has smoked. The antibiotics and steroids last year column is the number of rescue packs used for an exacerbation in the last year. The hospital last year column is the number of times the patient-user was hospitalised for an exacerbation in the last year. The type column is the severity of the previous exacerbation 2 = moderate, 1 = mild, 0 = did not have one in the last year.

Table 3.4: The baseline characteristics merged onto the end of the dynamic DataFrame.

reliever	gender	age	GOLD group	mMRC dyspnea	smoking status	pack year	antibiotics and steroids last year	hospital last year	type
8	2	72	3	1	1	30	2	1	1
0	2	65	2	2	1	36	1	0	1
0	0	69	4	1	1	20	1	0	1
0	1	61	1	1	1	26	0	0	0
1	1	61	1	1	1	26	0	0	0

All features included in the final DataFrame to be used for analysis, visualisation and machine learning include CAT score, symptom score, tile clicks, regular medication use, reliever medication use, gender, age, mMRC dyspnea score, GOLD group, smoking status, smoking years, pack year, antibiotics and steroids used in the last year, hospitalised for an exacerbation in the last year, and the severity of the last exacerbation which can be input to the app by the patient or is linked by medical records.

3.2.3.1 Feature engineering

A subset of features was developed from the initial features. The first feature developed was stable CAT which was created by taking a patient user's mean CAT score when their symptom score entry was 1. The next was a change in the CAT score which is the difference between current CAT score and stable CAT. This process was repeated for medication and tile clicks.

3.2.4 Demographic analysis

Demographic analysis began by counting the number of symptom scores both labelled ($n=1,758$) and non-labelled ($n=11,273$) patient users entered and visualising each symptom score and the number of entries in a bar chart. Gender, GOLD group, mMRC dyspnea score, and smoking status were plotted using a bar chart comparing labelled and non-labelled patient users. Age, pack year, antibiotic and steroid use times, hospitalisation times, and stable CAT were plotted in a line graph.

3.2.5 Feature analysis

For feature analysis the mean of each dynamic variable (CAT score, CAT change, regular medication, regular change, reliever medication, reliever change, tile clicks, tile change) for each day 14 days before either a stable or exacerbation label was plotted on a line graph. Gini impurity and relative importance were used to identify the potential of features for predicting AECOPD. Gini impurity is used to measure the purity of a split in a decision tree. Minimising Gini impurity increases the purity of a split meaning a feature has better separated a future exacerbation event from a future stable event. Relative importance indicates how frequently a feature was selected by a model due to its ability to minimise Gini impurity.

3.3 Results

3.3.1 Label visualisations

The first attempt at creating stable and exacerbation labels resulted in many exacerbation labels being part of an ongoing event rather than at the initiation of the exacerbation. Similarly, many stable labels would appear to be a momentary improvement within an ongoing exacerbation rather than a period of stable health. The visualisations in Figures 3.2 and 3.3 were produced from the initial label (7 days no symptom score of 3 or 4). As seen in these visualisations the exacerbation label does not capture the start of an exacerbation event. On day -14, there is a symptom score of 3 which suggests this could be the start of the exacerbation or potentially even earlier. In the stable label, the patient-user inputs a symptom score of 3 at -13 days before the stable. The patient-user inputs two more symptom scores of 3 before the stable label and 4 more within 14 days, one immediately after the stable label. This suggests the label is capturing momentary symptom alleviation in an ongoing exacerbation rather than a period of stable health. Moreover, the incorrect labels go against the newest definition of AECOPD in which symptoms worsen over 14 days before AECOPD [22].

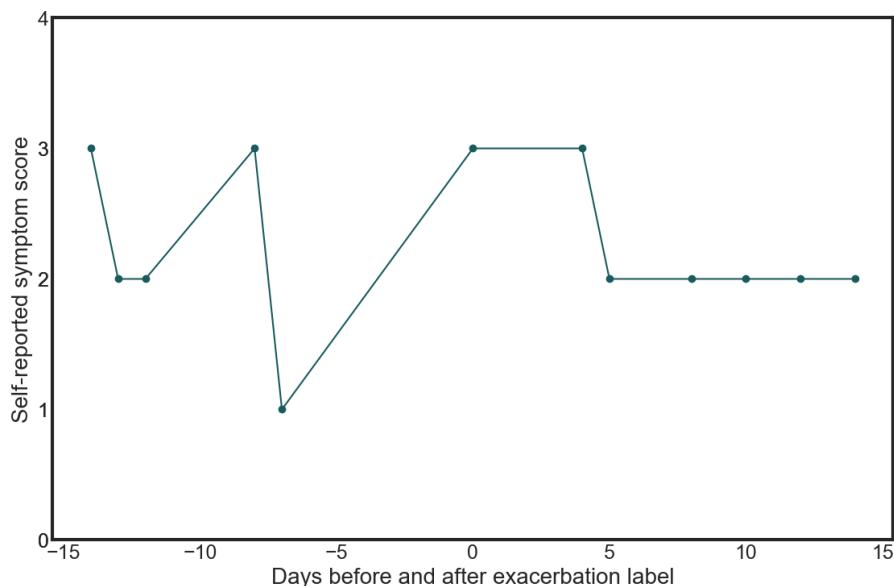


Figure 3.2: Ongoing exacerbation within a single patient-user's exacerbation label. The exacerbation label shows the patient-user reported a symptom score of 3 at -14 days, -8 days, and 4 days in relation to the exacerbation label on day 0. Because the patient-user reported a symptom score of 3 at -14 days and the exacerbation label is between two other symptom scores of 3 the exacerbation label has captured an ongoing exacerbation rather than the initial point the exacerbation started.

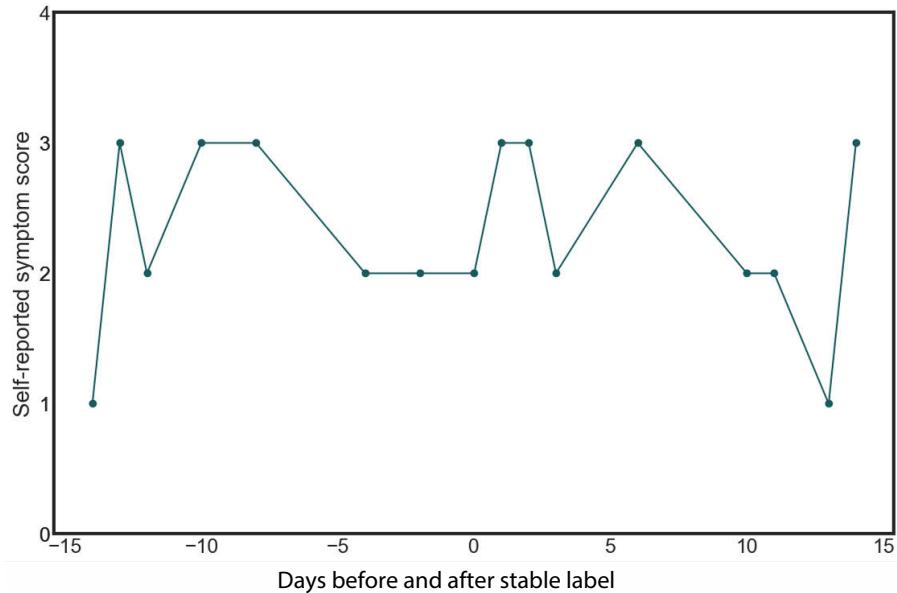


Figure 3.3: Ongoing exacerbation within a single patient-user's stable label. The label shows the patient-user has reported multiple symptom scores of 3 around the stable label at day 0. There are 3 reports of symptom scores 3 before the label and 4 after. This is all within a one-month window. The stable label has captured a period of better health but during an ongoing exacerbation rather than a period of stable health.

Symptom stability over 7 days did not enable appropriate labelling, both this and the new definition of AECOPD strengthen the validity of the decision of the final successful exacerbation label where there is no symptom score of 3 or 4, 14 days before a label. Visualisations of the final label demonstrating its ability to correctly capture stable health and exacerbation events are shown in Figures 3.4 and 3.5

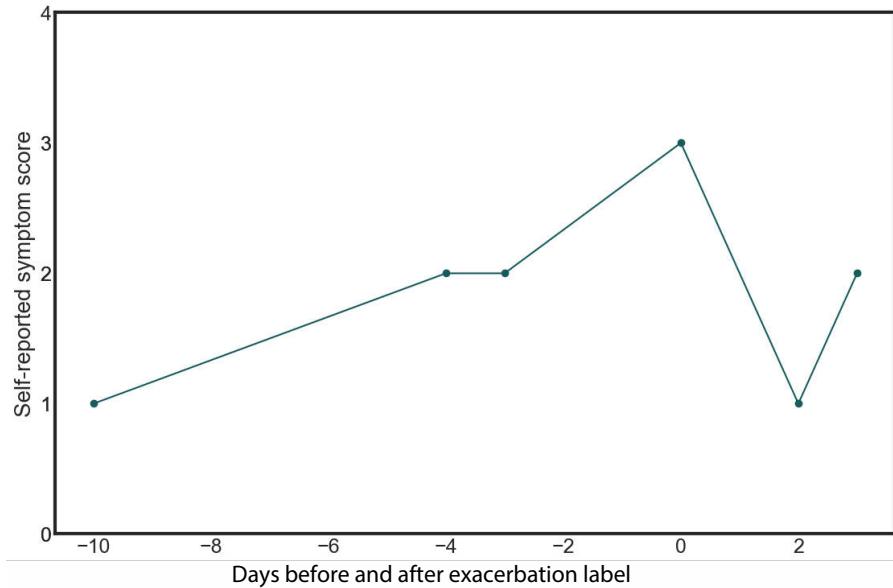


Figure 3.4: Successful identification of an exacerbation label in a single patient-user. The label shows the patient-user reported a symptom score of 1 at -10 days then a symptom score of 2 at -4 and -3 before reporting a symptom score of 3 at day 0 the exacerbation label. The steady worsening of symptoms over time suggests this label correctly captures the start of an exacerbation.

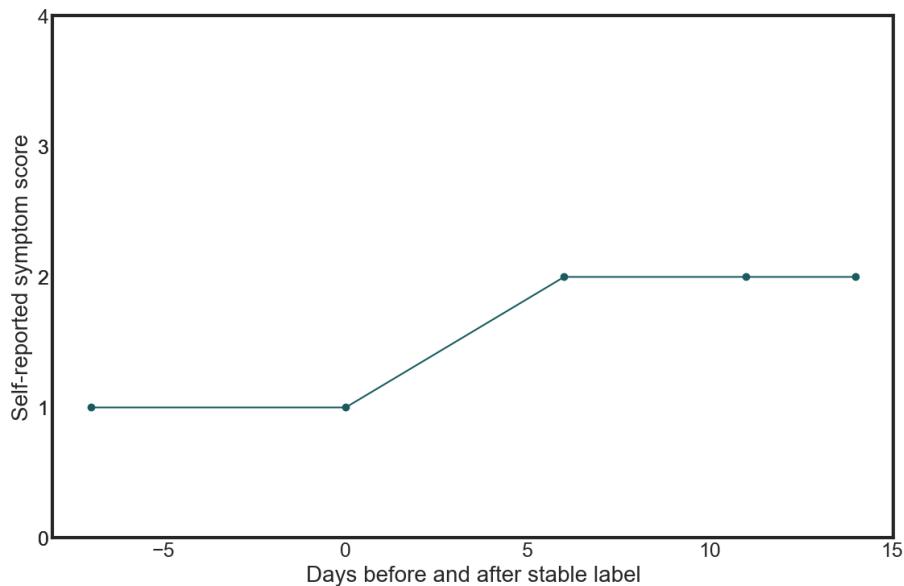


Figure 3.5: Successful identification of a stable label in a single patient-user. Day 0 is less than 3 and the surrounding symptom scores are less than 3 suggesting this is a period of stable health and accordingly, day 0 is a stable label.

3.3.2 Engagement analysis

Table 3.5 shows that on average, participants in the labelled group have used the myCOPD app more, a lower percentage of people have used the app less than 5 times and they have been using the app for far longer. This information suggests that participants in the labelled group have a greater proportion of people who use the app regularly and for longer and have greater user retention.

Table 3.5: myCOPD app usage comparing the labelled group to the non-labelled group.

Patient users	The number of separate days accessing the myCOPD app, mean (SD)	Percentage of users with less than 5 separate days accessing the myCOPD app	The length of time using the myCOPD app in days, mean (SD)
labelled (n=1,758)	107.5 (180.8)	17%	307 (318)
non-labelled (n=11,273)	13.9 (53.5)	65%	107 (193)

Cooper *et al.* use symptom scoring frequency (app entries) to classify myCOPD patient-users into engagement groups: low is <1 time every 100 days, moderate is 1-5 times every 100 days, high is 6-20 times every 100 days, and very high is >20 times every 100 days [107]. Table 3.6 shows how many users in each group (activated, non-labelled and labelled) use the app >20 times every 100 days (very high engagement). This table shows that more patient users in the labelled group are classified as “very high engagement”. There is a large number of users (n=448) who have very high engagement in the non-labelled group, however, proportionally the corresponding fraction is much smaller than the labelled group.

Table 3.6: myCOPD comparing engagement using the Cooper *et al.* definition of very high engagement (>20 times every 100 days) between all users who activated the mycopd app, the labelled user group and non-labelled users.

Patient users	Total users	The number of users with very high engagement (% of total)
Activated	13,031	1,134 (8.70%)
Non-labelled	11,273	448 (3.97%)
Labelled	1,758	686 (39.0%)

3.3.3 Demographic visualisations

Despite labelled patient users being a significantly smaller group than non-labelled (n=1,758 vs. n=11,273), labelled patient users have more symptom score entries as seen in Figure 3.6. As symptom scores must be input into the app before use on each new day, this would suggest that labelled patient users are using the app much more than non-labelled users. Moreover,

3.3. RESULTS

due to the greater frequency of symptom score 2, there is an indication that labelled patient users are recording declining health much more frequently. This would make sense as in the labelled patient group the user must have at least one entry 1-8 days before an exacerbation label. During this time patient users are more likely to experience worsening symptoms and use a reliever inhaler, which they would report when logging in and selecting “mild deterioration”.

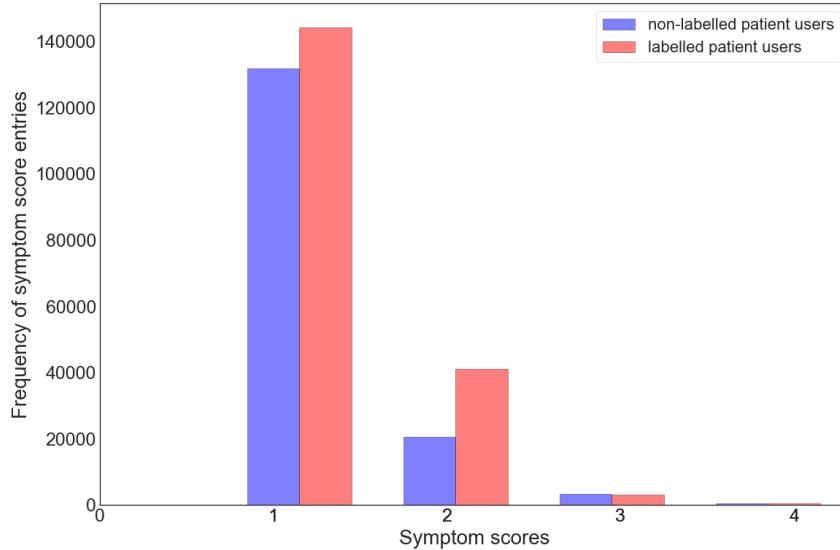


Figure 3.6: Comparison of symptom score entry frequency from labelled patient users against non-labelled patient users, using all available patient data.

Between the groups, there appears to be no discernible difference in the distribution of age and gender (Figures 3.7 and 3.8). The Kolmogorov-Smirnov (KS) test statistic (0.0153) suggests a small discrepancy in age distributions between the two datasets, while the p -value (0.862) exceeds the typical 0.05 significance level. Consequently, the KS test results indicate no significant difference in age distributions. A T-test revealed there is some evidence to suggest a difference between males and females in the two groups. Still, it is not strong enough to be considered statistically significant at the conventional 0.05 significance level ($P=0.0532$).

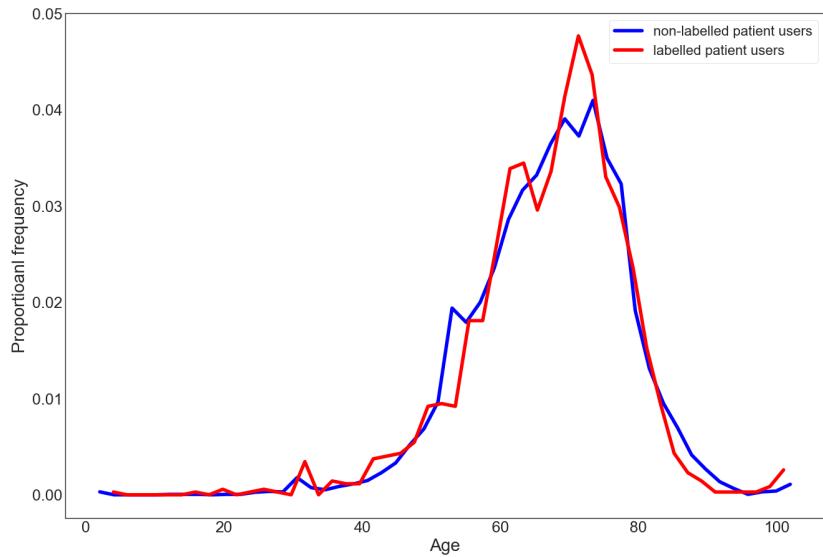


Figure 3.7: The proportional age distribution comparison between labelled and non-labelled patient users, using all available patient data

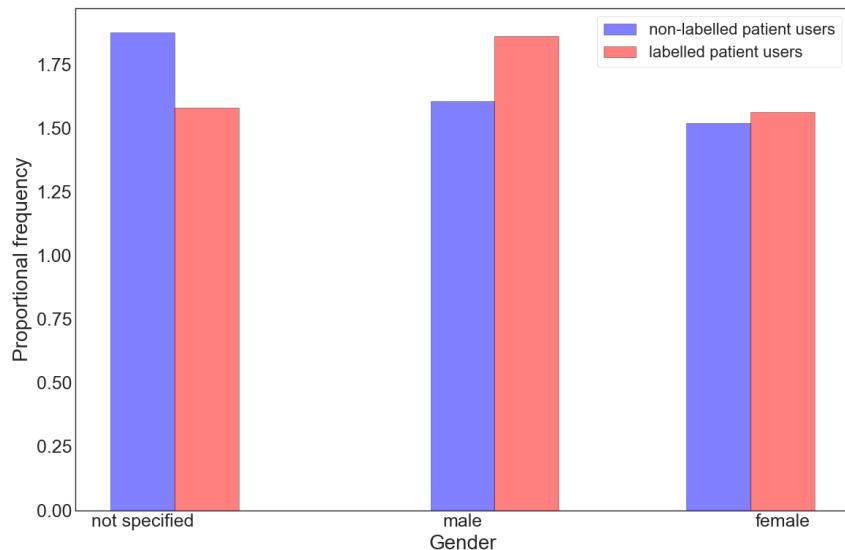


Figure 3.8: The proportional gender distribution comparison between labelled and non-labelled patient users, using all available patient data.

Interestingly, there are some key differences in the distribution of GOLD groups. This work predates the update involving the GOLD classification (A, B, and E, with C and D combined into E), and thus aligns with the original A, B, C, and D categorisation. There is a significantly greater proportion of the labelled patient users that are in GOLD group D compared to non-labelled patient users where the vast majority are in A and B. This suggests a greater proportion of labelled patient users have more severe COPD. A chi-squared test identified there is a significant difference in GOLD groups ($\chi^2 = 49.6$, $p = < 0.0001$). The number of

3.3. RESULTS

exacerbations in the previous year is a feature that is used in determining the GOLD group, so, this group likely has more exacerbations. It appears that both the labelled and unlabelled groups have very few members within group C. This could potentially be a coincidental finding. However, it does bring into question the validity of the group C stage. The low number of people within group C could suggest a greater number of people transitioning into group D from A and B or that group C is a rarer phenotype of COPD. Group C requires patients to have poor FEV₁ (<30% of predicted) and high exacerbation risk (>2 in the past year) but fewer symptoms (0-1 mMRC dyspnea scale and <10 CAT score). It seems likely it is rare for individuals to have very poor lung function, a high risk of exacerbation and yet minimal symptoms. With regard to the mMRC dyspnea scale, both groups are skewed to the right with slightly more labelled patient users being higher on the scale indicating a greater proportion of patient users experiencing more severe breathlessness. However, a chi-squared test revealed that the difference is non-significant ($\chi^2 = 5.54, p = 0.237$). Visualisation of GOLD groups and mMRC dyspnea scores can be seen in Figures 3.9 and 3.10.

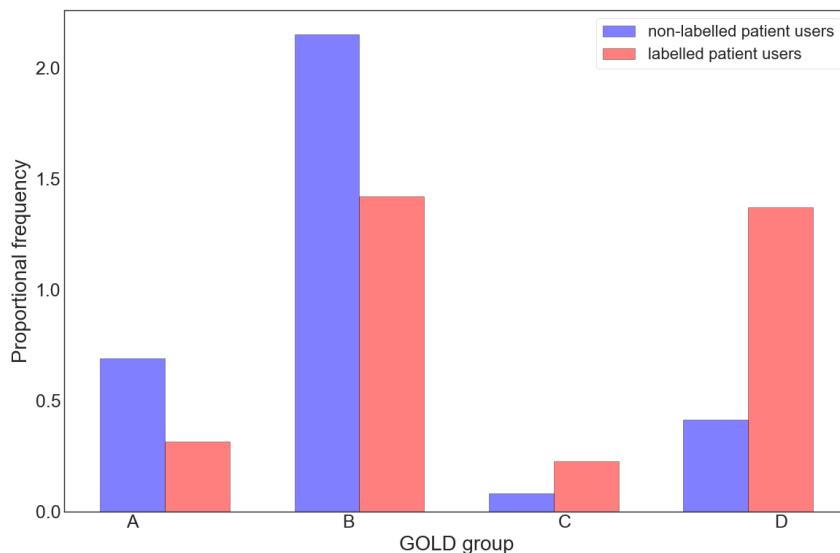


Figure 3.9: The proportional GOLD group distribution comparison between labelled and non-labelled patient users, using all available data patient data.

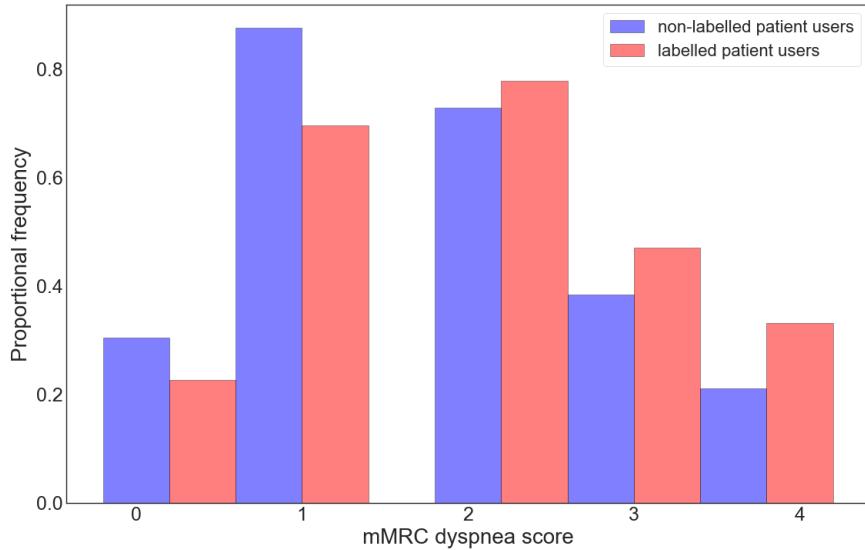


Figure 3.10: the proportional mMRC dyspnea scores distribution comparison between labelled and non-labelled patient users, using all available patient data.

With regard to pack year and antibiotic and steroid use, it appears that labelled patient users generally have a higher pack year and antibiotic and steroid use in comparison to non-labelled patient users. This can be seen in Figures 3.11 and 3.12. KS tests revealed that the difference in distributions was non-significant for pack year but was significant for antibiotic and steroid use, $p = 0.838$ and $p = < 0.0001$, respectively.

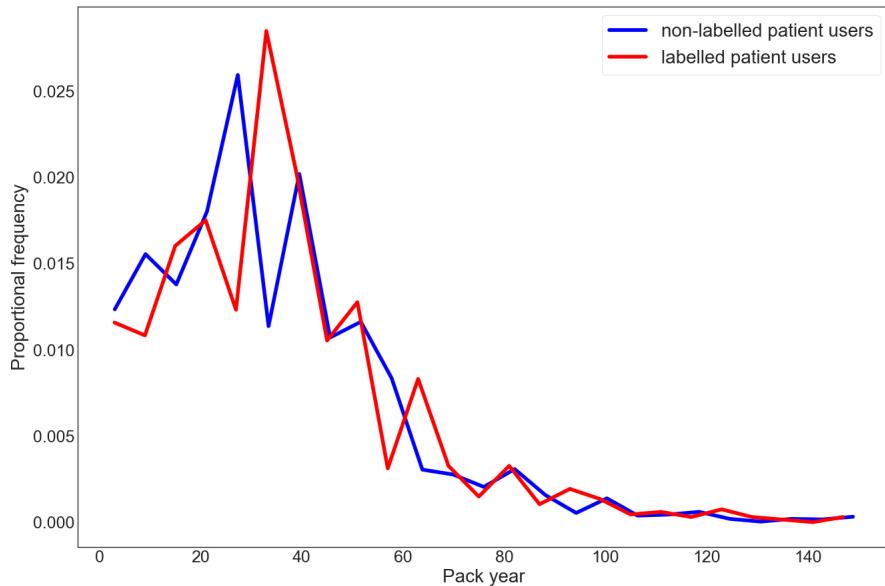


Figure 3.11: The proportional pack year distribution comparison between labelled and non-labelled patient users, using all available patient data.

3.3. RESULTS

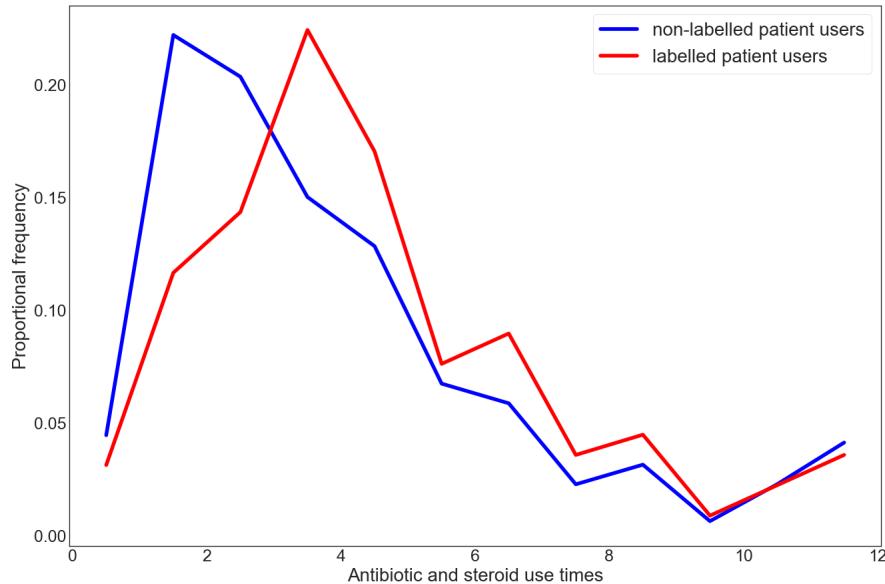


Figure 3.12: The proportional antibiotic and steroid use distribution in the last year between labelled patient users and non-labelled patient users, using all available patient data.

The distribution of stable CAT scores is slightly more negatively skewed in labelled patient users indicating that even in stable health, COPD symptoms are having a more severe impact on life in labelled patient users. This can be seen in Figure 3.13. The KS test indicates there is a statistically significant difference in the distribution ($p = < 0.0001$).

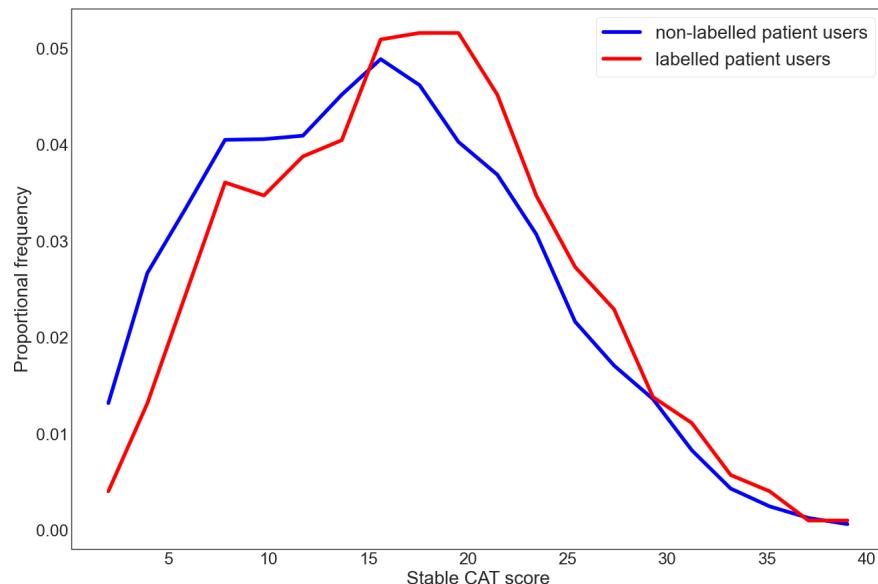


Figure 3.13: The proportional stable CAT score distribution between labelled patient users and non-labelled patient users, using all available patient data.

This analysis of demographics suggests that labelled patient users have more entries into the

myCOPD app and use the app more when their health is declining. Moreover, this group likely has more patient users with severe COPD, reports greater disability, and tends to have more exacerbations/use of antibiotics and steroids.

3.3.4 Feature visualisations

Before a labelled stable or labelled exacerbation event, the dynamic features (CAT score, tile clicks, regular medication use, reliever medication use, and symptom scores) are always higher before an exacerbation event compared to a stable event. This can be seen in Figures 3.14, 3.15, 3.15, 3.16, 3.17, and 3.18. Figure units for CAT scores represent scale points reported for that day, while for tile clicks (user interactions with graphical elements or "tiles" within the app's user interface), they indicate the number of tiles clicked within the app on that day. Similarly, for regular and reliever medication, the units correspond to the number of times a regular or reliever inhaler/medication was used on that specific day. It's noteworthy that tile clicks, regular medication use, reliever medication use, and symptom scores are close in range between exacerbations and stable health. Typically, these parameters are within a range of less than 1 or 1.5 points of each other, persisting consistently until the occurrence of, or in proximity to, an exacerbation. However, the mean of CAT scores differs by nearly 4 points 14 days before an event. This gap steadily increases rising to between a 6-8 point difference 1 to 2 days before an event.

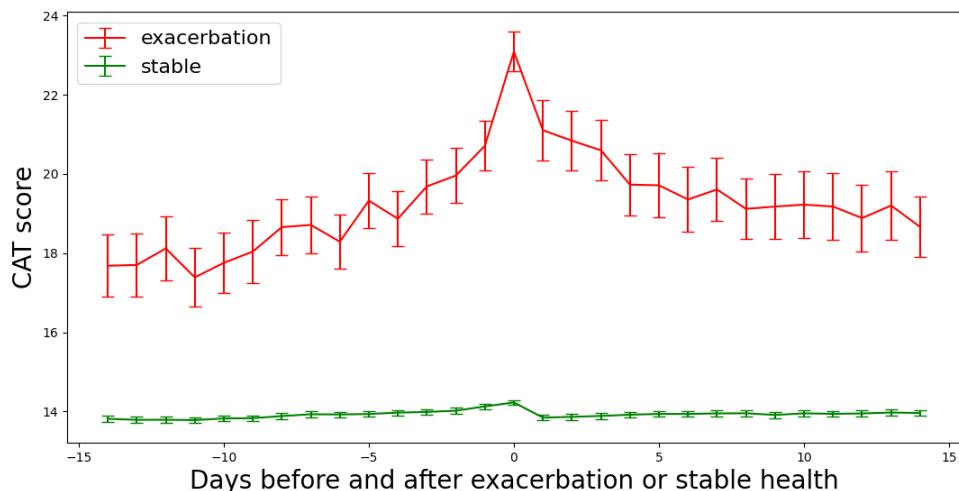


Figure 3.14: The mean **CAT scores** from all patients, recorded 14 days before an exacerbation or stable label. Error bars signify 95% CI intervals.

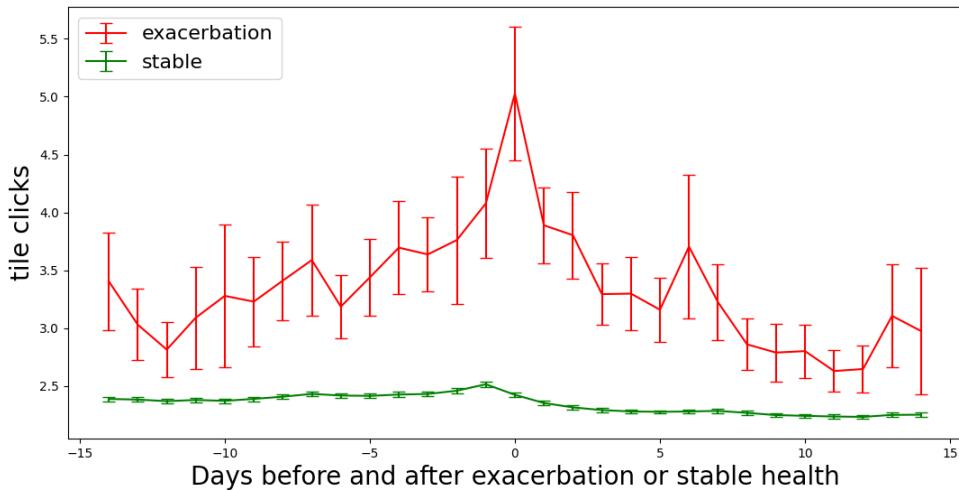


Figure 3.15: The mean **tile clicks** from all patients, recorded 14 days before and after an exacerbation or stable label. Error bars signify 95% CI intervals.

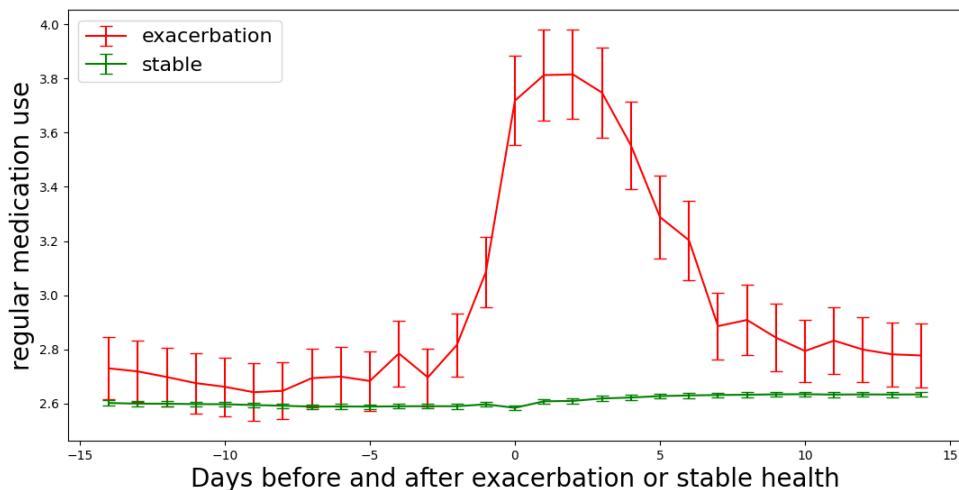


Figure 3.16: The mean **regular medication use** from all patients, recorded 14 days before and after an exacerbation or stable label. Error bars signify 95% CI intervals.

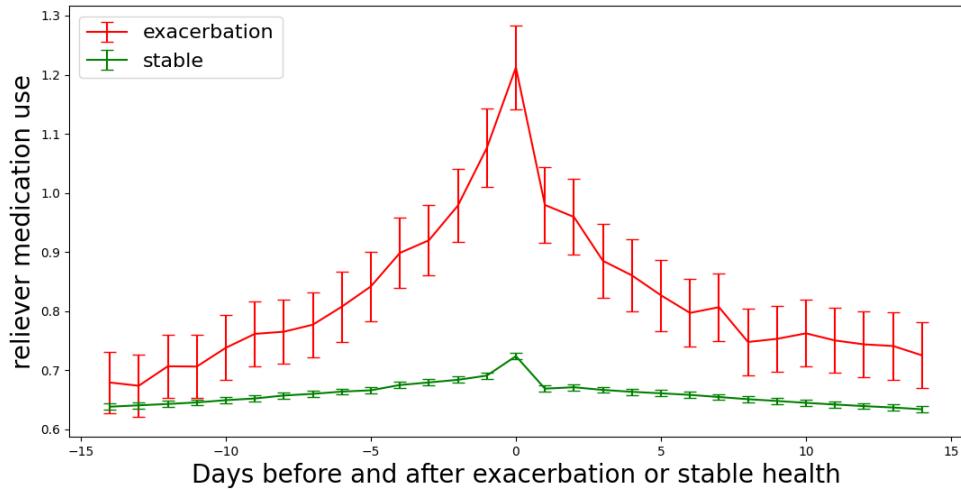


Figure 3.17: The mean **reliever medication use** from all patients, recorded 14 days before and after an exacerbation or stable label. Error bars signify 95% CI intervals.

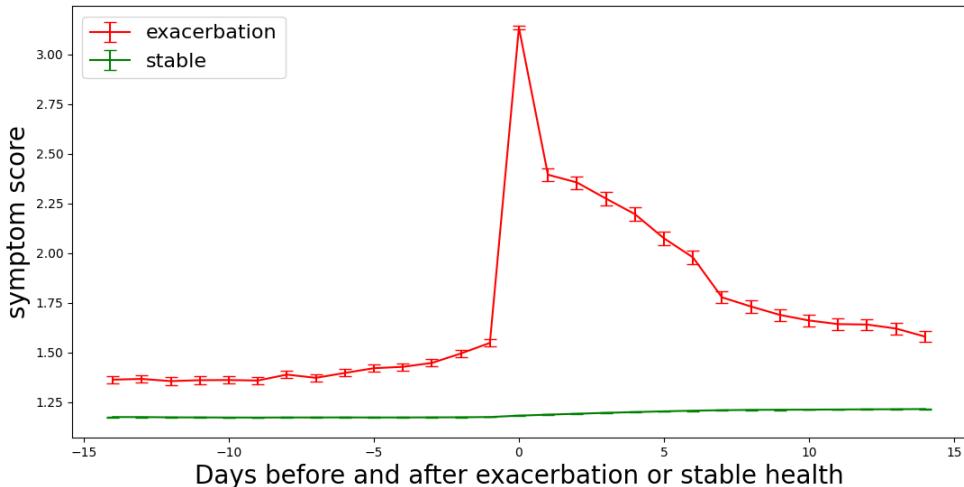


Figure 3.18: The mean **symptom scores** from all patients, recorded 14 days before and after an exacerbation or stable label. Error bars signify 95% CI intervals.

Upon evaluating these features based on the deviation from a stable baseline mean, the outcomes are less striking. Despite the lack of overlap in confidence interval intervals, the distinction between stable and exacerbation time points is minimal, except for the change in CAT score from a mean CAT score during stable health, which exhibits a noteworthy difference. These differences can be seen in Figures 3.19, 3.20, 3.21, and 3.22. The average change in CAT score increases to 2, four days before an exacerbation label. Notably, the minimum clinically important difference (MCID) for the CAT score is 2 according to Kon *et al.* [108]. This provides evidence for the notion that the change in CAT score from a stable baseline can serve as a key

3.3. RESULTS

predictive feature for AECOPD.

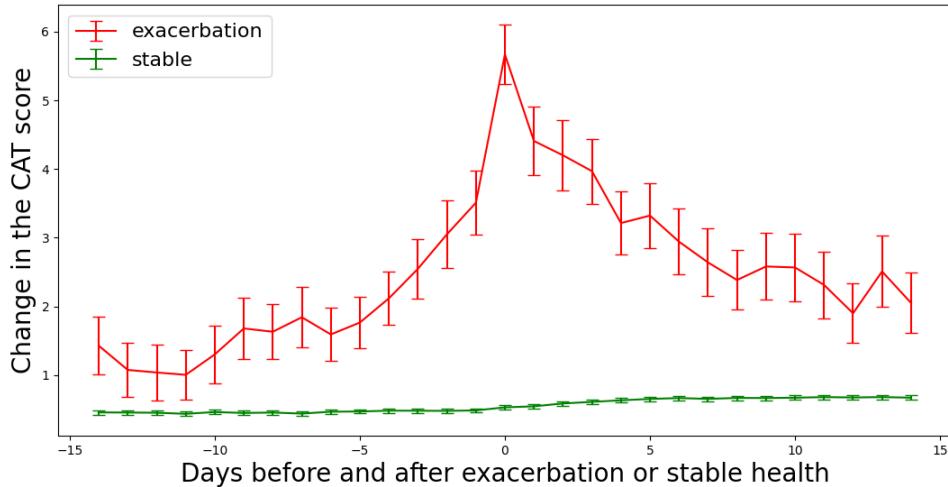


Figure 3.19: The figure illustrates the mean of the **change in CAT scores** relative to the mean stable CAT score for each patient, calculated 14 days before and following an exacerbation or stable label. Error bars represent 95% confidence intervals.

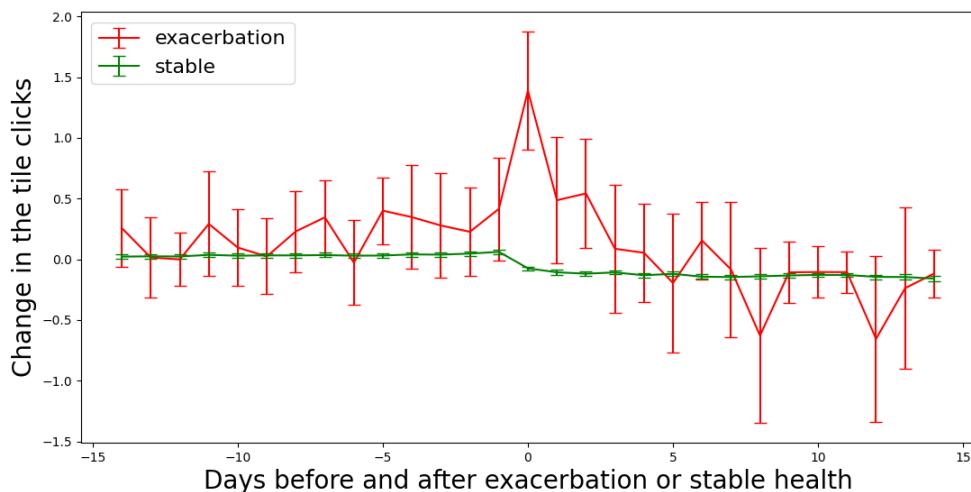


Figure 3.20: The figure illustrates the mean of the **change in tile clicks** relative to the mean stable tile clicks for each patient, calculated 14 days before and following an exacerbation or stable label. Error bars represent 95% confidence intervals.

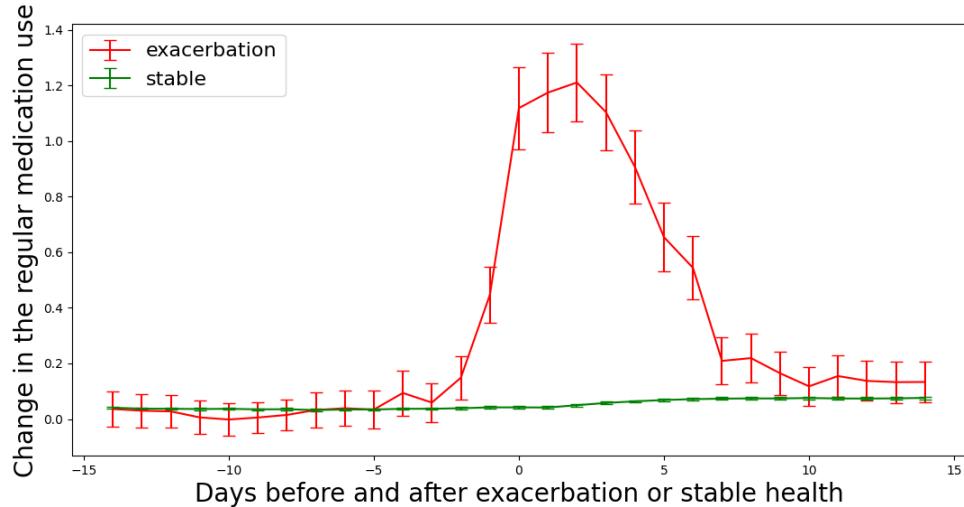


Figure 3.21: The figure illustrates the mean of the **change in the regular medication** relative to the mean stable regular medication use for each patient, calculated 14 days before and following an exacerbation or stable label. Error bars represent 95% confidence intervals.

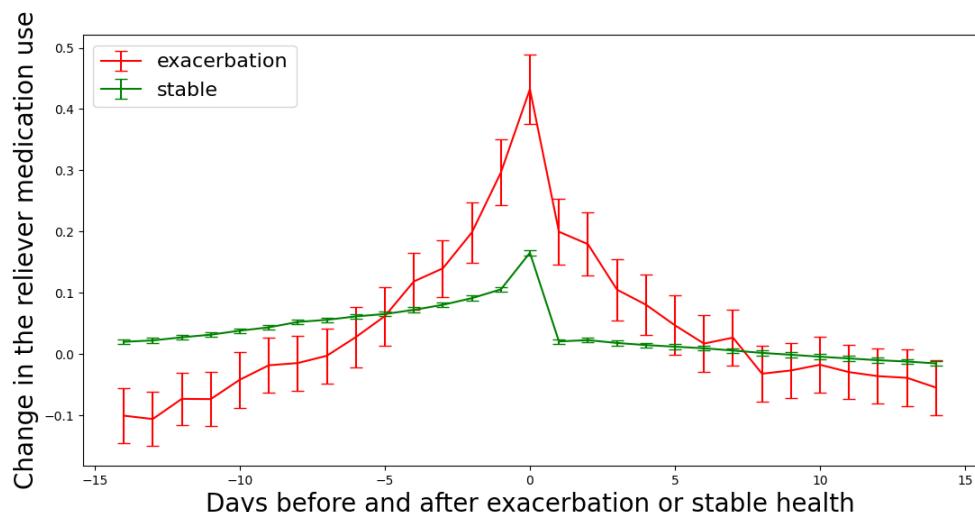


Figure 3.22: The figure illustrates the mean of the **change in the reliever medication** relative to the mean stable reliever medication use for each patient, calculated 14 days before and following an exacerbation or stable label. Error bars represent 95% confidence intervals.

Using Gini impurity and feature importance, the dynamic features were compared to determine their potential for AECOPD prediction, as shown in Figure 3.23. Change in CAT score has the highest relative importance indicating it results in better separation of exacerbation vs. stable classes.

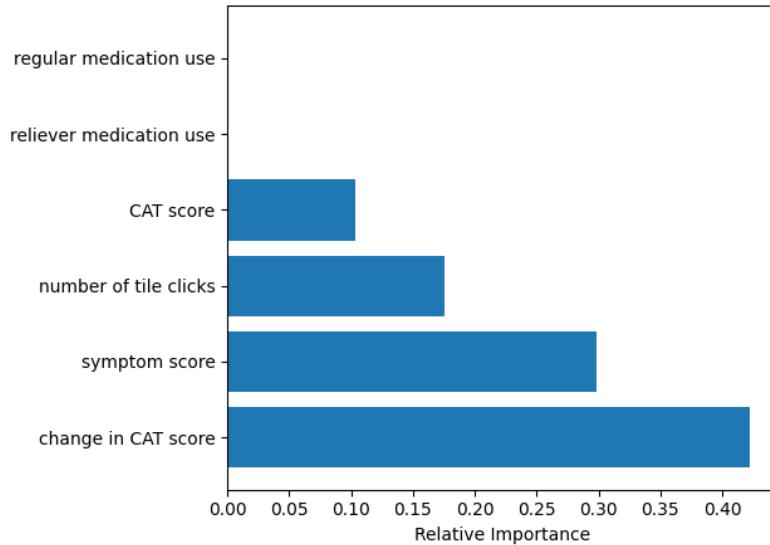


Figure 3.23: The relative importance of dynamic features for AECOPD prediction calculated from Gini impurity.

3.4 Discussion and conclusions

In this Chapter, I described the methodology I employed to process and transform the myCOPD patient-user data. The methods section details the process of preparing, organising and transforming the raw CSV files into a pandas DataFrame that is suitable for analyses and the application of machine learning. Part of this methodology describes the application of a label to capture stable health and the initiation of exacerbations. The exacerbation label was created by selecting data subject to the criterion that no symptom score of 3 or 4 was present for 14 days, there was at least one entry for 1-8 days, and on the day the patient-user input a symptom score of 3 or 4. The stable label works the same except on the day the patient-user entered a symptom score of 1 or 2. The visualisations of the labels matched the clinical definition as agreed upon by consensus in a co-creation discussion. The demographic analysis indicates that the labelled group are in general more engaged in using the myCOPD app and tends to have slightly more severe COPD. Visualisation of dynamic features in the myCOPD data suggests all features have the potential for predicting AECOPD. However, the CAT and more importantly the change in CAT score from a stable CAT score is likely to be the most important predictor.

3.4.1 Wrangling

The purpose of wrangling the myCOPD data was to generate a DataFrame for the application of machine learning to develop AECOPD prediction models. With this in mind, decision-making about how to structure the data was important as this early decision heavily influenced the type of machine learning methods that could be applied. I opted to establish a feature window encompassing the 14 days before each entry. The labels were generated using this 14-day window and visualised through a 30-day window. Subsequently, this label was assigned to the 1-8 days preceding an upcoming exacerbation or stable label. Patient data was then attached to this label. The choice to wrangle the data in this way creates a binary classification problem. The machine learning algorithm is trained on a set of data (CAT score, tile clicks, age, mMRC dyspnea score, pack year, etc.) with a label of either exacerbation or stable. The original raw CSV files of the dynamic data were organised as a time series. However, due to the sparsity of the data it was necessitated that the myCOPD data was restricted to a DataFrame as cross-sectional.

3.4.2 Labeling

As discussed in Section 2.3.4, there is heterogeneity in the literature when it comes to labelling exacerbations. Symptom-based labels have the advantage of capturing a large number of self-reported exacerbations that can lead to models that are perhaps more useful for patient self-management. Event-based labels capture the most severe events which as a result may have variables that produce stronger signals prodromal to exacerbations, resulting in better-performing predictive models. We believe that basing our definition on when patients self-report that their symptoms have significantly worsened and they have or will use their rescue pack or contact primary healthcare or emergency services is appropriate as this fits closely to GOLD's definition and clinical consensus. Moreover, these definitions indicate the exacerbation label is within the remit of both a symptom and event-based label.

Interestingly, earlier in the chapter where a description of the methodology for label creation was discussed in Section 3.2.2, it was explained that the label was created from the original definition “acute worsening of respiratory symptoms that results in additional therapy”. So, patients’ self-reported use of antibiotics and steroids or search for further support and treatment through GP or 999 emergency services in symptom scores 3 and 4 supports this definition. Clinician and engineer consensus in the co-creation meetings established the 14-day window of no symptom scores of 3 and 4 which was developed to prevent capturing ongoing exacerbations. Interestingly, the 14-day exacerbation-free period ties in closely with GOLD’s new exacerbation definition in which COPD symptoms (dyspnea, cough, sputum) worsen over 14 days. The

3.4. DISCUSSION AND CONCLUSIONS

14-day exacerbation-free window used in our label is similar to the recent update in definition which I think strengthens the validity of our decision-making regarding the label.

There are some unavoidable drawbacks to this approach, chiefly the lack of external validation concerning exacerbation self-report. There is a subjectivity to the myCOPD symptom score reporting in which it is determined by the patient-user that they are exacerbating. These self-reports are not validated by clinicians and other than the CAT there are no objective measures to support or verify the patient user's decision-making. This is problematic as potentially the label developed during this PhD lacks generalisability to other labels that were defined by clinician diagnosis or hospitalisation, thus reducing external validity. Moreover, if a patient has registered the use of their rescue medication (antibiotics and steroids) but in the following days realises they did not experience an exacerbation and it was just an isolated bad day, there is no way to remove that report from the symptom scores. This issue could have been resolved by programming into the label that a patient must register another symptom score of 3 or 4 within the next 14 days. However, this risks losing data where a patient-user did not use the myCOPD app when exacerbating or potential rapid recovery in the case of mild exacerbations. Conversely, there is no gold standard diagnostic test for AECOPD and clinicians will often diagnose based on a patient's symptoms and vital signs. So, there is not a complete lack of external validity as whilst the patient user's report is not being verified, clinicians will often make a diagnosis using patient symptoms as is done with symptom score reporting in the myCOPD app.

Many machine learning studies, as found in Chapter 2, also adopt a similar approach, such as Shah *et al.*, who used data from the EDGE COPD clinical trial [64]. In the study, exacerbations were defined by changes in medication or contact with healthcare practitioners due to worsening symptoms. They employed a finite state machine (FSM) model with three states: normal, transitional, and exacerbation. Patients moved between these states based on medication changes. Their data was segmented into 7-day stable or prodromal periods with 7-day guard bands to allow symptoms to return to normal before transitioning. Similarly, Mohktar *et al.* identified patients at high risk of exacerbations based on specific criteria [61]. These criteria included symptom changes and reliever inhaler use, with a 2-week gap from the previous high-risk period. While some studies like Shah *et al.* benefited from clinical oversight in an RCT context, enhancing label validity, clinical oversight has limitations, including potential inapplicability to the broader COPD population due to sample size constraints and research setting biases compared to real-world RPM data collection methods.

Chmiel *et al.*, in their model development and validation study using myCOPD data, defined exacerbations as symptom scores of 3 or 4 and stable health as symptom scores of 1 and 2

[76]. Importantly, they established and predicted exacerbations within a 3-day window. The label we developed, in line with GOLD's current definition, requires a 14-day exacerbation-free window to prevent prediction events during ongoing exacerbations and has an 8-day prediction horizon. Chmiel *et al.* used various symptom scores and exacerbation counts as features in their model. However, using symptom scores this way during exacerbations may affect the model's performance in identifying the initial onset of exacerbation events. Their model is more likely to predict an exacerbation if there have been recent scores of 3 or 4, making it less effective in the prodromal period before the first instance of an exacerbation. The goal of exacerbation prediction, as indicated by patients, is to identify when an exacerbation will occur. Chmiel *et al.*'s model is more suited to identify the likelihood of sustained exacerbation symptoms and may not generalise well to data where an individual is in a stable period with an impending exacerbation. Some machine learning papers in the review use a 14-day window to separate stable health and exacerbations, which is essential for effective prediction.

3.4.3 Demographics

There were significantly fewer patient users who had at least one event that matched the stable label and one event that matched the exacerbation label ($n=1,758$) compared to those who didn't have both labels ($n=11,273$). Analysis and later modelling were restricted to this small portion of regular users. We did consider and discuss this carefully at the time, and while this approach may limit generalisability to all patient users, there was substantial foundation to justify our reasoning. A key justification was that a final exacerbation prediction model would only be used by those who both have exacerbations and use the app while exacerbating. That means in future research a developed intervention would be suitable for this group of people. This argument makes sense in the context of both future model development and validation for clinical utility in RCT. For future model development, if the intention is to incorporate sensors to enhance prediction performance, the patient population will be required to have several exacerbations and log those exacerbations within the myCOPD app so models can be trained. Similarly, in RCT, the effectiveness of the intervention will be measured through exacerbation-related outcomes (hospitalisation, exacerbation frequency, severity and readmission). Incorporating individuals who do not exacerbate will affect the power calculation requiring much larger sample sizes, resources and time. Therefore, we decided at this stage to model patients who use the myCOPD app acutely and whilst stable. Overall, we were interested in the predictive potential of the platform, however, in the future, a larger number of patient users could have greater engagement as access to an exacerbation prediction tool could drive behaviour change itself.

It was important to demographically analyse these two groups to identify if there was a significant enough difference to influence the performance and generalisability of the AECOPD

3.4. DISCUSSION AND CONCLUSIONS

prediction model. From use analysis, there was a clear division in that labelled users are much more engaged with the myCOPD app. Despite the difference in the size of the populations, the labelled group has slightly more symptom scores meaning there are slightly more new day logins into the app. Importantly, there was almost double the number of symptom score 2. This suggests that labelled patient users are using the myCOPD app more as their health declines from their normal state which could be why they are part of the labelled group as they are using the myCOPD app before exacerbations. Moreover, analysis of app usage suggests on average the labelled group has used the myCOPD app more, a lower one-time usage or app churn rate, and a greater average length of time using the app. Using Cooper *et al.*'s definition of high engagement as previously mentioned in 3.3.2, the proportion of very highly engaged users is much greater in the labelled group. Nevertheless, there is still a large number of users with very high engagement in the non-labelled group. This group of patient users engages in the use of the myCOPD app frequently, however, they either do not suffer from exacerbations or do not report exacerbations in the app and so did not meet the requirements for the labelled group. The group has been termed "non-labelled" despite the presence of some individuals with stable labelled data because this category comprises users who lack either an exacerbation label, a stable label, or both. To be classified as a labelled user, both an exacerbation and a stable label must be present. In terms of demographic differences, there is no real observable difference in age or gender. However, the GOLD group, use of antibiotics and steroids, and stable CAT scores differ with statistical significance. It appears that labelled patient users tend to have more severe COPD likely having worse symptoms and more frequent exacerbations. This may affect the generalisability and validity of the results. If the AECOPD prediction model is trained for users with more severe disease it could lack the sensitivity to predict exacerbations in those with milder symptoms and potentially less severe events. Conversely, those with COPD who have less severe disease may not find a use for exacerbation prediction or exacerbation support. From the differences described in the two groups, I would contend, that the process of developing and deploying models should be directed towards distinct subpopulations within the context of COPD, each characterised by specific phenotypic attributes. From the demographic analysis, it is apparent the group of patients selected for modelling are likely phenotypically different; they are more prone to exacerbations, have worse symptoms and engage much more with the myCOPD digital therapeutic to support their health. Therefore, when testing the AECOPD prediction model in either a validation study or RCT, the population recruited needs to have these characteristics as such a model will likely be considerably less effective in patients with mild or moderate disease who exacerbate rarely.

3.4.4 Features

From the visualisations of the different dynamic features, two features stand above the rest. CAT score and change in CAT score demonstrate the greatest separation between stable and exacerbation labels before the event. The other features are either separated by less than one point or in the case of change, often overlap. This evidence suggests CAT score and more specifically change in CAT score will likely be the most predictive of AECOPD with regards to modelling. Figure 3.23 quantifies the contribution of change in CAT score over other features. The CAT score encourages patient engagement and communication about their symptoms. A change in CAT score can mean an individual is coughing more, producing more phlegm, experiencing greater breathlessness and having affected sleep all of which are related to deterioration before AECOPD. It is for this reason that change in CAT likely outperforms CAT itself and the other features. This is corroborated by Lin *et al.* in their recent paper finding change in CAT is a better predictor for AECOPD than CAT [109]. Surprisingly, the symptom score before an exacerbation label and a stable label is similar. A symptom score of 2 indicates symptom worsening and the use of a reliever inhaler so it was expected that the mean symptom score would be closer to 2 before an exacerbation. The reason for this apparent difference between exacerbation and stable health is unclear but may be particularly explained by the sparsity of data before an exacerbation compared to stable health.

3.4.5 Conclusions

In this chapter, I have described the process I undertook to structure the myCOPD data for modelling. I explained how this data was labelled to apply machine learning and provided insight into the differences between the patient users whose data met the labels and those that didn't. Finally, I present some initial visualisations of the dynamic features to develop an understanding of the potential predictive capacity of the features. The work presented in this chapter provides the foundation for which I apply machine learning to generate models for predicting AECOPD. My approach to this problem is detailed in the next chapter, Chapter 4.

Chapter 4

Exacerbation Predictive Modelling

The preceding chapter described the data preparation procedure for developing a dataset amenable to the application of machine learning algorithms. This chapter elucidates an investigation conducted to apply machine learning methodologies to the largest available real-world dataset, endeavouring to ascertain the feasibility of predicting AECOPD for the enhancement of prospective patient outcomes. The primary aim of this study was the development and validation of a prediction model, with the capacity to predict AECOPD occurring within a prediction horizon spanning 1 to 8 days before the actual exacerbation event. This predictive model leverages patient-entered data extracted from the myCOPD patient database. Within the scope of this study, various adaptations of the AdaBoost algorithm were employed as the principal machine learning approaches. The dataset under examination comprised records of 506 patients, with a range extending from 2017. Within this dataset, there existed 55,066 instances of stable health and 1,263 instances of exacerbations as defined by participants' responses when asked to report their symptoms on logging into the myCOPD app. The input features considered in the model encompassed the CAT scores, symptom scores, smoking history, and exacerbation frequency. The analytical framework featured the utilisation of the EasyEnsemble Classifier, yielding a Sensitivity of 67.0% and a Specificity of 65.0%, translating to a positive predictive value (PPV) of 5.0% and a negative predictive value (NPV) of 98.9%. Subsequently, a more advanced model, incorporating an AdaBoost algorithm with a cost-sensitive decision tree, was implemented, yielding a Sensitivity of 35.0% and a Specificity of 89.0%, thereby resulting in a PPV of 7.08% and an NPV of 98.3%. This preliminary analysis demonstrates that real-world data from a widely deployed digital therapeutic has the potential to predict AECOPD. Nevertheless, further enhancement of the models will likely be necessary before delving into efficacy studies within a patient population.

4.1 Introduction

A comprehensive dual literature review (Chapter 2) was conducted and identified that machine learning has the potential to enable accurate predictions of AECOPD to improve patient outcomes. Part of this review identified the potential burden to patient users when actively

engaging in the use of several sensors daily. Moreover, interventions that require substantial oversight by a clinical monitoring team are likely resource-intensive and potentially unfeasible [110].

This chapter builds upon the previous work from Chapter 3 using the data that was transformed into a suitable format for machine learning. The core of this chapter lies in the exploration and selection of appropriate data modelling methodologies. Machine learning algorithms, predictive analytics, and data-driven insights will be considered in constructing a robust framework capable of handling the diverse and dynamic nature of myCOPD patient data. The users of the myCOPD app use the app to support their own disease management in their own capacity. Modelling from the myCOPD self-management data means there is a minimal burden to the user for monitoring as the myCOPD app is already in use to support their health. We chose this approach because the ability to deploy a predictive exacerbation model within an app reduces the need for extensive engagement from the patient user and a team of healthcare professionals to monitor patient data daily.

4.2 Methodology

4.2.1 Data Extraction and Preparation

The myCOPD data extract created on December 7 2021, contained 13682 activated patient users (users who have activated their account and logged into the app). We then trained models on these previous entries to make predictions on the patient's exacerbation status in the future. We then conducted co-creation workshops between engineers and clinicians to define an exacerbation label. Exacerbations were forecasted within a 1-8 day timeframe based on clinical consensus derived from these workshops, indicating that this window is deemed suitable for timely and effective intervention.

Patients were not included in the sample if they were missing any one of the baseline characteristics such as age, GOLD group, and pack year ($n=3,998$). Lastly, we removed myCOPD users who had not reported at least one exacerbation event that matched the label to ensure the sample only included patients who experienced both a stable state and an exacerbation state and also had CAT scores available ($n=1,721$). This is because the current modelling approach is focused on patient users who are highly engaged in the app and use the system for support when they are stable and when they are exacerbating. This resulted in 506 myCOPD patient users with 55,066 stable state records and 1,263 exacerbation state records. Details of the characteristics of the study population are included in Table 4.1.

Table 4.1: Summary statistics of study participants.

Characteristic	Participants n=506
Age in years (mean, SD)	68.5 (10.2)
Sex (male, female, not available)	221 (44%), 193 (38%), 92 (18%)
GOLD classification of severity of airflow limitation (number of participants)	
A	44 (9%)
B	137 (27%)
C	51 (10%)
D	274 (54%)
mMRC dyspnea score (number of participants)	
0	42 (8%)
1	150 (30%)
2	167 (33%)
3	86 (17%)
4	61 (12%)
Smoking status (number of participants)	
Never smoked	14 (3%)
Ex-smoker	440 (87%)
Current smoker	52 (10%)
Pack year (mean, SD)	35.1 (25.6)
Number of hospital admissions in the previous year (mean, SD)	0.332 (0.971)
Number of rescue packs in the previous year (mean, SD)	2.28 (2.48)

4.2.2 Study Design

We selected and engineered 16 variables: health statuses, disease severity, and demographic features available from myCOPD patient user data. These variables were subjected to one-hot encoding to render them suitable for utilisation with machine learning algorithms. The feature engineering consisted of creating a mean stable state CAT score and a change in CAT score. The mean stable state CAT score was generated by taking a mean of the CAT scores for each patient on days they reported a symptom score of 1 or “normal for me”. The change in the CAT score was calculated by subtracting the mean stable state CAT score from their most recent CAT score. Once the data had been prepared, we trained the machine learning algorithms on the training sample and generated exacerbation prediction models. We assessed the performance of the models on a test set using receiver operating characteristic (ROC) curves and confusion matrices. A diagram illustrating the study design, as depicted in Figure 4.1, illustrates our approach to randomly allocating data points into training and testing sets. Rather than only applying random division to the data, we also employed a random assignment of participants into two distinct groups: the training sample (n=354) and the testing sample (n=152). These

groups contained 38,932 and 17,397 records, respectively.

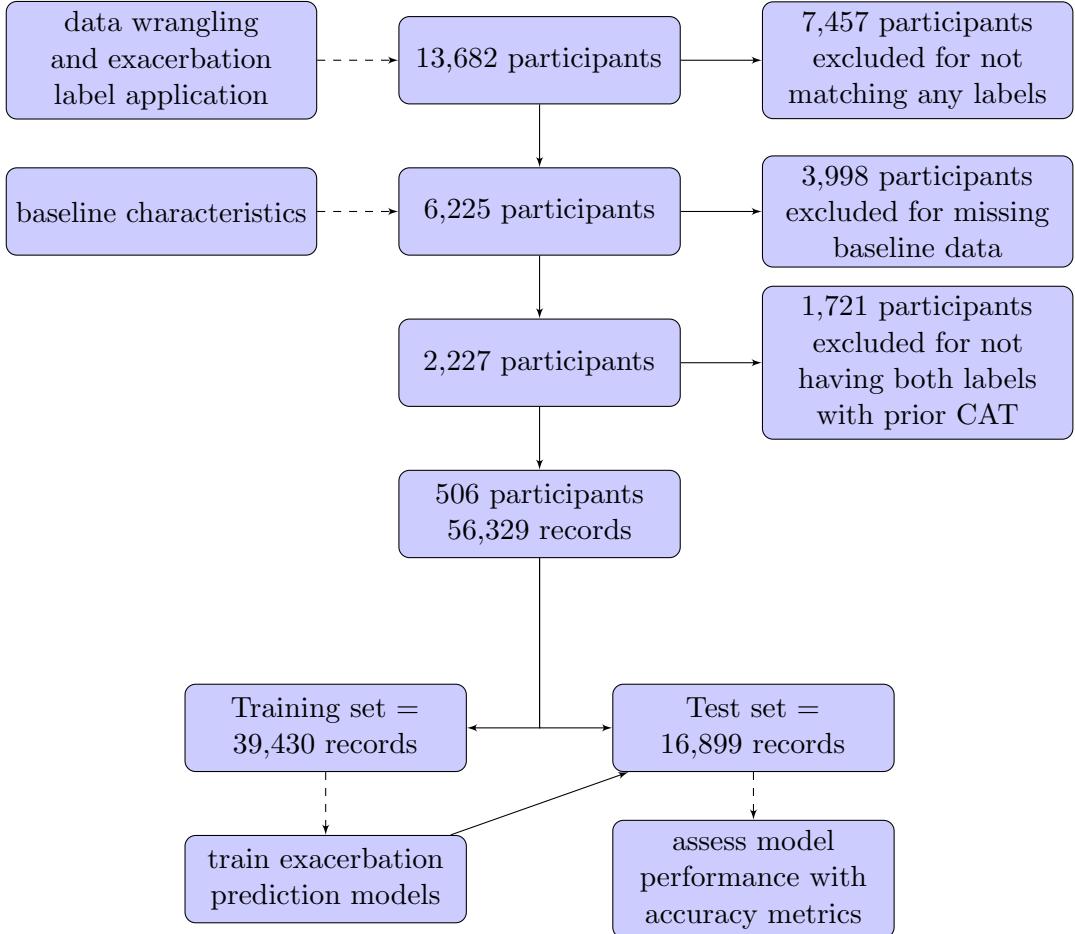


Figure 4.1: Schematic of participant selection and study design. A total of 506 participants with 56,329 records in myCOPD had their data randomly split into training (records = 39,430) and testing (records = 16,899) sets, leaving data from all participants spread across both sets. The machine learning algorithms were applied to the training dataset. Both models were tested in the testing set of the myCOPD sample.

4.2.3 Initial testing for predictive capacity

To understand the predictive ability of the data some initial modelling was undertaken which included the application of scikit-learn’s KNN [111] and then subsequently the combination of the imbalanced-learn’s Near-Miss algorithm [112] with KNN. KNN is a popular non-parametric classification and regression technique widely applied in machine learning. The KNN algorithm classifies an unlabeled instance by considering the majority class of its k nearest neighbours in the feature space. It relies on distance metrics, nearest-neighbour identification, and decision boundaries. Distance metrics quantify dissimilarity between instances and are crucial for KNN. Euclidean distance and Manhattan distance are commonly used for continuous features, while

Hamming distance is suitable for categorical data [113]. Euclidean distance is a measure of the straight-line distance between two points in a multi-dimensional space. Manhattan distance, also known as city block distance, is a measure of the distance between two points in a grid-like space, calculated by summing the absolute differences of their coordinates along each dimension. Hamming distance is a measure of the difference between two strings of equal length, calculated by counting the positions at which the corresponding elements differ. Manhattan distance is often favoured over Euclidean distance in situations involving high data dimensionality. Hamming distance quantifies the separation between categorical variables. Since the myCOPD data is not high dimensional and includes continuous variables, Euclidean distance was initially employed for modelling. Euclidean distance $d_{\text{Euclidean}}$ between instances x and y is computed as:

$$(4.1) \quad d_{\text{Euclidean}}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

KNN identifies the k nearest neighbours of an unlabeled instance by computing distances between it and labelled instances in the training dataset. The k nearest neighbours are determined based on the minimum distances. The algorithm assigns a label to the unlabeled instance based on the majority class among the neighbours. For regression, the predicted value is often the mean or median of the target variable values of the k nearest neighbours.

Decision boundaries separate classes in the feature space and arise from the distribution of labelled instances. In KNN, decision boundaries are implicit and equidistant from the nearest instances of different classes. The complexity of decision boundaries depends on the value of k . Smaller k captures finer-grained patterns but may be sensitive to noise and outliers leading to overfitting, while larger results in smoother boundaries which may generalise better but may oversimplify and reduce the model's ability to capture local variations. Moreover, KNN assumes that all features have equal importance, which might not hold in real-world scenarios where some features are more relevant than others. Lastly, KNN's performance can degrade if the data has imbalanced classes, as it tends to favour the majority class in classification tasks. However, the aim of KNN in this work was to identify initial predictive capacity, therefore, its limitations are not of great importance. The determination of k involved using the test data and error rate, wherein the error rate was computed as $1 - \text{accuracy_score}(y_{\text{test}}, y_{\text{pred}})$. The accuracy score function, sourced from scikit-learn, assesses the accuracy of predictions. This method resulted in the selection of $k = 28$ when employed in tandem with an undersampling algorithm, detailed further in Section 4.2.3.1. This particular approach was applied to address the challenge of class imbalance.

In summary, KNN was used to initially test the predictive capability of myCOPD data, a

suitable k was selected through error rate and the data was balanced before implementation. The results of this analysis can be seen in Section 4.3.1

4.2.3.1 NearMiss for under-sampling

After applying labels for exacerbations and stable health, it became evident that a significantly larger number of data points were categorised as stable health. Upon closer examination, it was apparent that there existed a substantial class imbalance between the exacerbation state and the stable state, with a ratio of 1:44. Kuhn and Johnson in their book on applied predictive modelling state “An imbalance occurs when one or more classes have very low proportions in the training data as compared to the other classes” [114]. To address this issue initially, a decision was made to employ a sampling algorithm in order to rebalance the class distribution and enhance the accuracy of the predictive model. For managing the class distribution, the chosen approach was to implement the NearMiss algorithm. NearMiss is an under-sampling technique designed to reduce the number of instances in the majority class (those labelled as stable), thereby equalising the class distribution. The algorithm achieves this by selecting a subset of the majority class instances based on their proximity to the minority class instances. NearMiss operates in a manner reminiscent of the K-Nearest Neighbors (KNN) algorithm, it calculates the distances between each majority class instance and its k closest neighbours from the minority class. Subsequently, these majority class instances are arranged in ascending order based on their distances to the minority class instances. Those instances with the shortest distances are deemed to be the most informative. Finally, a subset of the majority class instances is chosen based on predefined criteria, effectively addressing the class imbalance issue.

I used the imbalanced-learn package for their NearMiss algorithm [112]. There are three versions of the NearMiss algorithm:

1. (**version=1**) - **Majority class samples with the closest average distance to the three nearest minority class samples.** This version of NearMiss selects the majority class samples whose average distance to the three nearest minority class samples is the smallest. It aims to select samples from the majority class that are closest to the minority class.
2. (**version=2**) - **Majority class samples with the farthest average distance to the minority class samples.** In contrast to Version 1, Version 2 selects majority class samples whose average distance to the three nearest minority class samples is the largest. It focuses on selecting samples from the majority class that are farthest from the minority class.

3. **(version=3) - Majority class samples with the closest distance to each minority class sample.** Version 3 of NearMiss selects majority class samples that have the smallest individual distances to each minority class sample. In other words, it picks the majority class samples that are closest to any minority class sample.

I chose to use NearMiss version 2 in conjunction with KNN, aiming to remove the majority class instances that were farthest away from the minority class. This approach was motivated by the intention to address the class imbalance issue effectively and improve model performance by focusing on challenging-to-classify instances. The resulting set of selected majority class instances, combined with all minority class instances, constitutes the newly balanced dataset, ready for machine learning model training. NearMiss serves as a valuable tool for mitigating class imbalance, ensuring that the model avoids a bias towards the majority class by under-sampling its instances. The improvements NearMiss made to the KNN algorithm can be seen in Section 4.3.1.

4.2.4 The application of AdaBoost models

In their paper on class imbalance learning, Liu *et al.* emphasise the advantages of ensemble techniques for addressing class imbalance while enhancing overall performance [115]. Ensemble methods, such as bagging, boosting, and stacking, play a pivotal role in achieving this balance. Bagging, exemplified by Random Forest, constructs multiple base models by training them on various subsets of the training data through resampling with replacement. It then combines their predictions to mitigate overfitting and enhance generalisation. Boosting, represented by algorithms like AdaBoost (adaptive boosting), XGBoost, and LightGBM, builds a series of models sequentially. Each model focuses on correcting the errors made by its predecessors by assigning higher weights to misclassified instances. This dynamic approach adapts the model's attention to challenging data points, resulting in improved predictive performance. Stacking takes the concept further by employing a meta-model to combine predictions from diverse base models optimally. This method captures complex interactions in the data and often outperforms individual models.

These ensemble techniques harness the strengths of multiple models to enhance predictive accuracy and robustness, making them valuable tools for addressing complex real-world scenarios. For modelling the myCOPD data, the AdaBoost classifier was chosen from the boosting class due to its effectiveness in imbalanced classification tasks compared to bagging algorithms [116]. As previously stated the AdaBoost algorithm aims to create a strong classifier by iteratively training weak classifiers and assigning weights to training samples:

Weight Update (D_{t+1}):

$$(4.2) \quad D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

At each iteration t , AdaBoost updates the weights of the training samples ($D_{t+1}(i)$) to emphasise the samples that were misclassified by the previous weak classifier and de-emphasise the correctly classified ones. The weight update formula is shown above, where $D_t(i)$ is the weight of the i -th sample at iteration t , α_t is the weight assigned to the weak classifier $h_t(x)$, y_i is the true class label of the i -th sample, $h_t(x_i)$ is the prediction of the weak classifier h_t for the i -th sample, and Z_t is a normalisation factor.

Final AdaBoost Classifier ($H(x)$):

$$(4.3) \quad H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

After training T weak classifiers and updating weights, AdaBoost combines these weak classifiers into a final strong classifier, denoted as $H(x)$. The final classifier is a weighted sum of the weak classifiers, as shown above, where T is the total number of iterations, α_t is the weight assigned to the weak classifier $h_t(x)$, $h_t(x)$ is the prediction of the t -th weak classifier for the input x , and sign is the sign function. In the final AdaBoost classifier $H(x)$, each weak classifier $h_t(x)$ contributes to the prediction with a weight α_t . Weak classifiers with higher accuracy (lower weighted error) are given higher weights, making them more influential in the final prediction. The sign function ensures that the final prediction is binary, typically representing the class labels +1 and -1 for binary classification.

AdaBoost was used in two forms, as the EasyEnsemble Classifier or using an alternate weak learner to the decision stump. In the paper by Liu *et al.*, they report the use of a modified AdaBoost algorithm, the “EasyEnsemble Classifier” [115] available from imbalanced-learn. The EasyEnsemble Classifier is an ensemble of AdaBoost learners trained on different balanced bootstrap samples. The EasyEnsemble Classifier works by randomly undersampling the majority class from the original dataset every time a decision stump is trained. This means the ensemble prediction model is trained on balanced classes but has sampled many cases from the majority class, so the final model has more knowledge of the entire dataset.

As an alternative to the EasyEnsemble approach, which employs undersampling to address class imbalance, we utilised a class-weight strategy with a modified weak learner. Instead of the default decision stump utilised in AdaBoost, we opted for the Decision Tree Classifier (DTC)

from the scikit-learn package [117]. The DTC is a supervised machine learning algorithm that constructs a tree-like structure to make decisions regarding the target variable based on input features. It recursively partitions the dataset into subsets, guided by feature values, in a manner that maximises information gain or minimises impurity at each node. Impurity, in the context of decision trees, quantifies how effectively instances of different classes are separated within a node, influencing the tree's choice of splits to achieve improved class separation. The resulting tree serves for classification by traversing it from the root to a leaf node, where the final decision is rendered.

The process commences with the selection of a root node, representing the feature that yields the highest information gain or lowest impurity. The dataset is then partitioned into subsets based on the chosen feature. This feature selection and split point determination process recurs for each subset, progressively growing the tree until a predefined stopping criterion is met or when further splits do not significantly enhance classification performance. Subsequently, leaf nodes are assigned class labels based on the majority class within the respective subset or a “soft” classification with a probability assignment to each class based on the fraction of examples from that class.

DTC relies on metrics such as Gini impurity or entropy to assess the homogeneity or purity of data at each split. In this instance, entropy was employed as the criterion in our experiments. Entropy measures impurity or disorder within a node and is calculated as the sum of the negative logarithms of class probabilities. A lower entropy value signifies greater homogeneity among instances concerning their class labels. Decision trees aim to minimise entropy by selecting splits that result in more pure and well-separated child nodes. Notably, entropy was preferred over Gini impurity due to its superior performance in this specific model configuration. When calculating entropy for a specific subset S_{subset} within a decision tree, you need to consider both the classes and examples within that subset:

$$(4.4) \quad \text{Entropy}(S_{\text{subset}}) = - \sum_{i=1}^m \sum_{j=1}^n P_{ij} \log_2(P_{ij})$$

Here, i iterates over classes (e.g., class 1, class 0), j iterates over examples within the subset, and P_{ij} represents the class probabilities for each example j within the subset S_{subset} . The ranges for i and j are specified as i ranging from 1 to m and j ranging from 1 to n , where m and n are the respective numbers of classes and examples in the subset. When building a DTC, the goal is to minimise entropy by making decisions that result in more pure subsets. The process involves evaluating the entropy of different ways to split the data based on different

features and their threshold values. The decision that reduces entropy the most is chosen as the optimal split.

The information gain (IG) is used to measure the reduction in entropy after a split and is calculated as the difference between the entropy before the split and the weighted average of the entropies of the resulting subsets:

$$\text{Information Gain (IG)} = \text{Entropy before split} - \text{Weighted average of entropies after split}$$

In scikit-learn's DTC, the class weight parameter is a valuable tool for handling imbalanced datasets, where certain classes have significantly fewer samples than others. This parameter mathematically influences how the DTC accounts for misclassifications across different classes during training. Specifically, the class weight parameter can be incorporated into the impurity calculation (whether it's Gini impurity or entropy) used to determine the best splits at each node. By assigning appropriate class weights, the decision tree classifier becomes more effective at addressing imbalanced datasets, resulting in enhanced performance, especially in the minority class. The classifier will pay more attention to correctly classifying the minority class, even if it means making some mistakes on the majority class.

Class weighing can be determined by subject experts, hyperparameter tuning, or specified using a well-established best practice. In this case, tuning through grid search was used. Grid search works by trying out different combinations of settings specified for a model enabling visualisation of all the possible combinations of class weights and identifying the impact they have on model performance. For the metric of performance in the grid search, F1 was selected as it combines a model's precision and recall to provide a balanced evaluation of its performance in binary classification tasks. The grid search found the class weight ratio balance that should be applied in the stable class is 0.03 whereas the exacerbation class is 0.97.

Importantly, the DTC was selected to have a maximum depth of 3. This means that the tree will have a maximum depth of 3 levels from the root node to the leaf nodes. Each level represents a decision based on a feature, and the depth determines how many sequential decisions are made before reaching a final classification. It was necessary to use a small tree depth as the weak classifier of an AdaBoost is designed to perform only slightly better than random guessing and be computationally efficient; complex algorithms might lead to overfitting and counteract the boosting process's intended benefits.

To summarise, this strategy differs from the EasyEnsembleClassifier: whilst the AdaBoost algorithm is still used, the AdaBoost with CST does not undersample the dataset. Alternatively,

it uses a different weak learner; a DTC with a greater depth than the default weak learner (the decision stump). As we still had to mitigate the class imbalance, the class weight parameter was modified to give the exacerbation class significantly more weight (0.03:0.97). This CST was used as the weak learner within the AdaBoost to train the model.

4.2.5 Application of t-SNE

After the modelling phase, an examination was conducted to assess whether the data reached its limits either as a consequence of model complexity or inherent limitations in its predictive capabilities. This evaluation was performed on both the machine learning-prepared data and the data after the application of machine learning methods. Further details on this analysis are elaborated in Section 4.3.4. t-SNE (t-distributed Stochastic Neighbor Embedding) was applied for nonlinear dimensionality reduction to enable visualisation.

The t-SNE algorithm aims to create a low-dimensional representation of high-dimensional data points while retaining the inherent structure and relationships between them. It achieves this by modelling pairwise similarities between the data points and optimising an objective function that minimises the divergence between the high-dimensional and low-dimensional similarities. Unlike linear methods, t-SNE focuses on preserving local relationships and is particularly adept at capturing non-linear structures. It achieves this by modelling the probability distributions of pairwise similarities in both the high-dimensional and lower-dimensional spaces. During optimisation, t-SNE minimises the Kullback-Leibler (KL) divergence between these distributions. The KL divergence is a measure of how one probability distribution differs from a second, reference probability distribution. This works to make similar data points be pulled together while dissimilar points are pushed apart. This behaviour results in distinct clusters and clear separation in the visualisation, making t-SNE useful when exploring complex datasets and identifying hidden patterns.

Let $X = \{x_1, x_2, \dots, x_N\}$ represent the high-dimensional data points, with each x_i being a vector of d dimensions. The pairwise similarity between two data points x_i and x_j is computed as:

$$(4.5) \quad P_{ij} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i}^N \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)}$$

where σ_i is the bandwidth or perplexity parameter that determines the scale of the Gaussian distribution for each data point. The summation is taken over all indices k such that k is not

equal to i . The objective of t-SNE is to minimize the KL divergence between the high-dimensional similarities P and the lower-dimensional similarities Q . The cost function C is defined as:

$$(4.6) \quad C = \sum_i KL(P_i||Q_i) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

where $KL(P_i||Q_i)$ represents the KL divergence between the i -th row of P and Q , and q_{ij} denotes the similarity between embedded points in the lower-dimensional space. The double summation is taken over all indices j such that j is not equal to i . By iteratively optimising the objective function, t-SNE generates a lower-dimensional representation of the data that effectively captures the pairwise similarities and facilitates the visualisation and interpretation of complex structures.

t-SNE was used to reduce the dimensions in the initial dataset on the stable and exacerbation classes but was also used to reduce the dimensions of the classes after model application (true positive, false positive, true negative, false negative). This was chiefly done to understand if the predictive accuracy of models was not improving due to model application or due to the relationships between the classes in the feature space.

4.3 Results

Demographic characteristics of participants, markers of disease severity (including admissions in the previous year), and smoking status are displayed in Table 4.1. The participants in the modelling sample tend to fall into the GOLD stage D and the average number of rescue packs used by participants in the previous year was 2.28.

4.3.1 Initial exacerbation modelling

The KNN model was trained on a small subset of data, specifically, data collected 3 days before an exacerbation or stable health. Patient data was randomly split between training and test leaving some patients in both sets. The performance of the KNN models is shown in two separate confusion matrices in Figure 4.2 and Figure 4.3. The confusion matrix for the KNN model alone shows very low sensitivity which is due to the class imbalance. The sensitivity for the KNN with NearMiss is significantly higher but at a tradeoff for specificity which drops to 60.64%. This is likely because the NearMiss sampling strategy has removed examples from the majority class preventing the model from accurately classifying the stable state. The sensitivity,

specificity, PPV and NPV are displayed in Table 4.2. The model only performs slightly better than random but suggests the data has some predictive capability.

		ACTUAL	
		POSITIVE	NEGATIVE
PREDICTED	POSITIVE	TP (1)	FP (0)
	NEGATIVE	FN (11)	TN (564)

Figure 4.2: Confusion Matrix for the KNN model. TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative.

		ACTUAL	
		POSITIVE	NEGATIVE
PREDICTED	POSITIVE	TP (11)	FP (222)
	NEGATIVE	FN (1)	TN (342)

Figure 4.3: Confusion Matrix for the KNN model with NearMiss. TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative.

Table 4.2: Measures of performance of KNN exacerbation prediction models.

Model	Sensitivity	Specificity	PPV	NPV	Accuracy
KNN	8.33%	100%	100%	98.09%	98.09%
KNN with NearMiss	91.67%	60.64%	4.72%	99.71%	61.28%

4.3.2 Development of ensemble exacerbation prediction models

The AdaBoost models were trained using all available data from 1 to 8 days before an exacerbation or during stable health. Patient data was randomly partitioned into training and testing sets, with some patients included in both sets. In a specific case, the data was divided between two distinct groups of patients, with the training set containing 354 participants and the test set containing 152 participants. The ROC curves for the two models when all patient

data is split randomly across training and test sets and when the dataset is split into two groups of distinct patients are shown in Figure 4.4 and 4.5, respectively. The AUROC is different for the EasyEnsemble Classifier (0.730) compared to the AdaBoost with CST (0.626). Much of the improvement seems to be at a decision threshold leading to higher false positive rates. These results suggest the EasyEnsemble Classifier is a better predictive model on the dataset. The ROC curves for the two models when split by participant are shown in Figure 4.5. There is a difference in the AUROC for the EasyEnsemble Classifier (0.659) compared to the AdaBoost with CST (0.564). There is a reduction in AUROC for both models when split by participants indicating that the models may have less inter-patient generalisation.

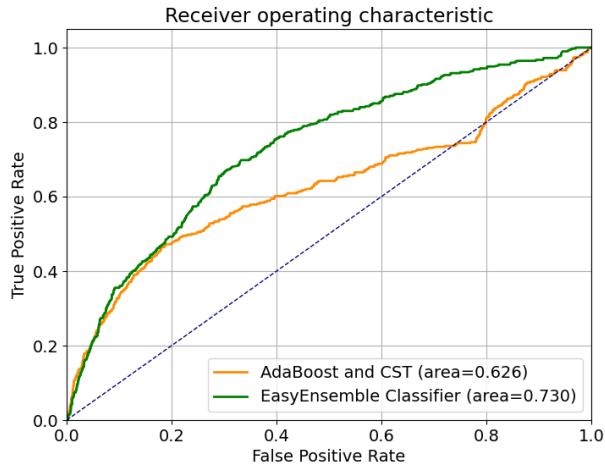


Figure 4.4: ROC curve comparing the EasyEnsemble model against the AdaBoost with a CST model. Models were trained on 70% and tested on 30% of the myCOPD sample.

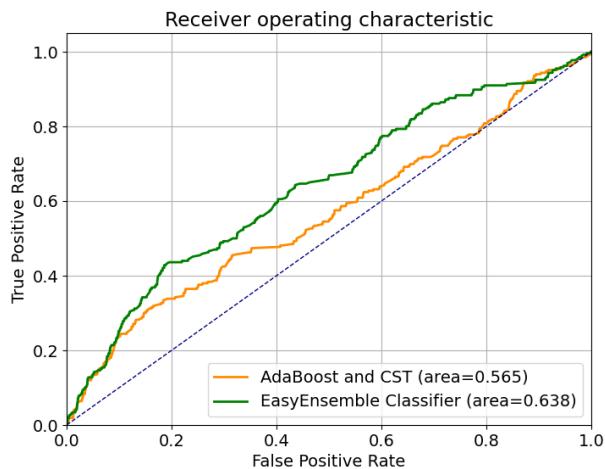


Figure 4.5: ROC curve comparing the EasyEnsemble model against the AdaBoost with a CST model. Models were trained on data from 354 participants and tested on data from 152 participants of the myCOPD sample.

The performance of the models when the dataset is split randomly is shown in two separate confusion matrices in table 4.6 and table 4.7. The confusion matrix for the EasyEnsemble Classifier shows a significant number of false positives, especially in comparison to the results of the AdaBoost with CST. However, the number of false negatives for the AdaBoost with CST is considerably higher than that of the EasyEnsemble Classifier. The sensitivity, specificity, PPV and NPV are displayed in Table 4.3. Whilst the ROC curves indicate the EasyEnsemble Classifier should be the better model, descriptive statistics calculated from the confusion matrices demonstrate that the AdaBoost with CST has significantly higher accuracy.

		ACTUAL	
		POSITIVE	NEGATIVE
PREDICTED	POSITIVE	TP (264)	FP (5038)
	NEGATIVE	FN (130)	TN (11467)

Figure 4.6: Confusion Matrix for the EasyEnsemble Classifier model. TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative.

		ACTUAL	
		POSITIVE	NEGATIVE
PREDICTED	POSITIVE	TP (138)	FP (1810)
	NEGATIVE	FN (256)	TN (11695)

Figure 4.7: Confusion Matrix for the AdaBoost with a CST model. TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative.

Table 4.3: Measures of performance of AdaBoost exacerbation prediction models.

Model	Sensitivity	Specificity	PPV	NPV	Accuracy
EasyEnsemble Classifier	67.0%	69.5%	5.0%	98.9%	69.4%
AbaBoost with CST	35.0%	89.0%	7.08%	98.3%	87.8%

The relative feature importance of the AdaBoost with the CST model is shown in Figure 4.8. CAT change, CAT mean, pack year, CAT score, and the number of rescue packs used in the last year are ranked as having the highest relative importance. The feature importance is an attribute of the AdaBoost algorithm derived from its base classifier. Therefore, feature importance is determined by the average feature importance provided by each CST. In the algorithm, CST relies on the concept of entropy. It selects the best split point based on the feature with the lowest entropy. Entropy is at its peak when the probabilities of the two classes are equal, and a node is considered pure when entropy reaches its minimum value, which is 0. This means that the change in CAT score has the lowest entropy and is most used by the tree to separate the classes. The change in the CAT score being the most important feature suggests the model is making timely predictions rather than classifying from the baseline characteristics.

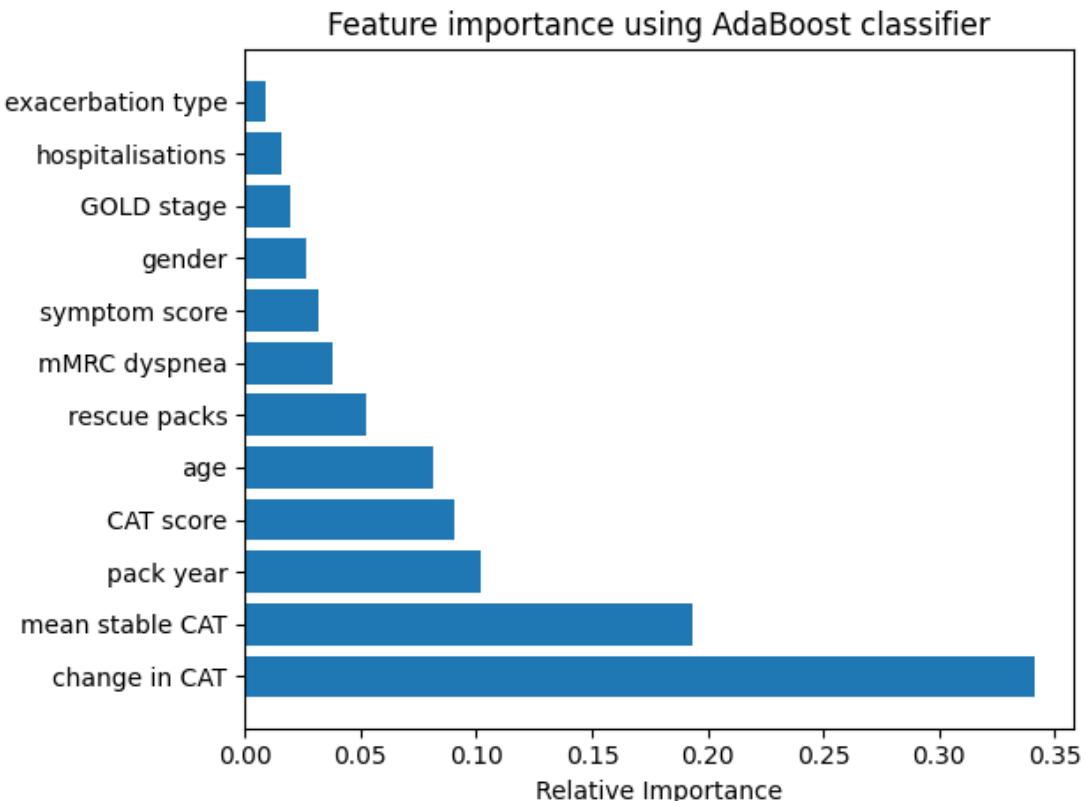


Figure 4.8: Feature importance computed from the AdaBoost with CST on the myCOPD dataset

We conducted a rigorous sanity check using the AdaBoost with CST algorithm. In this evaluation, the model was trained on data from the day of exacerbation and three days leading up to the event. Testing was performed on an independent set of unseen patients ($n=107$), comprising 8501 data points and 56 exacerbations. The obtained AUROC for this analysis was notably higher at 0.801, as depicted in Figure 4.9. The model showcased a sensitivity of 48.21%, a specificity of 94.21%, and an overall accuracy of 93.94%.

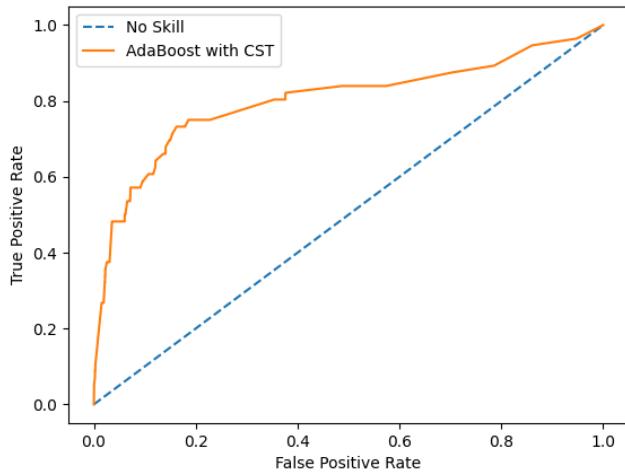


Figure 4.9: ROC curve for the AdaBoost with a CST model. The model was trained on the day of an exacerbation and three days prior. The training data contained 426 participants and was tested on data from 107 participants from the myCOPD sample.

The incorporation of the day of exacerbation into the training significantly enhanced predictive accuracy, serving as a robust sanity check. It's important to note that perfect accuracy wasn't achieved, likely attributed to variations in patient data, potential misreporting, and other inherent complexities associated with real-world datasets.

4.3.3 Misclassification analysis

The time leading up to exacerbation appears to be a crucial factor in misclassification. Figure 4.10 illustrates data from the same patient user. In the first row, data from 8 days before an exacerbation is present, while the second row contains data from 2 days before an exacerbation. Interestingly, the first row is misclassified as stable, whereas the second row is correctly identified as an exacerbation. Examining the correctly classified instance reveals a higher CAT score, an increased symptom score, and a positive change in CAT score, indicating a decline in health. These distinctions are likely temporal, suggesting that closer to the event, symptoms worsen. This implies a need to reduce the prediction horizon for more accurate predictions.

CAT score	symptom score	gender	age	GOLD stage	mMRC	dyspnea	pack year	rescue packs	hospitalisations	exacerbation type	mean stable CAT	CAT change in CAT
5	1	2	74	3	1	37	3	0	1	8.66667	-3.66667	
22	2	2	74	3	1	37	3	0	1	8.66667	13.33333	

Figure 4.10: Two rows of data extracted from a pandas DataFrame representing a single patient-user. The data was part of the testing set for the AdaBoost with CST model. The first row is data 8 days before an exacerbation and the second is 2 days before an exacerbation. However, the first row is misclassified as stable, while the second is correctly identified as an exacerbation by the model.

4.3.4 t-SNE for dimensionality reduction and visualisation

The results of the t-SNE 2D map for the stable and exacerbation label and the confusion matrix classes are shown in Figure 4.11 and Figure 4.12, respectively. For stable and exacerbation labels, t-SNE constructs a map in which the separation between the two labels/classes is minimal. There is some but it is complex, non-linear, and localised. Similarly, t-SNE with confusion matrix classes produces a map in which there is no clear boundary between different groups. The lack of separation is indicative that the current features for modelling have a limited capacity to predict AECOPD vs. stable health with a high degree of separation or accuracy.

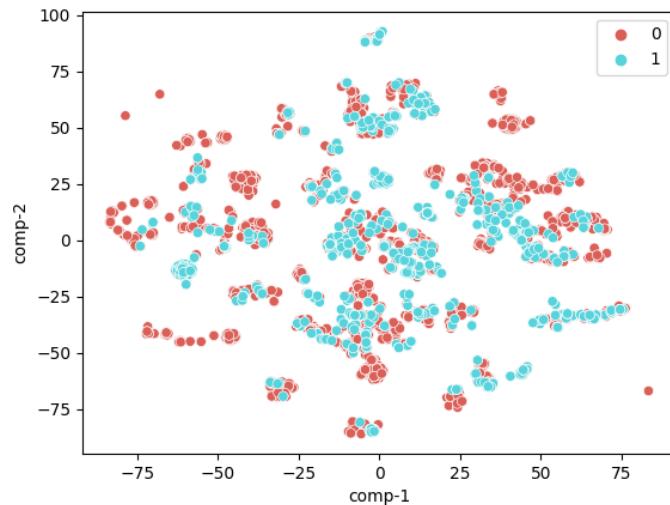


Figure 4.11: Data visualisations in 2D by t-SNE for exacerbation and stable labels. Colour: blue for exacerbation labelled data and red for stable labelled data.

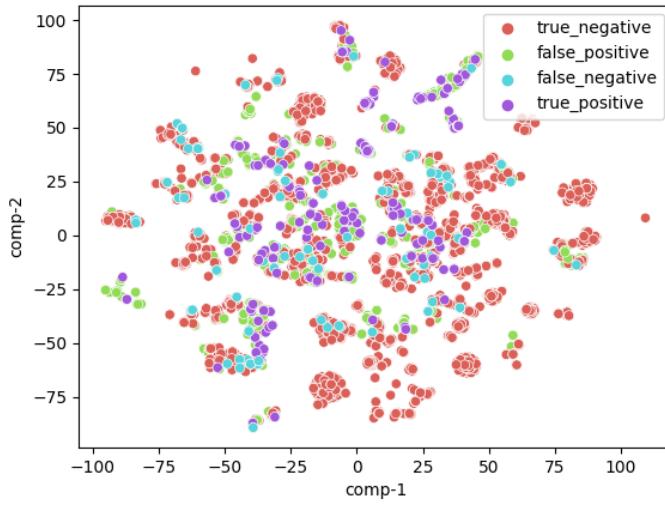


Figure 4.12: Data visualisations in 2D by t-SNE for confusion matrix classes. Colour: red for true negative, green for false positive, blue for false negative, and purple for true positive.

4.4 Discussion and conclusions

AdaBoost with a CST and an EasyEnsemble Classifier model were used to identify exacerbations using data extracted from a widely deployed digital therapeutic. AdaBoost with CST and EasyEnsemble Classifier models achieved a sensitivity of 67% and 35% and a specificity of 69.5% and 89%, respectively. Sensitivity and specificity play a pivotal role in shaping the false negative and false positive rates of exacerbation prediction models. False positives, stemming from low specificity, signify inaccurate predictions of exacerbation events, potentially causing unnecessary interventions, escalating healthcare costs, and heightening patient anxiety. Conversely, false negatives, associated with low sensitivity, indicate a failure to predict actual exacerbations, leading to missed opportunities for timely intervention and potentially worsening health outcomes. Striking a balance between these errors is vital for optimising model performance. Given the app’s prevalence of healthy periods among patients, prioritising increased specificity becomes crucial to minimise false positives. A higher false positive rate, in my view, poses a greater risk of overtreatment and technology rejection compared to a lower sensitivity.

In this work we conducted two different data splits; we split the data randomly into training and test sets and also split by participant, training the data on one group of participants and testing on another. Model development and validation studies in the literature split data for model training and testing using different strategies; splitting by participant, training models on individuals before deploying them on the same individuals or splitting on records across the dataset [64, 68, 74]. Splitting by participants is more realistic indicating the ability of a model to generalise. Not splitting by patient may provide overly optimistic results for the myCOPD

dataset and others, and it is likely that there is a need for a significantly larger patient pool for better inter-patient generalisation.

We used two variations of the boosting ensemble classification algorithm known as AdaBoost. The machine learning approach for this study is classification predictive modelling. Our goal was to classify patient data into two classes: the stable state and the exacerbation state. However, there is a significant disparity in the class sizes, leading to an imbalanced classification problem. We encounter a minor imbalance when the distribution of examples is slightly uneven, as in the case of a 4:6 ratio. A Severe imbalance will have a significantly larger difference in the distribution, such as a ratio of 1:100 or more. For the data in this study, the ratio is 1:44, meaning the problem is moderate to severe. Modelling with imbalanced data presents several challenges. Firstly, predicting the minority class is more difficult due to the limited number of instances available for that class. Furthermore, the abundance of cases in the majority class can overshadow the minority class, causing an imbalance. Most classification algorithms are tailored for scenarios with roughly equal class distributions, leading to a bias in the learning process towards the majority class [118]. This bias causes the final model to neglect the minority class, which, with regards to predicting exacerbations of COPD, is of greater interest.

A large proportion of machine learning approaches developed to overcome imbalanced classification problems have consisted of sample techniques, cost-sensitive learning and ensemble methods [119]. We used two modifications of the AdaBoost algorithm (an ensemble method). AdaBoost uses a process of fitting multiple weak learners (models that perform slightly better than random) that combine their output to create a weighted voter or a strong learner. This method has the advantage of preventing underfitting and overfitting. The first modification is the EasyEnsemble Classifier which deals with the classification problem by a sampling technique. The EasyEnsemble algorithm randomly undersamples the majority class on each iteration of training the week learner. As a result, the final predictive model has been trained on balanced classes so is more sensitive to the minority class (exacerbation state). Moreover, on each iteration of weak learner training, the algorithm randomly undersamples different parts of the data capturing a greater variety of the majority class. The second modification is adjusting the weak learner for cost-sensitive learning. To do this we incorporated a cost-sensitive tree that assigns higher weights to the class with a higher misclassification cost (exacerbation state). Whilst the AdaBoost algorithm is effective at reducing bias, it can be sensitive to outliers/noisy data because at each stage of learning it adds another weighted model to the overall classifier it is learning. The issue lies in the fact that misclassification penalties increase dramatically with the predictive function output's magnitude. This means that a single outlier or mislabeled point could yield a substantial influence over the final learned model. Nevertheless, since our dataset comprises highly engaged users, the likelihood of encountering outliers is low. Consequently,

4.4. DISCUSSION AND CONCLUSIONS

there exists a notable disparity in the predictive performance of the two models.

The specificity of the AdaBoost with CST is significantly higher (89.0%) than the EasyEnsemble Classifier which also translates to higher accuracy. The reason for the differing performance is likely due to the approaches selected for tackling the imbalanced classification problem. The EasyEnsemble Classifier uses an undersampling technique to balance the classes to improve model performance. Whilst the AdaBoost algorithm improves the scope of the data used for training the models, a significant amount of data from the majority class is still left unseen. This could have resulted in the specificity being lower than the AdaBoost with CST. However, the alternative cost-sensitive approach caused a significant drop in sensitivity. The weighting of the CST algorithm is designed to prevent the misclassification of the minority class. However, because the class imbalance is so large it seems that the majority class still overwhelms the training of the model. The myCOPD data consists of real-world data, so, there will likely always be a significant trade-off between sensitivity and specificity. Patients use the myCOPD app how they see fit and will modify their use based on their interpretations of the severity of their COPD.

The dataset used for modelling in this study consisted of both dynamic and fixed features. A model that ignores the dynamic features (CAT score, CAT change, and symptom score) is not making timely predictions. The feature importance indicates the variables the model is likely to use for making a prediction. Crucially, the most important feature was the change in the CAT score.

Key strengths of the approach include the duration of data collection, the large number of patients and the large number of exacerbations captured. This is one of the largest datasets to date, implying that the predictive models generated in this study are more generalisable to the wider population. Furthermore, generating models from an already established digital therapeutic creates the potential to deploy an exacerbation support tool more widely providing a greater impact on patient care. Many studies monitor sample sizes for 3-6 months and as a result, only capture a handful of exacerbations [53, 91]. This study's data set spans 4 years, with 506 participants and 1,263 exacerbations captured.

The main limitation of this study is the lower-than-desired sensitivity of the Adaboost with CST. Moreover, because of the large number of stable state records, even with good specificity, the PPV is low. A low PPV implies that a considerable number of positive predictions made by the model are false positives. Clinically, this means that a substantial portion of patients identified by the model as at risk for exacerbation may not actually experience one. This can lead to unnecessary interventions, increased healthcare costs, and potential patient anxiety. Comparatively, a handful

of papers in the literature demonstrate very high predictive performance such as Patel *et al.*'s COPDPredict™ [74]. The system identifies COPD exacerbations at a median of 7 days before a clinician-defined episode. The researchers found that COPDPredict™ has a sensitivity of 97.9%, a specificity of 84.0%, a PPV of 38.4% and an NPV of 99.8%. Moreover, the researchers identified that the alerts generated by the COPDPredict™ system (providing early warning to both patient and clinician) reduced hospitalisations by 98% (90 vs 2, p<0.001) compared to the previous 6 months before the study. While the observed reduction of COPD exacerbations by 98% in the 6 months following the implementation of COPDPredict™ is noteworthy, it's crucial to consider the possibility of regression to the mean. The initial high exacerbation rates in the previous 6 months may have been an anomaly, and the subsequent decrease could be a natural consequence of the statistical tendency for extreme values to normalise over time. Additional analyses, such as control groups or statistical tests accounting for potential confounding factors, would be essential to ascertain the true impact of COPDPredict™ on exacerbation rates. Additionally, the study was composed of patients with severe COPD who are frequent exacerbators. This means the patient group is less generalisable to the wider COPD population.

In the study by Patel *et al.* participants were required to use COPDPredict™ every day for 6 months [74]. This involved uploading self-assessments daily using the COPDPredict™ app, measuring FEV1 using connected spirometers and recording C-reactive protein using finger-prick testing. It is yet to be demonstrated that people with COPD will track their health consistently for so long. Nonetheless, this study underscores the possibility of integrating sensing technologies to augment existing models and enhance predictive performance, although the clinical utility of PPV for an exacerbation remains modest at 38.4%. The modest PPV has implications for overtreatment with steroids and antibiotics. The consequences of this include the development of antibiotic resistance, potential side effects such as disruptions to the microbiome and systemic impacts, increased healthcare costs, and the risk of diminishing the efficacy of future treatments. Furthermore, patients may experience heightened anxiety and a reduced quality of life due to unnecessary medication burden. The strength of the approach set out in this analysis is the minimal patient burden; therefore, the focus of merging sensing technologies will be to continue the burdenless approach.

The reliance on the myCOPD app for symptom recording in COPD exacerbation prediction introduces a potential bias, as events occurring during periods of non-engagement with the app go unrecorded. This intermittent data collection creates gaps in the dataset, limiting the model's understanding of a participant's health during those unmonitored periods. Such a limitation not only reduces the model's sensitivity to early indicators but also questions the generalisability of findings to real-world scenarios where consistent digital health monitoring

4.4. DISCUSSION AND CONCLUSIONS

may not occur. Future model iterations could address this by incorporating alternative data sources and improving participant engagement strategies to ensure a more comprehensive and continuous view of an individual's health status.

Importantly, the visualisations from the t-SNE suggest that the performance of the model will likely be limited. In Figure 4.11, it is clear there are many overlapping points between the exacerbation and stable classes with no clear structure cluster representing a single class. Similarly, when t-SNE was applied to the predictions of the AdaBoost with CST model in Figure 4.12, there is no clear pattern emerging. This suggests for any further significant improvements to model performance new features will have to be developed to apply to the model.

This study demonstrates the potential exacerbation predictive value of patient-entered data in a real-world digital therapeutic. While the accuracy of the models is still relatively low, the negative predictive value is high indicating potential utility in supporting patients when to avoid rescue packs unnecessarily. Patients can benefit from reliable guidance, reducing the risk of unnecessary medications and associated costs. This not only boosts patient confidence but also minimises potential health risks, ultimately contributing to improved health outcomes. However, it is unclear how acceptable these accuracies might be to patient users but it seems clear further work is required to improve positive predictive performance. This could be through incorporating sensing technologies in co-design partnership with patients followed by further clinical studies to validate a digital early exacerbation warning system which results in improved clinical outcomes in COPD. The following chapters detail the inception of these co-design partnerships, where we collaborate with myCOPD patient users to assess the acceptability of exacerbation models, warnings, and enhanced monitoring using remote sensing devices.

Chapter 5

Sensing, Predictions, and Alerts in COPD Exacerbations Study I

A preliminary analysis using machine learning demonstrated that real-world data from a widely deployed digital therapeutic (myCOPD) has some predictive capabilities for AECOPD. This study aims to gather information on myCOPD patient-user perceptions and attitudes towards the exacerbation prediction models (EPM) previously developed and their deployment in the future with regard to an exacerbation warning system (EWS). Two focus groups were conducted online via Microsoft Teams with myCOPD patient users and were analysed thematically. The six myCOPD users were recruited through the myCOPD app service. Patient preferences were explored in two focus groups and 5 themes were defined. These themes include: with further improvement the exacerbation prediction models will be acceptable, the exacerbation prediction models aren't useful in the real world, seeking understanding and trust in exacerbation prediction models, an exacerbation prediction would trigger preparation and caution, and insights and desires in the context of an exacerbation warning. Participants expressed greater acceptance of the EPM from the AdaBoost with CST algorithm, with a specificity of 89% and sensitivity of 35%. There was disregard for the EPM based on the EasyEnsemble Classifier. Participants explained they would build trust in the EPM through trial and error but would quickly lose that trust with inaccurate predictions. For EWS, participants expressed a want for the language to be an "indication of" or "suspected" exacerbation and also to be able to choose from different types of warning. They also identified the warning would give them time to prepare before an exacerbation but expressed concerns about potential support or lack thereof after a warning. They also had mixed feelings about the intensity of the warning with one participant feeling concerned over the response to a red alert whereas others felt this was necessary. While participants were very supportive of the work for the SPACE studies, they expressed dissatisfaction with the current performance of EPM. However, they felt the models didn't need to be much better until they would begin using/trusting them. Participants also communicated their satisfaction with some preliminary EWS designs requesting the ability to have access to several options with a percentage risk of certainty of prediction.

5.1 Introduction

After completing the work from Chapter 4, it was clear that the accuracy of the AECOPD prediction models was still low. Before proceeding to further develop the performance of EPM it was determined that the next stages of this work should involve patient and public involvement (PPI) to better inform the design of an exacerbation prediction system built into the myCOPD app.

This chapter focuses on the first study of the SPACE “Sensing, Prediction, and Alerts in COPD Exacerbations” studies. The overarching aim of the SPACE studies is to develop a system that combines machine learning with myCOPD data to create a decision support tool for myCOPD users. This system would inform users when they are at risk of an AECOPD and can guide their decision-making with regard to treatment. This chapter presents SPACE Study I which aims to discuss previous modelling work undertaken (Chapter 4) with users of the myCOPD app to determine what they perceive as a valuable level of accuracy needed from an automated exacerbation prediction tool and what a good exacerbation warning might look like.

myCOPD app users were invited to participate in our focus groups. The focus group took place online via Microsoft Teams in February 2023. In the focus groups, we explored how the myCOPD users might respond to exacerbations typically and what are the possible consequences (positive or negative) of different prediction models e.g., over, or under-estimating risk of exacerbation. We also asked users to identify what actions they might take if they received a warning that they were at risk of an exacerbation and if there were any negative or positive implications related to using an automated exacerbation warning system.

Here is a more detailed breakdown of the activities:

1. Focus groups were around 1 hour and were scheduled in accordance with the participant’s needs.
2. Participants were in groups of 4-6 with 3 researchers.
3. During the focus groups we learnt more about the use of myCOPD, exacerbation management, opinions on exacerbation prediction and intervention, what kind of warning might illicit a suitable response, and perspectives on sensors.
4. Visualisations were used to explain how prediction works and the accuracies of those predictions.

5. Balsamiq (a user interface design tool that allows informal visual representation of software interface) was used to display different designs of exacerbation alerts.

The participant information was held under the Data Protection Act, and was stored in a confidential form and later anonymised. Audio recordings of the sessions were also kept confidential and securely stored on one of the University of Bristol servers, and only I had access to them. All data was completely anonymised and was not attributable to any participant.

5.2 Methods

This study received approval from the Faculty of Engineering Research Ethics Committee (Ref: 12398). All participants provided informed written consent before the focus group took place.

5.2.1 Participants

Participants were patient users of the myCOPD self-management app with a diagnosis of COPD. Participants were recruited via a mailing list from the myCOPD database. The research nurses at my mhealth proactively reached out to participants by sending invitation emails to those who met our criteria. Participants were not excluded; instead, they were deliberately chosen based on specific criteria, employing a purposeful convenience sampling approach. The selection was focused on individuals with a demonstrated high level of engagement in the myCOPD app and a history of exacerbations. This method allowed for targeted participant sampling, aligning with the research objectives and enhancing the relevance of the study cohort. In this hybrid approach, we strategically selected participants based on specific criteria (purposeful sampling) while also taking advantage of the convenience and accessibility of patient users who engage with the myCOPD app (convenience sampling). Patients in our sample who expressed interest in participating were guided to an online form. This form included the participant information sheet and fields where they could input their email address, phone number, and their preferred method of contact to register their interest. Subsequently, I reached out to them and provided an online consent form along with a questionnaire. Once they consented and completed the questionnaire, we scheduled an interview.

5.2.2 Focus groups

Focus groups were conducted virtually using Microsoft Teams software. Interviews took place between February and March 2023 and lasted roughly 1 hour. All focus groups were conducted by 3 researchers with Glyde leading the sessions. The focus groups were supported by Beth Cliffe and Caitlin Morgan. At this time Glyde was a doctoral student in engineering, Cliffe was

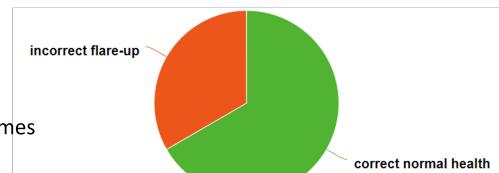
CHAPTER 5. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS STUDY I

a postdoctoral researcher in psychology with considerable qualitative experience, and Morgan was an airways clinical research fellow. Glyde has previous training and research experience conducting focus groups, Cliffe has published research in which she conducted focus groups and interviews, and Morgan has a PG Cert in Medical Education and has been trained for interviews. The interview schedules featured open-ended questions, accompanied by prompts designed to stimulate participants to share comprehensive responses and engage in discussions with each other. The full interview schedules for the separate focus groups are available in Appendix File C. The first focus group schedule explored patient experiences with myCOPD, how exacerbations are managed generally, perspectives on the accuracy of current EPM, the usefulness and usability of the information provided by EPM, and their attitude towards the use of sensors to improve EPM. Representations of the two models developed in the previous model work shown to participants are in Figures 5.1 and 5.2, model 1 being the EasyEnsemble Classifier with a sensitivity of 67% and a specificity of 69.5% and model 2 being the AdaBoost with CST with a sensitivity of 35% and a specificity of 89%.

MODEL 1

When normal health:

- Correctly identifies normal health 2 in 3 times
- Wrongly suggests you're at risk of a flare-up 1 in 3 times



When at risk of a flare-up :

- Correctly identifies flare-up 2 in 3 times
- Wrongly suggests normal health 1 in 3 times

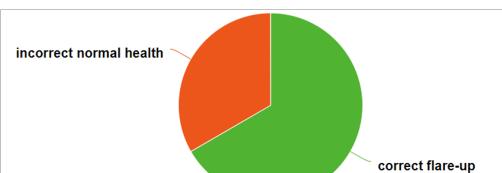
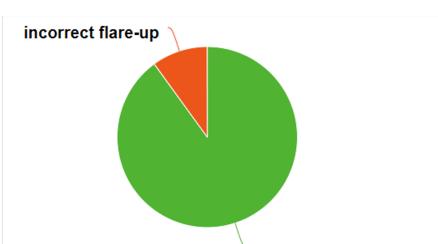


Figure 5.1: Statistical representation of model 1.

MODEL 2

When normal health:

- Correctly identifies normal health 9 in 10 times
- Incorrectly identifies at risk of a flare up 1 in 10 times



When at risk of a flare up :

- Correctly identifies flare up 1 in 3 times
- Incorrectly identifies normal health 2 in 3 times

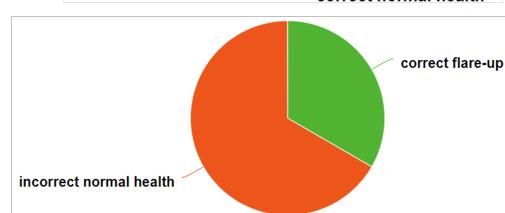


Figure 5.2: Statistical representation of model 2.

5.2. METHODS

The second focus group also delved into experiences related to myCOPD and the management of exacerbations. In this focus group, we investigated the participants' perceptions regarding a warning of an impending exacerbation, the communication of and language used in the warning, and the potential impact such warnings could have on their well-being. Different designs of EWS were used to develop their opinions further. These designs can be seen in Figures 5.3, 5.4, and 5.5. The focus groups were recorded using the built-in functionality of Microsoft Teams and then transcribed for analysis.

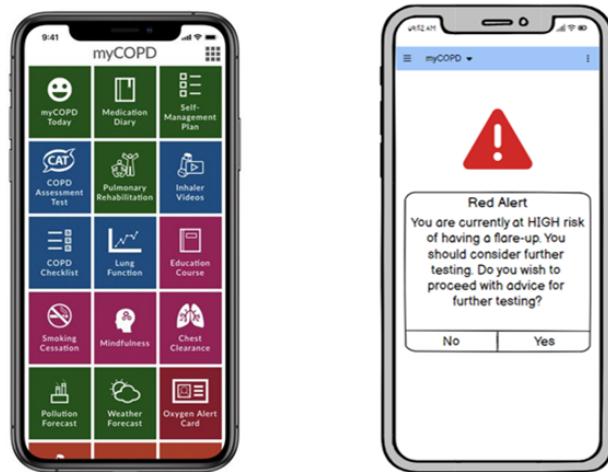


Figure 5.3: Design of a passive alert.

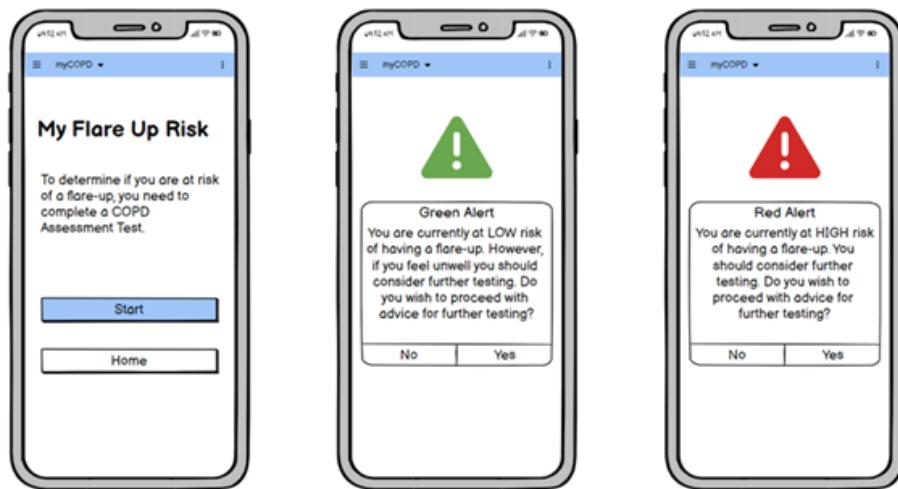


Figure 5.4: Design of an active alert.

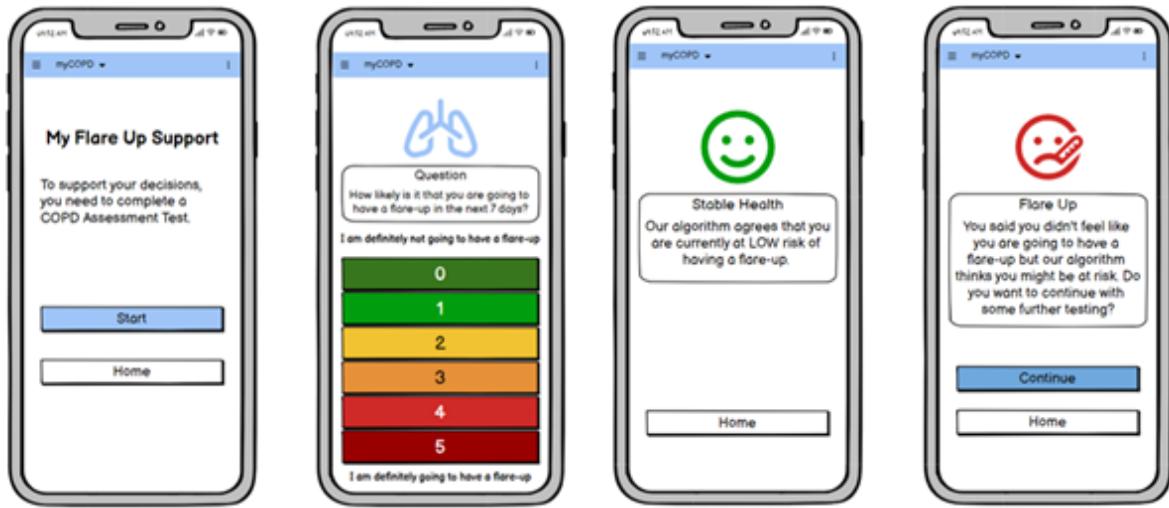


Figure 5.5: Design of a support tool.

5.2.3 Thematic analysis

We conducted an inductive and reflexive thematic analysis, following the six phases outlined by Braun and Clarke [120]. These phases encompassed: phase 1 - familiarising yourself with your data, phase 2 - generating initial codes, phase 3 - searching for themes, phase 4 - reviewing themes, phase 5 - defining and naming themes, and phase 6 - producing the report. I began by immersing myself in the data, a process that included manually transcribing it. I meticulously cross-checked these transcriptions against the audio files to ensure accuracy and then conducted proofreading to finalise the transcripts. Codes were generated from the data using NVivo 12, and these codes were organised into potential themes. This was an iterative process, allowing for the refinement of codes and themes until they accurately represented the transcribed data. Following this, I created a thematic map, which served as a framework for reviewing the emerging themes. I revisited the transcripts to ensure that they captured the nuanced meanings present in the data. Subsequently, I assigned names to the themes and provided a concise overview of each theme. Importantly, all stages of this thematic analysis were carried out independently by me.

It is important to understand how my experience with the research on the project to date will influence my interpretations and shape the positionality of this analysis. I had previously worked on developing exacerbation prediction models through analysis of myCOPD data. Subsequently, I was instrumental in the development of the research plan for this study and led the focus groups. Moreover, at this time I had planned the subsequent SPACE Study II and have further aims to continue the SPACE Studies into an interventional study. The intimacy I have with the work in this project will have strongly influenced my interpretation of the data and would likely not be shared amongst those who have not had the same role as myself.

5.3 Results

22 myCOPD users expressed interest in the study, with 14 providing consent. Subsequently, 12 participants accepted invitations to the focus groups, and a total of six myCOPD patient users actively participated in both sessions. Four attended the first and four attended the second with some crossover. myCOPD patient users were all >50 years, white and were mostly male (4/6, 67%). They all stated they used the myCOPD app daily. Four participants stated they were moderately skilled in digital literacy (I can function and participate fully in a digital world and can use technology) apart from two who felt they had no skill in digital literacy (I can not function in the digital world or use technology). They all had experienced AECOPD at a range of severities (rescue pack use, hospitalisation, frequent exacerbations).

EPM and EWS were explored in the two focus groups and 5 key themes were defined: (1) With further improvement the exacerbation prediction models will be acceptable (2) The exacerbation prediction models aren't useful in the real world (3) Seeking understanding and trust in exacerbation prediction models. (4) Insights and desires in the context of an exacerbation warning. (5) An exacerbation prediction would trigger preparation and caution. The themes and their sub-themes are outlined below with quotes. See Figure 5.6 for a thematic map.

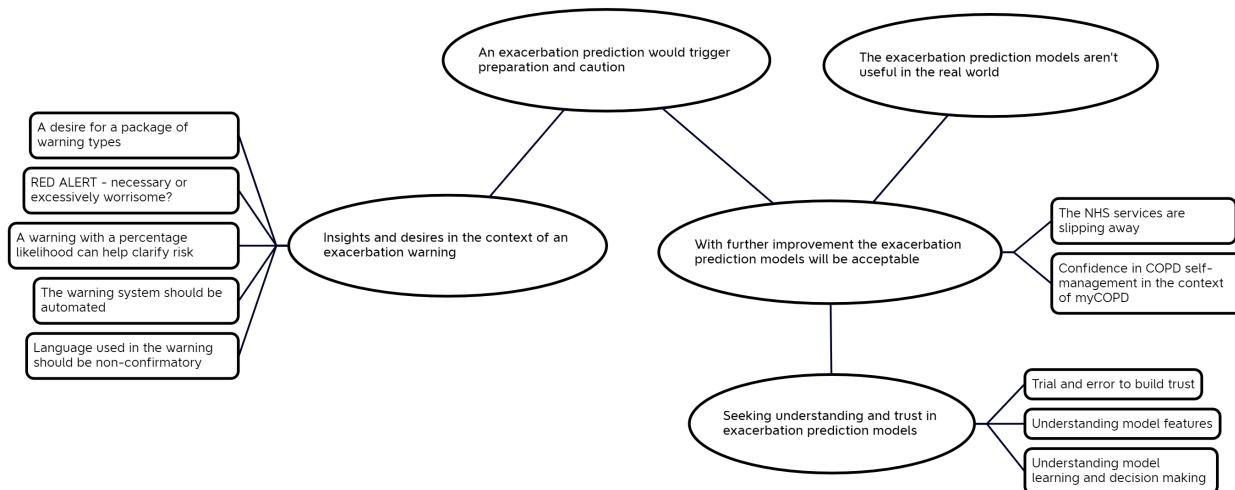


Figure 5.6: Thematic map from the focus groups.

5.3.1 Theme 1: With further improvement the exacerbation prediction models will be acceptable

When participants expressed an acceptance of the EPM they spoke in a future tense that one day, with further improvement, there will be acceptance. This acceptance focused more on model 2.

All participants were supportive of the development of EPM:

“I would be 100% behind developing something, which is what you’re trying to do along the lines of what we’re talking about today. Because I think it would be invaluable if we can get it to where we actually need it to be” - *participant 1*

Participants spoke with regard to acceptance in the future tense. The EPM will be accepted once it reaches a requisite standard of accuracy that the patient-user deems acceptable to them. There was a greater acceptance of model 2 and as a result, when participants discussed future acceptance they focused more on this model and gave their insight:

“What I would want to know is, am I at risk of a flare-up? That would be the bit that I would be interested in. And in this instance, we’re saying it correctly identifies a third, one in three times. So we’re back to the 33% again. The bit that’s gonna be really useful to me.” - *participant 1*

Because model 2 had near-perfect specificity (89%) they often remarked on focusing work on sensitivity (35%) and bringing it up to a similar standard to reach acceptability:

“The good health was satisfactory because I’m not worried about that, but the one in three would be the one that I’d focus on. That would be the one that I want to improve to the same sort of level as good health, I want nine out of 10.” - *participant 2*

When discussing the acceptability participants remarked that model 2 was much more acceptable:

“That looks a lot better yea, results are significantly improved.” - *participant 3*

Despite the consensus that work should focus on improving sensitivity, the participants felt that the increased specificity, identifying good health 9 in 10 times was a significant improvement in comparison to model 1:

“The second one is for me, far more accurate. As *participant 3* quite rightly says, which correctly identifies normal health 9 in 10 times.” - *participant 1*

5.3.1.1 Sub-theme: The NHS services are slipping away

Healthcare support concerns were a recurring theme in both focus groups. In the first group, discussions revealed a divide: some participants expressed frustration with their interactions with healthcare practitioners, while others harboured concerns about the ongoing reduction of

5.3. RESULTS

their standard care. In contrast, the second group primarily voiced worries about the availability of support following a warning. Frustration with care often stemmed from perceptions of receiving subpar healthcare:

“I’ve got a real problem with my GP because they just keep prescribing amoxicillin and it fails every time.” - *participant 2*

This frustration would revolve particularly around a feeling of being unheard or ignored. They would ask their GP or healthcare practitioner for a certain treatment or care and they felt their concerns weren’t heard.

With regards to fear of losing care, one participant described the NHS slipping away:

“In my experience, the NHS service that provides the cover for us as we’re trying to progress through this situation has been slipping away bit by bit year after year.” - *participant 1*

A consensus emerged within the group, highlighting a noticeable decline in the quality of care received in the past few years since the onset of COVID-19. Numerous services and routine checkups, once commonplace, were suspended during the pandemic and have yet to resume, fueling suspicions that they might never return. This discussion predominantly revolves around the perceived lack of support, which seems to be a compelling factor in determining the acceptability of future models. As healthcare services continue to be scaled back, patients are seeking innovative approaches to self-manage their health. The availability of an exacerbation decision support tool could potentially reduce their reliance on the NHS, fostering greater independence and improved symptom control.

In the second focus group, lack of support was an issue but this time the concern focused on the lack of support in place after a warning. There was a concern about being able to access services after a warning and also the fear it might incite in individuals who can’t access support or intervention:

“I would need to know because you know within about four or five hours I’d be calling an ambulance and in London this is impossible. You can’t get hold of a doctor. And you can’t get hold of ambulances because there’s everybody wanting to get to the hospital and the last thing I want to do is get this and be in A&E at the moment.” - *participant 6*

The participants’ apprehensions about post-exacerbation support carry substantial weight and must be thoughtfully addressed when contemplating the type of assistance the myCOPD

app can offer patients following an exacerbation warning. This consideration is of paramount importance because an exacerbation warning without adequate support has the potential to induce significant anxiety among isolated COPD patients:

“But I think there’s a very valid point about scaring people because I think we’ve all mentioned that we live very isolated at times. And if you’re isolated, you can be conscious of a lack of support.” - *participant 1*

5.3.1.2 Sub-theme: Confidence in COPD self-management in the context of myCOPD

The participants reported using the myCOPD app very similarly to each other. By far the most common use of the myCOPD app among participants was for medication tracking. Participants used educational videos to support their COPD such as the pulmonary rehabilitation course. Another common use was for weather, pollution and pollen tracking. The app was also used to track weight and log exercise/physical activity.

All study participants were queried about their use of the myCOPD app for exacerbation management. The consensus among them was negative, with most stating that they did not use the app to support them during an AECOPD:

“As for managing flare-ups themselves; no, I don’t use the app in managing them. My wife has got COPD. We have a ready supply of steroids, etcetera, etcetera. So I don’t need to refer to anything.” - *participant 1*

Whilst preparedness is a reason for not using the myCOPD app there was a perception that there was nothing of use that could support them during exacerbations:

“But I don’t use the app to manage exacerbations because I don’t really think it would make any difference, frankly.” - *participant 2*

Conversely, one patient expressed a high level of confidence in their COPD management, believing that their extensive experience with the condition rendered them largely self-sufficient and minimally reliant on external references or resources.:

“I’ve been doing this since 2005, you know and so I know what my body’s doing.” - *participant 2*

However, in the development of an AECOPD intervention, it will be crucial to incorporate

5.3. RESULTS

valuable tools within the myCOPD app that patients can readily access when facing an exacerbation. This is especially vital because there is a unanimous agreement among patients that the app currently lacks features they find beneficial or relevant to their needs during an AECOPD.

The only aspect currently available to patient users that might have a potential connection to exacerbation management is the utilisation of the CAT to monitor their symptoms over an extended period and assess the effects that exacerbations have had on their overall health:

“I also record my CAT score every day. Which I find very interesting because that shows exacerbations and everything else because it goes up and down like a yoyo, you know.” - *participant 2*

“I can look back particularly to see when I last had one and I can also monitor my general well-being on that as well, that just gives me an indication of how things are going and long term.” - *participant 3*

In terms of exacerbation, the use of rescue packs is common to everyone, taking a course of antibiotics and steroids when they believe they are exacerbating. However, much of management is exacerbation prevention. This prevention involves avoiding going out and close contact with people to prevent catching a respiratory tract infection:

“Well I haven’t had an exacerbation for some time, probably because I’m very careful where I go now. I don’t go out unnecessarily.” - *participant 5*

“I like some of you, keep a very low profile. I remain at home as often as I can, I only go out for medical appointments or dentists or something like that.” - *participant 6*

The approaches participants take to avoid exacerbations have led to a feeling of loneliness with one participant remarking that COPD itself is a lonely disease:

“And of course, what happens a few weeks after we’ve got back together? I get an exacerbation again. So it just burns out. COPD is a very lonely life, isn’t it?” - *participant 2*

While participants currently rely on the app to assist in managing their COPD, it seems they handle exacerbations independently and are confident in their self-management abilities. Nevertheless, incorporating guidance for AECOPD within the app may lead to enhancements in their self-management by providing access to tools they currently lack.

5.3.2 Theme 2: The exacerbation prediction models aren't useful in the real world

Whilst there was some indication of acceptance (in the future), both models at present-day were unacceptable, with a shared understanding that they wouldn't be useful in the real world. This perception of impracticality was focused much more on model 1. Model 1 was disregarded with participants expressing concern when discussing the specificity of model 1 and the likelihood of false positives:

"We talked earlier about the accuracy of the prediction and how it would affect you. Yeah, how the inaccuracy of it would affect you. And I'll mention the fact that I don't wanna be worried unnecessarily. If it's predicting something, then it's not actually sure it's going to happen." - *participant 1*

Participant 4 summarised the general perspective of model 1:

"OK, well, when I first saw the two models together, I wasn't very enamoured with model one." - *participant 4*

The sensitivity and specificity of model 1, which were characterised by a 2 in 3 chance of making accurate predictions, were met with significant scepticism and disregard. It became immediately evident that there was strong disapproval of this model due to its perceived inadequacies in terms of accuracy.

"Well, I wouldn't be very... I don't think the accuracy is particularly stunning." - *participant 3*

The participants often interpreted the results of the model in relation to use-ability in the real world and felt a margin of error of one-third was excessively substantial for real-world applications.

"If you're gonna get something wrong, thirty-three to an even larger percentage of the time, it's not going to be much use in the real world because people are not gonna want those odds." - *participant 1*

The accuracies raised concerns about the prudence of making decisions or taking actions relying on such a significant margin of error, thereby limiting its practical utility in their everyday lives:

“I’d want it to be a lot more accurate if I was going to take any serious action about it.” - *participant 1*

Whilst a low sensitivity of model 1 means it is likely to miss exacerbations, participants interpreted this information in a way that the prediction of AECOPD would likely be wrong (two-thirds of AECOPD predictions are false positives):

“Well, I wouldn’t want to start taking antibiotics and steroids, based on a one in three.” - *participant 2*

The mistrust may have been the way sensitivity was presented, as a 1 in 3 chance of correctly predicting AECOPD. Participants may have interpreted this as 2 out of 3 predictions are false positives rather than 2 out of 3 AECOPD are not predicted. nevertheless, the participant’s response still suggested a rejection of model 2 because the low accuracy prevents trust in a prediction. I assumed the participants had understood the initial description of the prediction models and on reflection, I believe some of the information may not have been fully conveyed or correctly interpreted.

5.3.3 Theme 3: Seeking understanding and trust in exacerbation prediction models

During the initial focus group, the participants sought a deeper comprehension of the exacerbation prediction models. Through this enhanced understanding, they delineated the steps required to establish trust in these models within the context of managing their COPD.

5.3.3.1 Sub-theme: Understanding model features

The participants would frequently ask for the features used in the model to be repeated:

“What did you get the scores from? Explain again.” - *participant 1*

Given ample time for contemplation, participants independently formulated their own interpretations, gradually developing an understanding of the elements involved in constructing a model that facilitates accurate predictions.

“So the CAT scores, in theory, are a historical thing, aren’t they? So it’s what? It’s more really what happened yesterday.” - *participant 2*

5.3.3.2 Sub-theme: Understanding model learning and decision making

Participants were eager to delve deeper into their understanding of the exacerbation prediction models, posing inquiries about the mechanisms by which these models acquire knowledge and make determinations regarding what constitutes an exacerbation and what doesn't:

“I was wondering whether these algorithms learn as time goes by and it could improve?” - *participant 3*

For decision making they would ask more intricate and complex questions. One participant explained their symptoms before exacerbations aren't always consistent so was unsure how this would result in accurate predictions:

“How can you correctly say that we're gonna have a flare-up when sometimes our mucus doesn't always change colour? So, you can't predict when we're gonna have a flare-up?” - *participant 4*

Despite having this knowledge of inconsistencies in symptoms before exacerbations, there was uncertainty about why AECOPD prediction wasn't 100% accurate:

“But what I would do is I'd wonder why it wasn't correctly anticipating the events fully when you're nearly there in Model 2.” - *participant 1*

One participant tried to rationalise the lack of perfection by interpreting the algorithm as having inbuilt parameters that prevent 100% accuracy:

“But I'm sure the algorithms have got some sort of level of uncertainty built into them so that could be why we're not getting 100%.” - *participant 3*

5.3.3.3 Sub-theme: Trial and error to build trust

Participants described how they would come to trust an EPM (through trial and error) and also explained their current thoughts on trusting models 1 and 2. The consensus among participants was that an EPM, once deployed, wouldn't automatically garner trust. Instead, the model would have to prove its trustworthiness through accurate predictions:

“If it did prove to be an exacerbation over a number of events, then obviously you'd start to believe it, wouldn't you?” - *participant 2*

One participant suggested this trust-building would occur through trial and error:

“Trial and error would be something that might indicate whether it was working well.” - *participant 3*

The participants’ discussion of building trust through trial and error indicates they will not respond to an early warning without previous evidence of correct prediction. It follows that, from the patient’s perspective, inaccurate predictions would likely cause a loss of trust in EPM.

When participants described putting trust in the models this was closely related to their perception of the accuracy:

“It’s all about having confidence, isn’t it? Confidence in the prediction. And if it’s nine out of 10, there is a chance that you’ll believe it and go with it.” - *participant 2*

When referring to model 2 specifically, there was more trust but it still lacked enough accuracy to fully trust its use in the real world:

“The results are significantly improved. So yeah, I trust it a lot more.” - *participant 3*

“I wonder why it doesn’t correctly predict 100% of the time because you’re getting 90%. What’s that? What? What’s that 10% that’s going wrong, that’s causing it not to umm not correctly evaluate completely? So I think for me, although it’s much better, I wouldn’t trust it.” - *participant 1*

So, even before beginning a trial and error process to build trust, the accuracy has to be suitably high enough to engage with the EPM. This trust-building process makes logical sense. Whilst we understand how a model works in a closed system making retrospective predictions, it is unclear how effective such a system would be in the hands of patients.

5.3.4 Theme 4: Insights and desires the context of an exacerbation warning

The Participants were uncertain about what the warning should look like or say but they did provide insight into some key characteristics that should make up a warning.

5.3.4.1 Sub-theme: Language used in the warning should be non-confirmatory

When participants described the language that should be used in the warning design they preferred it to be non-confirmatory:

“So it must be able to give you a suspected warning of something going wrong with your lungs.”
- *participant 2*

“Something, anything that just says over a gadget that says you may be having an exacerbation, you may expect a flare up.” - *participant 5*

“There is an indication that there is a risk, something like that. There is an indication that you’re at a higher risk than usual.” - *participant 1*

I interpret this non-confirmation as coming from two places the first being an understanding that the prediction may be wrong. Thus, to the user, it makes more sense that the language used is not definitive but potential - the likelihood is higher but it may not happen. The other origin of non-confirmatory language comes from a place of concern. A definitive exacerbation warning has the potential to cause concern, language focusing on an increased probability could be less stressful:

“Something that’s a little less? Oh my God, you’re going to go into some sort of problem right in this moment, you know?” - *participant 1*

I believe the preference for non-confirmatory language is likely attributed to potential model inaccuracies, driven by the belief that EPM will likely never attain perfect accuracy. It appears relevant that participants’ awareness of these imperfections influenced their language preference for warnings.

One participant suggested that to alleviate such concern the language currently being used “alert” should be shifted to “warning” as this feels less worrisome:

“Umm, I just like to hop back to the previous discussion briefly on instead of alert, I would much rather prefer warning. You know high warning or level, you know, high level warnings. You know, some action is necessary or you must take action. I think that’s far less scary, but maybe that’s coming from a military background cause alerts got a very specific connotation to it in that respect. And so yeah, probably a warning might be for me better.” - *participant 1*

5.3.4.2 Sub-theme: The warning system should be automated

The exploration of design functioning as automation was examined comprehensively from various perspectives. The initial aspect of this automation involves the automatic generation of warnings, providing alerts without requiring human intervention or manual control:

“It would be best if it was an automated text message but it would have to be very clever, wouldn’t it?” - *participant 2*

The participants desired the automated warning system to convey its operational status and its capacity to alert individuals about an exacerbation even when not using the myCOPD app:

“I think that the indication should be something that makes you aware of it without you having to switch something on to check out if it’s aware.” - *participant 1*

“Well, it will be great to have a real signal, wouldn’t it? From somewhere that doesn’t depend on you turning the COPD app on. Because the COPD app, you only use for 10 minutes tops in the morning. It’s turned off the rest of the day, so it’s not a lot of use to give a warning, is it?”
- *participant 2*

This indicates that it is preferable for patient users to have background monitoring with automated warnings instead of active engagement with a monitoring system. A participant mentioned that it is important for such a warning to be repeated until it is recognised:

“You should be able to switch it off once you’ve got it, but it keeps coming back until you do switch that particular alert off. So if you’re asleep and miss it. Yeah, it keeps coming at you, you know, so you have a repetitive sort of process where you go. Ohh yeah, I’ve gotta look. I’ll press that button there. Yeah, so that’s all I’d say about it.” - *participant 1*

Participants also expressed a eagerness to have the automated warning appear as a wearable alert or potentially a vibration:

“It’s on the phone and I’ve got alert set on that, one little buzz and you’ll feel it. You don’t have to hear it, you’ll feel it. I feel it. Buzz on my wrist and I’m looking, ohh, that’s what it’s telling me, you know? So I think something like that would be ideal.” - *participant 1*

“I will respond to a buzz. So if something could be passed through the watch rather than through an app on the screen. I think that would be my first sign and then I would immediately go to

the oximeter and monitor it.” - *participant 6*

“I would expect it to be something on a gadget, something that would just warn me in general terms.” - *participant 5*

5.3.4.3 Sub-theme: A warning with a percentage likelihood can help clarify risk

Having a warning communicated as a percentage risk was important to the participants:

“I think numbers give an exact representation of where you’re at, so rather than saying you’re at high alert, what does that mean? But if it tells me I’m at 80%, I know exactly what that means.” - *participant 5*

“I’m very much a numbers percentage man. I want to know if there’s a 60% problem, 70%, 80% something like that.” - *participant 1*

Participant 1: “That’s the trouble with interpretation as well. If you say high, people will interpret that high differently to other people. I think I might be 95. Somebody else might think it’s 75, you know. But if you get the numbers, there’s no, there’s no argument about what, what the situation is. So, I’m very much behind the numbers.”

Participant 2: “Well, I agree with all that. That’s good logic.”

From what the participants suggested, incorporating a percentage likelihood of exacerbation risk would enable better self-management decision-making with regard to taking rescue medication than a verbal warning alone.

5.3.4.4 Sub-theme: A desire for a package of warning types

The three designs received a favorable response, leading to intriguing discussions about the preferred warning types and the significance of passive/automated monitoring. The consensus within the group was to amalgamate all three designs (active, passive, and supportive), providing patient users with the flexibility to select the warning type that suits their preferences:

“I liked them. I thought they were all relevant. You could use all of them. I don’t know whether you expect us to pick and choose some of them, but I thought they were all relevant and you could have run them all in a sequence.” - *participant 2*

“Yeah, you know what? I’d agree with *participant 2*. I think the whole group, all of those things you’ve shown us together would be fantastic to have as a sort of package that we can access if we want to or even have it done automatically if that’s what we like to do.” - *participant 1*

One participant felt strongly the support tool design shown in Figure 5.5 is very similar to the myCOPD opening screen where they report their symptoms through four options. This participant suggested because of this similarity, the applicability of the support tool warning design may be redundant:

“Well, I just, when you turn myCOPD on, it asks you how you are, doesn’t it? And it gives you four options. You’ve already got in theory the self-driven alert. It’s already there. The only trouble is it doesn’t do anything, does it? So it doesn’t notify anybody, and nobody’s gonna help you. So what? It is already there in theory.” - *participant 2*

5.3.4.5 Sub-theme: RED ALERT - necessary or excessively worrisome?

The concept of a “red alert” was thoroughly examined as a significant topic among the participants. The core of the matter revolved around whether such an intense warning was essential to prompt an appropriate response to an impending exacerbation, or if it might lead to unnecessary distress. A particular participant expressed concerns that a “red alert” could potentially induce panic in individuals.

“However, I think that’s an area you should take extreme caution with because to alarm somebody unnecessarily when you’re saying you might have something in your 7 or 8 days time, I wouldn’t want something alert, alert, red, red. You know you. You’ve gotta get sorted, cause I would, you know, that would just be panic stages.” - *participant 1*

Others felt that the greater the severity of the warning might be needed to get them to take action:

“Opposite to *participant 1*, I quite like the shock. When you suddenly got a text message that comes up red and says if you don’t press this button, you’re gonna die in 10 minutes. I quite like that, you know. So, we’re all different, aren’t we?” - *participant 2*

Another participant agreed that such an alert was needed to ensure they respond correctly and gain access to suitable services:

“Well, I wouldn’t be frightened by reading a red alert because honestly, I’ve got to such a stage

in this COPD that if it's going to make me think and look at my accelerometer all day and you know go up the stairs for example, just to test my breathlessness or start doing some exercises I would need to know because you know within about four or five hours I'd be calling an ambulance and in London, this is impossible. You can't get hold of a doctor." - *participant 6*

5.3.5 Theme 5: An exacerbation prediction would trigger preparation and caution

During the initial focus group discussion on how participants would react to an AECOPD prediction, they examined the issue from two perspectives: an emotional one, considering how the warning would make them feel, whether it would evoke concern or leave them unfazed; and a practical one, pondering what actions they would take based on this information.

One of the participants felt like receiving the information they were going to exacerbate could make them concerned:

"I think it would be a little bit of a worry." - *participant 1*

The others had a different perspective believing that they probably wouldn't be too worried:

"I don't think I'd be really worried about the warning, that really wouldn't worry me too much."
- *participant 3*

But this might have been due to interpretations about the length of the prediction horizon and having time to prepare and care for oneself prior to symptomatic decline:

"It's not a worrying thing because if it's all 10 days ahead, you've got 10 days to worry about this flare-up that's coming down the line." - *participant 3*

This is a significant concern, as existing EPMs are not inherently equipped for temporal predictions, and it is highly unlikely that they will ever attain the requisite level of accuracy needed for such predictions. Therefore, deployment of current models in real-world scenarios will predict an AECOPD occurrence anywhere from the next day to 8 days time, prompting the need for rescue medication at that point in time. It should be reinforced to myCOPD patient users that the exacerbation model detects if an individual has entered the prodromal phase of an exacerbation, marked by a decline in respiratory symptoms. If the model accurately identifies these early warning signs, it's crucial for patients to take immediate action and intervene to harness any potential benefits that early intervention can provide. There must be

5.3. RESULTS

careful consideration in the future when translating the workings of an EPM to ensure this is understood.

Participants conveyed that upon receiving a prediction of an AECOPD, they would take steps to prepare, such as ensuring they had a rescue pack readily available or obtaining one if needed. Additionally, they expressed their intent to exercise greater caution in their daily activities and take better care of themselves.

“I can make sure that I have actually got my rescue pack handy because it’s sometimes, you know, you can use it and forget to order another one.” - *participant 3*

“I would take care of how I’m living my life in a way it’s supposed to be a bit grandiose about it, but I’d be careful about not going out in cold and such like conditions.” - *participant 1*

In the second focus group, the reaction to a warning centred very much around caution, preparation and the specifics the patient users would do after a warning to manage their health. Preparation involved ensuring they had rescue packs/enough medicines to manage their health and other coping tools they may use in an AECOPD:

“It would help me prepare. It would help me get my coping tools in line.” - *participant 5*

Caution is about taking care of yourself, avoiding contact with people, and avoiding going out in potentially harmful conditions such as high pollen, high pollution, or especially cold days:

“Well I might be careful or what I might do, what I might be limited in doing and how I might behave.” - *participant 5*

Part of this caution involves a heightened vigilance using sensors to ensure the physiological variables didn’t drop to dangerous levels especially relating to pulse oximetry:

“I think that would be my first and then I would immediately go to the oximeter and monitor it, and if it’s below what it ought to be, then I know I’ve got a problem.” - *participant 6*

Participants also discussed how they would medically treat themselves in response to a warning by taking the medications they would normally take during exacerbations such as their rescue pack consisting of antibiotics and corticosteroids.

“Well, I also have a nebuliser and that would probably be the first thing to start if I was feeling

particularly breathless, I think I would use the nebuliser, you know three times a day. I've got plenty of Prednisolone here as part of my rescue pack and to see if that opens the airways and makes one feel less breathless." - *participant 6*

5.4 Discussion and conclusions

This study sought to understand myCOPD users' opinions on the accuracy of recently developed EPMs and provide insight into the design of an EWS. The primary conclusion drawn from this study is that myCOPD users are unlikely to express interest in utilising the current EPMs. This reluctance stems from the perception that the current accuracies are not deemed practically useful in the real world, reflecting participants' understanding of the technology's limitations. An essential point of concern was the chance of an incorrect prediction was too high for them to want to take their rescue pack (antibiotics and oral corticosteroids). Despite this standpoint, the participants of the first focus group made it clear that the EPM could become acceptable with improvement in accuracy. Participants expressed greater acceptance of the EPM from the AdaBoost with the CST algorithm, with a specificity of 89% and sensitivity of 35%. Participants' perceptions regarding the described accuracies led them to believe that 2 out of 3 predicted exacerbations would be incorrect. This misunderstanding likely stems from misinterpreting the sensitivity rate (35%) as 1 in 3 predicted exacerbations is accurate (2 out of 3 are false positives). In reality, the EPM is more likely to miss exacerbations (2 out of 3 AECOPD will be predicted as stable health), with only a 1 in 10 chance of a false positive. In such cases, it is crucial to emphasise that an exacerbation prediction should be taken seriously and acted upon, as the likelihood of an incorrect prediction is much lower. This confusion underscores the importance of providing clear explanations to myCOPD users regarding how the EPM's accuracies translate to real-world scenarios, aiming to prevent misinterpretations and enhance trust in the EPM.

While it is evident that enhancing accuracy is essential to establish trust and promote the initial adoption and utilisation of an EPM, it is also apparent that trust would still need to be built throughout use. It was explained that participants would build trust in the EPM through trial and error. Observing over time correct predictions is a necessary step in the adoption of an EPM. But this brings into question what happens with incorrect predictions. It is unclear how many incorrect predictions it would be until a user would lose trust. Suzuki *et al.* conducted a study aiming to measure the connections between patients' or their family members' experiences of misdiagnosis and the level of trust they have in their clinician. [121]. Among 661 patients, 23.2% had a personal history of misdiagnosis. In a multivariable-adjusted general linear model, patients' misdiagnosis experiences were associated with lower confidence in their current physician (mean difference -4.3 , 95% CI -8.1 to -0.49). This was negatively correlated

5.4. DISCUSSION AND CONCLUSIONS

with patients' sceptical attitude toward medical care (Spearman's correlation coefficient -0.205 $p = <0.001$). Whilst prediction is not diagnosis this evidence highlights incorrect predictions have the potential to cause discouragement of use due to loss of trust in EPM.

For EWS, the participants had thoughts and opinions on the language of a warning, automating the system and incorporating percentages to convey risk. With regard to language, the participants described the information presented in a warning as non-confirmatory so an "indication of" or "suspected" exacerbation. Participants emphasising such language convey their awareness of the inherent impossibility of achieving 100% accuracy in predictions, acknowledging the existence of an inevitable margin of error. The term "suspect" denotes forming an impression of something's presence without absolute confirmation. The participants' recommendation to employ such language may be associated with the process of trial and error in establishing trust; utilising non-definitive language can assist in tempering expectations regarding the precision of exacerbation predictions. Moving forward, the language within EWS should underscore a potential escalation in risk rather than providing a conclusive confirmation of an impending exacerbation.

The participants were keen to have the EWS automated. The participants describe an EWS that operates automatically; the system does not require their active engagement and will continuously monitor their inputs and flag a warning external to the myCOPD app if changes in symptoms are prodromal to an exacerbation. Key insights on this automatic design were described by patients to include an awareness that the system is running, repeated notifications that require acknowledgement when an impending exacerbation has been detected, and a buzzing or vibrating on a connected wearable or smartphone to alert the user. For awareness that the system is running, this could be in the form of notifications to inform the user the system is continuously checking for early signs of an exacerbation. Alternatively, there could be an exacerbation tile built into the myCOPD app that allows users to turn on and off exacerbation tracking which indicates to the user that they're currently being monitored. The repeated notifications could be connected to the buzzing with a warning similar to an alarm that has to be opened and turned off to stop the warning. Once turned off the system could provide information about the exacerbation and signposting to what they should do next.

Percentage risk was determined to be a better way to convey risk than a verbal warning. One participant explained that users will likely interpret a verbal warning of risk differently than if the warning could be conveyed as a percentage likelihood. A study by Kunneman, Stiggelbout, and Pieterse agrees with this concept [122]. They worked to assess the numerical probabilities that individuals associate with frequently used verbal labels relating to treatment outcomes. They used 11 verbal labels (almost never, rarely, a small chance, very occasionally,

sometimes, occasionally, maybe, regularly, often, a fairly high chance, and most) extracted from 90 audio-taped decision encounters in oncology. Three hundred participants were asked to assign numerical probabilities or percentage likelihood to the labels in the context of nausea or cancer recurrence. There was considerable variation in how individuals interpreted verbal labels. The participants' estimates were significantly lower in the context of cancer recurrence than in nausea which suggests a tendency to underestimate a likelihood in the context of increasing severity of the outcome. The participants with lower subjective numeracy [123] differentiated less between labels which led to the authors concluding verbal labels for this may be more difficult to interpret or less helpful [122]. This study did not include individuals who had been recently diagnosed with cancer, the specific context for which the verbal labels were originally designed. As a result, there is a possibility that people not diagnosed with cancer may articulate their interpretations of the verbal labels differently. Similarly, in the context of COPD exacerbations, people who are at risk from an exacerbation may interpret "a fairly high chance" more consistently and with less variation than seen in this study. Additionally, it's important to recognize that individuals with COPD can face substantial cognitive impairment, as elucidated by Dodd, Getov, and Jones [124]. This cognitive impairment may subsequently affect their ability to interpret verbal risk effectively. Consequently, it is advisable for EWS to integrate a percentage-based risk or likelihood of exacerbation to assist myCOPD users in making informed decisions.

Upon presentation of three separate designs of EWS, there was a positive reception with participants requesting a package of options to be able to choose from different types of warnings. The designs differed by function; the first is a passive alert - it monitors the user in the background alerting them when detecting signs of an impending exacerbation, the second is an active alert - the user must engage with the system and conduct some tests, resulting in an answer about their exacerbation risk, and the third is a decision support tool - the user inputs how they are currently feeling and conducts some tests to be able to receive either confirmation on their suspected health status or potentially disagreement with alternative information. Whilst it was requested to have access to all three options, the participants seemed to prefer the passive alert. The passive alert shares many similarities with the participants' concept of automation. The earlier discussion of automation combined with a preference for the passive design suggests that participants are keen to incorporate a system that doesn't require their input and will monitor them in the background. One participant pointed out the passive alert shared similarities with the introductory symptom scoring screen asking "how are you feeling today?". Their perspective was the introductory screen was already a self-driven alert that didn't do anything. However, the rest of the participants felt they wanted all options to choose from. It is worth noting this participant has had COPD for the longest and stated that because of this he is very aware of his health:

5.4. DISCUSSION AND CONCLUSIONS

"This is only because I've been doing this since 2005, you know. And so I know what my body's doing." - *participant 2*

It may be that he did not find the support tool useful for his own management as he is more aware or more confident in his decision-making when he feels like he is at risk of having an exacerbation. The support tool design would be more useful to someone who has a greater uncertainty regarding how they are feeling and if they need to react to changing health. Laue, Melbye, and Risør conducted semi-structured interviews with a group of 19 COPD patients to investigate the decision-making process behind initiating antibiotic and/or oral corticosteroid (rescue pack) treatment during an AECOPD [125]. The researchers analysed the collected data using a thematic approach. They determined that patients were proficient through practical experience which enabled them to recognise and swiftly respond to worsening symptoms. However, the potential adverse effects of a rescue pack had a significant effect on their decision-making to start treatment which could result in hesitation. The patients preferred support from a healthcare practitioner when they reached the limit of their individual self-assessment ability. From the findings of this paper and the conclusions drawn from the thematic analysis, it is necessary to continue with the development of an EWS that operates in part as a support tool to assist in decision-making in the face of uncertainty. This is also a reminder that not all of the EWS options will be relevant or useable to everyone and the ability to choose what the user has access to can be emphasised.

In their study on early therapy in COPD exacerbations, Wilkinson *et al.* reported, "The median (interquartile range) treatment delay (time from the onset of exacerbation to the initiation of treatment) was 3.69 (2.00 to 5.57) days." This implies that the reference point for predicting exacerbations is within 3 days after the onset. However, it's essential to note that this study focused on exacerbations treated by physicians, introducing a potential delay associated with scheduling appointments. Given the importance of understanding how patients self-identify for exacerbations, a dedicated study is warranted to establish a benchmark for future research in exacerbation prediction, particularly in the context of patient self-management.

The participants had mixed feelings about the intensity of the warning with one participant feeling concerned over the response to a RED ALERT whereas others felt this was necessary. This is a complicated issue and can't be fully answered within a section of this study. The participant who did not like the concept of a RED ALERT explained that such a message could cause excessive worry and had the potential to ruin someone's day. Alternatively, the other participants suggested that the more severe the warning the better as this is needed to kick them into action. Unfortunately, there is currently a gap in the literature regarding the severity of a warning and responses. However, HD Hadjistavropoulos, Craig, and T Hadjistavropoulos explored the

CHAPTER 5. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS STUDY I

cognitive and behavioural reactions to health-related information in a sample of 192 patients with varying degrees of subclinical health anxiety [126]. The participants received feedback on diagnostic assessments, and this feedback was randomly categorised as either indicating a positive, negative, or uncertain risk for health complications. The authors emphasise that individuals with higher levels of health anxiety are likely to experience more pronounced negative reactions and tend to catastrophise the implications of medical information. Additionally, they discovered that such individuals might struggle to initiate or complete necessary medical procedures, a phenomenon extending to individuals with even mild forms of health anxiety. While most participants in our study emphasised the importance of a more severe warning for exacerbations, there is a potential concern that such a heightened warning could lead to substantial distress and, in some cases, hinder an appropriate response to the warning, including the timely initiation of rescue medication. The appropriate severity level for an exacerbation warning remains unclear and warrants further investigation across individuals with varying degrees of health anxiety. Moreover, it is essential to identify if a milder warning can still elicit an appropriate response to encourage engagement in treatment.

Despite uncertainty regarding the severity of a warning, all the participants described effective and suitable responses to an exacerbation prediction or warning. These responses included intervention with their rescue medicine, enhanced monitoring, and increased precautionary measures such as avoiding harsh conditions and contact with potential pathogens, and in general taking care of themselves. These responses show promise that if exacerbation prediction can be offered to myCOPD users they will use the system in the way it was intended to be used - to improve self-management.

The participants across both focus groups shared the view that the NHS services they had been receiving have been declining. Participants reported that routine check-ups that had been dissolved during the COVID-19 pandemic to prevent the spread had not since returned. There was talk of frustration with their healthcare practitioners and emergency services and a real concern that there would not be support in the face of an exacerbation. The theme regarding the NHS services was not part of a direct question but was explored continuously in almost all the questions. As highlighted by Laue, Melbye, and Risør, COPD patients will often seek support from healthcare practitioners in the face of uncertainty regarding exacerbation intervention [125]. Access to an AI-enabled exacerbation prediction could potentially mitigate the need for clinician-based advice. I suspect the theme of the NHS services slipping away was explored as participants potentially see exacerbation prediction as alternative support in a time when the support is very much needed.

This PhD work used two qualitative data analysis approaches: inductive thematic analysis

and reflexive thematic analysis. Inductive thematic analysis involves a data-driven exploration of qualitative material. Researchers carefully examine the data to find patterns, recurring themes, and underlying meanings that naturally emerge. It prioritises participants' voices and allows themes to develop without predefined categories. This approach differs from deductive analysis, where predefined themes are applied to the data to see how well it aligns with those pre-established concepts. Inductive analysis explores themes, codes, and categories that arise from the data itself. Since there's limited research on deploying exacerbation prediction models, inductive analysis was more appropriate. It doesn't impose preconceived themes and allows themes and codes to be defined naturally from the data.

Reflexive thematic analysis adds a layer of researcher reflection to the process. It acknowledges that researchers bring their own biases and perspectives, which can influence the analysis. Researchers using this approach critically examine their own preconceptions and biases, leading to a deeper understanding of how their viewpoint shapes the interpretation of themes. This approach enhances the analysis's rigour by promoting transparency and self-awareness, contributing to a more nuanced interpretation of the data. It was especially relevant in this case because I'm closely involved in the entire project, from developing prediction models to planning their safety, efficacy, and deployment. Considering my own perspectives and potential biases was essential.

5.4.1 Limitations

This study may have been limited by its size and lack of diversity. There were 6 participants who were predominantly white males meaning these findings may not represent the experiences, thoughts and ideas of other myCOPD users. Whilst the largest group of people with COPD are white, a significant number of people from other ethnicities make up the population [127]. Their perspectives are essential to the development of this intervention to improve its suitability for people from other backgrounds. Moreover, this is a small study undertaken with a small group of myCOPD users and so the implications for an exacerbation prediction system may not represent the wider myCOPD population. Acknowledging that this study involved a small, predominantly white male group of 6 participants, it's important to recognise that these initial qualitative studies often start with limited samples to assess initial concepts. These findings, while not comprehensive, provide a foundation for future research with more diverse populations, considering the broader spectrum of COPD patients, which is necessary for intervention development. While the sample's demographics may not fully represent the diversity within the myCOPD user population, this study serves as an initial exploration. Future research can actively seek to include a more diverse participant pool to better capture the varied experiences and needs of different ethnicities, ensuring that the intervention is suitable

for a wider range of users.

A substantial limitation, initially overlooked but now apparent upon reflection, concerns how the models were presented to patients. The method of conveying performance results across different models in this study was unlikely to facilitate a genuine understanding of their potential impact in a clinical context. Specifically, the communication of a very low PPV was not sufficiently clear, and no data were presented to indicate that participants comprehended that out of 20 instances where the model signalled a risk of exacerbation, only 1 of these predictions would be accurate. This lack of clarity constitutes a significant limitation. In an earlier discussion, I noted that the false positive rate was stated as 1 in 10, a statistic that holds true but may not effectively convey the more impactful measure of the very low positive predictive value from both clinical and patient perspectives. Future research with myCOPD patient users should focus on conveying the PPV of the models and translating this to a meaningful context in terms of how it would affect their self-management.

5.4.2 Conclusions

Whilst the study group may be small and lack diversity there was a clear consensus amongst the participants that the current accuracy of EPM was unsuitable for use. However, they felt it could be acceptable to them in the near future with improvement in accuracy. As identified in the dual systematic literature review in Chapter 2, sensors have the ability to enhance exacerbation prediction as the current models only use patient-reported outcomes. Conducting a comparable study with a more extensive and diverse patient cohort, coupled with a clearer explanation of PPV, is evident. While some participants emphasised sensitivity in our modelling approach, it may prove essential to present a range of models featuring different sensitivities and specificities. This approach aims to elucidate an optimal balance that aligns with the perspectives and preferences of patients. Nevertheless, It follows that the incorporation of additional sensing measures is likely the direction to take to garner acceptability. However, as discussed in Chapter 2 in the section on patient factors adoption of sensors into daily life is unclear. Next, in Chapter 6, I describe SPACE Study II, determining the feasibility and acceptability of integrating sensing technologies with myCOPD users.

Chapter 6

Sensing, Predictions, and Alerts in COPD Exacerbations Study II

Findings from SPACE Study I indicated that the current predictive ability of machine learning EPM is not sufficient for trust and/or use in the real world. However, it is possible to use additional sensing technologies alongside the myCOPD app to enhance prediction. This study aims to gather information on myCOPD patient-user perceptions and attitudes towards the use of sensors in conjunction with the myCOPD app and perform a preliminary analysis to determine the usability of the data for improving the performance of EPM. Five myCOPD users were recruited through the myCOPD app service. Initially, a focus group was conducted online via Microsoft Teams with myCOPD patient users. Subsequently, a smartwatch, a pulse oximeter and a digital spirometer were sent to the five participants. At study initiation 5-days of daily interviews were conducted. A midpoint interview was conducted after two weeks, and a final interview was conducted as study completion (4 weeks). The focus group and interviews were analysed thematically. Sensor and myCOPD patient user data were extracted at study completion. The themes that were defined from the analysis of the data from the focus group and interviews include: seamless incorporation of daily sensing, accuracy is important to adherence, perceived inaccuracy and usability issues may hinder digital spirometry adherence, and adoption of increasing physical activity and calorie tracking. The key motivators for adherence to daily use include the establishment of routine, using sensor feedback to inform health management, using sensors to support weight loss and physical activity behaviour change, and the potential to have access to exacerbation prediction. The barriers to adherence include health anxiety and poor results, difficulty conducting tests (especially relating to the spirometer), and perception of sensor test results as inaccurate. Analysis of sensor and myCOPD data indicates near-daily use with usable data. Whilst the participant group of this study is homogenous and is known to engage with technologies, the findings suggest that myCOPD patient users will likely engage in sensor use daily with minimal disruption. Moreover, the data from the sensors is usable in conjunction with myCOPD data and so can be used to identify if the sensors can improve the performance of the current EPM.

6.1 Introduction

As previously discussed in Chapter 5, myCOPD users are willing to engage in the daily use of sensor technologies, if it means they have access to exacerbation prediction models. Sensors allow the collection of physiological and functional data remotely and daily which could enable the enhancement of current EPM. However, the successful implementation and adoption of such sensor-based systems heavily depend on patient acceptance and adherence to daily usage. Understanding patient perspectives, attitudes, and experiences regarding the use of these sensors is crucial to determine their feasibility and potential for widespread implementation. Additionally, investigating factors influencing patients' motivation (and barriers to adherence) to engage in daily sensor usage will allow for the development of strategies that optimise acceptance and adherence.

This proposed mixed-method study aimed to explore the acceptance and feasibility of daily sensor deployment in the patient users of the myCOPD app, with a focus on their willingness to utilise these sensors regularly. By integrating qualitative interviews and quantitative data analysis, this study sought to achieve a comprehensive understanding of patients' perceptions, concerns, and expectations surrounding the use of sensor technology for AECOPD prediction.

The qualitative component of the study involved 5 days of daily interviews a midpoint interview and a final interview after 4 weeks of sensor use. This allowed the development of evidence into their thoughts, beliefs, and experiences related to daily sensor usage. Thematic analysis of these interviews aimed to highlight the factors influencing patients' acceptance/motivation, barriers to adherence/potential discouraging factors for daily sensor use. Through understanding these factors, an intervention can be designed that increases the motivation to engage with sensor use whilst minimising barriers to adherence or factors that increase discouragement. The quantitative aspect involved analysing data collected from the deployed sensors, including physiological parameters and daily activity patterns. The implemented sensor suite comprises a smartwatch, a pulse oximeter, and a smart spirometer. The selection of the pulse oximeter and spirometer was based on their widespread use in existing literature. Additionally, the smartwatch was chosen for its capability to measure sleep variables intricately associated with respiratory diseases, including sleep respiratory rate and sleep SpO₂.

By assessing patients' actual engagement with the sensor technology, we can evaluate their compliance with daily usage protocols, identify patterns of usage, and potentially explore correlations between sensor utilisation and worsening health/symptoms that might suggest a sensor's potential to improve the predictive performance of an EPM.

The findings from this mixed-method study will provide evidence to guide the future deployment of sensor-based systems for COPD management and exacerbation prediction. Understanding patients' perspectives and behaviours will inform the development of patient-centred interventions, improve adherence, and ultimately enhance the potential of these technologies to transform COPD care.

The participant information, myCOPD data, and sensor data were held under the Data Protection Act, and were stored in a confidential form and later anonymised. Audio recordings of the sessions were also kept confidential and securely stored on one of the University of Bristol servers, and only I had access to them. All data was completely anonymised and was not attributable to any Participant.

6.2 Methods

This study received approval from the Faculty of Engineering Research Ethics Committee (Ref: 12403). All participants provided informed written consent before the focus group took place.

6.2.1 Participants

Participants were patient users of the myCOPD self-management app with a diagnosis of COPD. Participants were recruited via a mailing list from the myCOPD database. Participants were sampled by purposeful convenience sampling and were selected based on high engagement in the myCOPD app with a history of exacerbations. In this hybrid approach, we strategically select participants based on specific criteria (purposeful sampling) while also taking advantage of the convenience and accessibility of patient users who engage with the myCOPD app (convenience sampling). patient users from this sample who were interested in participating were directed to an online form that contained the participant information sheet and space to enter their email address and phone number and how they preferred to be contacted to register interest. I then contacted them and sent an online consent form and a questionnaire. Upon consenting and completion of the questionnaire, the focus group was arranged. Study inclusion criteria: Over 18 years old and confirmed COPD. Five participants (100% male) met inclusion criteria and provided informed consent.

6.2.2 Procedure

The study was an uncontrolled, prospective observational feasibility study with all participants offered three sensors (smartwatch, pulse oximeter, smart spirometer) to use. A focus group was conducted prior to the study initiation. Interviews were conducted every day for the first 5

days, and then two more interviews were conducted; one at the study midpoint (2 weeks) and one at study completion (4 weeks).

The focus group and interviews were conducted by Glyde. The focus group discussion lasted 1 hour and took place over Microsoft Teams. The focus groups were conducted by 2 researchers with Glyde leading the session with Beth Cliffe supporting it. At this time Glyde was a doctoral student in engineering and Cliffe was a postdoctoral researcher in psychology with considerable qualitative experience. Glyde has previous training and research experience conducting focus groups, Cliffe has published research in which she conducted focus groups and interviews. A clinician also was present during the focus group to explain how to use the digital spirometer. The interviews were conducted separately and by Glyde alone.

The interview schedules featured open-ended questions designed to prompt participants to share detailed responses and engage in group discussions. The full interview schedules are available in Appendix File D. The focus group explored patient perspectives and experiences with sensors generally. The interviews asked more targeted questions about why they have or have not used their sensors, any issues that have arisen, and motivations for continued or discontinuation of use. The focus group and interviews were digitally audio-recorded and then transcribed to prepare them for thematic analysis.

Sensor data was extracted at study completion for analysis. One participant did not use the wrist-worn wearable due to the use of the same wearable in a separate trial. One participant dropped out of the study after two weeks due to the need for treatment for a separate health condition.

6.2.2.1 Sensor technologies

The sensor technologies comprised the Fitbit Inspire 3, a Bluetooth-enabled pulse oximeter, and the Portable Digital Spirometer - MIR Smart One. The Fitbit Inspire 3 is a health and fitness wearable/smartwatch that contains a 3-axis accelerometer, an optical heart rate monitor, a vibration motor, red and infrared sensors for SpO₂ monitoring and an ambient light sensor. These sensors within the wearable monitor and record: steps, sedentary minutes, heart rate (HR), resting heart rate (RHR), heart rate variability (HRV), SpO₂, skin temperature (ST), and respiratory rate (RR). The wearable provides a mean for these physiological variables during sleep. The Fitbit Inspire 3 is worn on the user's non-dominant wrist and is connected to the Fitbit: Health & Fitness app via Bluetooth. The Bluetooth-enabled pulse oximeter uses red and infrared sensors to monitor and record SpO₂ when the device is attached to the user's finger and is connected to the smartphone app via Bluetooth. The MIR Smart One spirometer

uses a turbine sensor to measure PEF and FEV₁. The spirometer is Bluetooth-connected to the MIR SMART ONE app to monitor and record the measures.

Participants were provided with a crib sheet and contacted via the phone by Glyde to support the set-up and instruct them on how to carry out and record measurements. The participants were asked to use the sensors every day and to contact Glyde if they had any issues.

6.2.3 Data analyses

Sensor acceptability, motivation, barriers to adherence, and behaviour change were explored qualitatively using thematic analysis [120]. The analysis of both the focus group and individual interviews was conducted simultaneously to enhance the depth of comprehension. Inductive thematic analysis was used to explore themes in the data. All stages of the thematic analysis were undertaken independently. Codes were generated from the data using NVivo 12 and then organised into themes. Themes relating to the motivations and barriers of sensor adoption are reported below, with quotes to exemplify key insights.

Sensor use rate patterns were examined quantitatively. The potential for sensor measures to improve the performance of EPM is explored in visualisations incorporating myCOPD patient users' symptom scores and the values from the sensor readings.

6.3 Results

A total of 5 myCOPD patient users attended the focus group and used the sensors for 4 weeks apart from one who dropped out after two weeks due to a non-COPD-related health issue. myCOPD patient users were all >40 years old, white and were all male. Three of the participants (2, 3, and 4) participated in SPACE Study I. Participant 2 was participant 1, participant 3 was participant 2, and participant 4 was participant 3. Participants 2,3, and 4 used the myCOPD app daily, participant 5 used the app irregularly and participant 1 had not activated their account. They all stated they were moderately skilled in digital literacy (I can function and participate fully in a digital world and can use technology). They all had experienced AECOPD at a range of severities (rescue pack use, hospitalisation, frequent exacerbations).

In the focus group and interviews, we explored the perspectives on the use of sensors and how they affect myCOPD users in real-time and 4 key themes were defined: (1) Seamless incorporation of daily sensing (2) Accuracy is important to adherence, (3) Perceived inaccuracy and usability issues may hinder digital spirometry adherence (4) Adoption of increasing physical activity and calorie tracking. The themes and their sub-themes are outlined below with quotes.

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS
STUDY II

See Figure 6.1 for a thematic map.

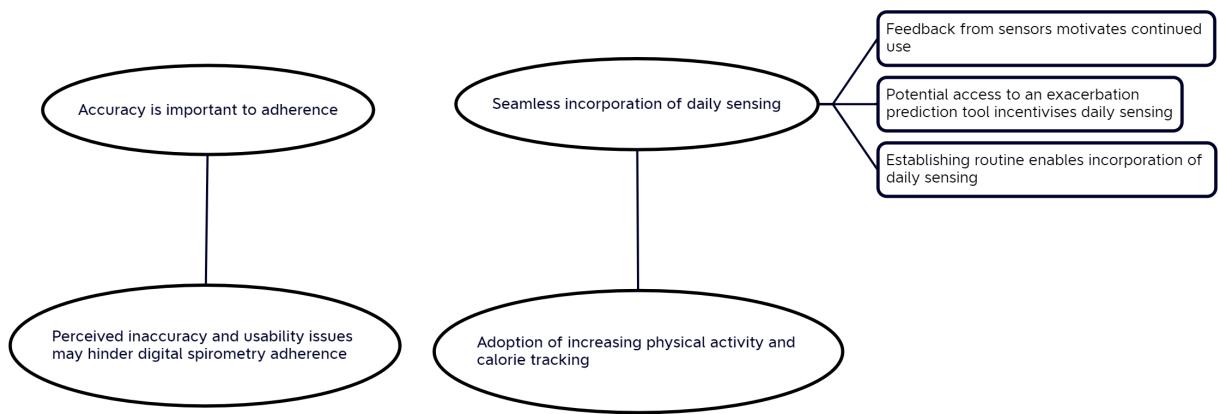


Figure 6.1: Thematic map from the focus group and interviews.

6.3.1 Theme 1: Seamless incorporation of daily sensing

Overall, participants demonstrated a willingness to engage in the use of sensing technologies. many were even excited about the prospect of gaining further insight into their health and management of COPD. During the focus group, participants were asked “Would you use a sensor if you knew it could improve an early flare-up warning?”. The consensus of participants was in strong support:

“Yes, of course I would. Yeah, because yeah, yes, definitely.” - *participant 1*

“Absolutely. I already use two of them.” - *participant 3*

“Ohh yeah, 100% 100%.” - *participant 5*

“I would use any sensor that would improve the chances of identifying another flare-up, yeah.” - *participant 4*

Due to the anticipation of access to an exacerbation prediction tool, when participants were queried about their willingness to incorporate the sensors into their daily routines, a similar enthusiastic response was observed:

“I've got no, no objections to using them at all. I'll use them on a daily basis. It is. It is. So I'm fine with all of it.” - *participant 2*

“Uh, I think I'd use them all daily.” - *participant 3*

“I just don't see that being a problem really because you know, I could have it. So, it's because I worked from this desk where we are quite a bit. I could have them probably up there and I'd probably just use it, just probably use it every now and again. There just wouldn't be an issue with it.” - *participant 1*

Throughout the focus group participants shared their own insights and outlook on using the sensors. A key insight was regarding the Fitbit and ease of use, in that it does not require much effort and can even be used during working hours:

“At this moment in time I'm probably better off with the Fitbit because I'm working and I'm driving, and it would be a lot more... how can I put it... easier for me to use than it would the other two. The other two wouldn't have an issue at home, but the other two wouldn't

be... how can I put it... as easy to use when I'm at work? Because I am basically all day." - *participant 5*

While discussions revolved around the convenience of Fitbit and wrist-worn wearables, the greatest interest was consistently directed towards the spirometer, primarily due to its potential to offer valuable insights into the participant's current health status:

"I think the one that I'll probably use a lot is the puffer one that 'blow blow blow' one because I'm always interested to see what it would be." - *participant 1*

"Yeah, I particularly like the respiratory one, the one for checking airflow and that would be really good." - *participant 4*

A thought-provoking perspective emerged regarding the absence of clear instructions for sensor usage and the crucial role that such guidance will play in the future. They felt there was a lack of instruction on how to use sensors to support their disease self-management and that if there were guidelines in place recommending their usage then the frequency of use would be much higher:

"I would look at it as part of a clinical trial and if that clinical trial had instructions, then I was to use this every day while that was ongoing. Then that's exactly what I would do. I'm not being facetious, I mean, you can lay the guidelines down for when I would want to use them, and I'd stick to those guidelines." - *participant 2*

6.3.1.1 Sub-theme: Establishing routine enables incorporation of daily sensing

Participants offered valuable insights into their motivations for sustaining daily engagement with sensory activities. This insight fell into three sub-themes: Routine, Sensor information, and exacerbation prediction. These three sub-themes are effectively the reasons to continue using sensors. Whilst, routine is not necessarily a motivator or incentive, sticking to an established routine was a key reason that participants continued to use the sensors. Participants described creating a new routine which incorporated sensing, or adding them to an already established routine:

"I do it as part of a routine, so I will do it when I first wake up in the morning and then later on in the day after I've had my tea around 7." - *participant 5*

"I use them every day, I will definitely use them tomorrow. I'm scheduled to use them; at certain

times of the day, I'll go through a process of using all the sensors." - *participant 4*

"I have a set time that I am going to do them and I'm always interested in what they're going to say." - *participant 2*

"I plan to continue using the sensors until we next speak it's easy now it's part of a routine. I do them at the same time every day between 10 to 11 and I plan to continue that until we next speak." - *participant 3*

One participant established a set of alarms to remind them to carry out their sensing:

"Normally I have alarms that go off to remind me like in the evening. I'm sitting here watching the TV or working on the computer, and the alarm will go off on my phone and I'll check it, Ohh, time to take them so I don't even have to think about them. It's great." - *participant 2*

To reinforce the argument for their routines as a motivating factor, participants frequently cited instances where disruptions, such as illness or appointments, hindered them from carrying out their sensory activities as usual:

"Now I'm starting to feel better I'm gonna get back to my routine tomorrow." - *participant 5*

"I forgot about them. Normally I do them in a schedule, I wake up do my exercise, do my sensors, and have a shower. But this morning I was, if you remember, I said I got to go to the doctor for appointments. So I go upstairs, got dressed, I got changed into my going out stuff and I never thought about it until a couple of minutes ago when I was downstairs." - *participant 2*

"I got called into work and I didn't have the time. I felt knackered and I wasn't feeling great so I just went to bed." - *participant 5*

"I didn't use the sensors this morning because I was feeling pretty unwell. I couldn't face it. But I got a bit better this afternoon, so I used them." - *participant 5*

6.3.1.2 Sub-theme: Feedback from sensors motivates continued use

The participants found the feedback provided by the sensors to be useful to them, especially regarding their management of COPD. They would often describe being interested to see what the sensor says tomorrow as a motivator:

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS
STUDY II

“I’m really interested to see what they have to say, like with the pulse oximeter I did one 96-97 then later on in the day it says 91-92. Not sure why that is. Maybe it’s because I had the inhaler? Tomorrow I’m gonna try one before the inhaler and one after the inhaler.” - *participant 1*

This motivation runs deeper than just interest. The motivators behind the sensor information lie in informing health management, tracking long-term changes, informing behaviour change, and affirming positive health. For informing health management the participants felt that staying on top of their health was made possible by the sensor information and thus were motivated to continue that through this insight:

“I noticed my oxygen levels were a lot higher than what they were yesterday when I looked at it this morning. Then come back down again when I took it about an hour later. So I want to keep track of it.” - *participant 1*

“But I find it useful to see what my oxygen and pulse and other things are.” - *participant 2*

“I need to look at my heart rate and my pulmonary function.” - *participant 4*

Participants also felt that being able to see these changes over the long term might help them:

“I’ve got my own oximeter which I’ve used in the past, but it’s nice to have this one that’s connected to an app on the phone because I’ve got historical data there without having to write it down. You know what I mean? And keep a log. Yeah. Then you can see how things are changing over time like that.” - *participant 4*

“But what I do find interesting is I do remember my results in a percentage format being 62% and I can see on this one that more results are of 57%. So that tells me in the last two years there’s been a change there.” - *participant 2*

Importantly, behaviour change in itself was a motivator to use the sensors as they were using the information from the sensors to inform the behaviour change:

“The sensors are particularly important with weight, steps, and calorie intake. I’m trying to hit 10,000 steps a day but I need that to be recorded.” - *participant 2*

“Pre-Christmas I was getting continual exacerbations and haven’t exacerbated recently. But I think it’s because I’ve been losing weight and getting healthier. I started this before the sensors

6.3. RESULTS

but using these sensors I have been able to make better targets." - *participant 2*

Affirming positive health was an interesting motivator. One participant described how when you're feeling well, getting that confirmation from a sensor is a major drive to not only continue using the sensors but to engage in greater activity that could benefit health:

"So, you know like today it was nice to see the oxygen levels were up and your puff was better as well. You know there's motivation to do that to look at that if you're feeling good. Because of those readings, I was thinking I'm gonna take advantage of this and do some more exercise." - *participant 1*

"I like to look at the pulmonary function I think that's interesting. I was digging up some of my records from past spirometry tests and you know I'm surprised at how well-maintained my COPD has been." - *participant 4*

The participants universally expressed a heightened sense of awareness regarding their COPD due to the insights gleaned from the sensor data:

"Oh yea I am definitely thinking about my health and COPD a lot more." - *participant 1*

"Using the sensors has made me a bit of an expert. I am more aware of my condition." - *participant 4*

"I am just thinking about my COPD a bit more and trying to be a bit healthier." - *participant 5*

One participant mentioned that they had not altered their behaviour during the study because they had already made necessary changes beforehand. However, they did indicate that the study had prompted them to contemplate their COPD and its management more deeply:

"No, I've been doing everything the same. But I'm going to be more aware or have a greater depth of exactly how things are going." - *participant 2*

"I was fairly cognisant of my COPD situation. I've used the sensors for further investigation. When dealing with COPD, the app fills a lot of my requirements. With this study, I looked at it as a chance to explore other areas and maybe look into issues I have. With the spirometer I've noticed my scores are low, so I want to know why. I've been making notes about my exercise and feeling unwell and trying to relate those results to the test and that sort of experimentation. The Fitbit has also been quite interesting so I compare it with the Garmin and feel more

assured. But I really like using the Fitbit wearable for tracking calories and meals.” - *participant 2*

6.3.1.3 Sub-theme: Potential access to an exacerbation prediction tool incentivises daily sensing

The purpose of this feasibility study was to determine the acceptability and usability of these sensors for exacerbation prediction. The participants were well aware of the ultimate goal. Not only were they aware they were hopeful of one day having access to this tool. So, when they were providing reasons for continuing to engage with the sensors it was apparent that they were driven to support the study and in turn, support the project of exacerbation prediction:

“Yea I’m planning to use them every day. My main motivation is obviously doing this study.” - *participant 1*

“I’m planning to use the sensors because I’m interested in this trial and the prediction of flare-ups.” - *participant 3*

“This trial is also important for me at the chance that this system is made but I think the support of the sensors is more important.” - *participant 2*

6.3.2 Theme 2: Accuracy is important to adherence

Whilst there was acceptance for the daily use of sensing technologies for supporting exacerbation prediction, participants still raised some concerns when exploring potential issues that might arise. In particular, participants seemed doubtful of the accuracy of sensors:

“The only concern I came up with was the Fitbit and its functionality. I used to have a Fitbit before I switched to the Garmin and one of the reasons why it was switched was because it wasn’t at an acceptable level of recording the information. So, you get things that did work, and things that didn’t work. Sometimes they’re good, sometimes some things were right, some things were wrong.” - *participant 2*

“So if I’ve got my heart rate up to 120 say, you’re in some severe exercise. My pulse oximeter it’ll drop below 100 but the Fitbit still going on at 120. I don’t know which is right and which is wrong. Both pulse oximeters do exactly the same thing, even though they’re from different companies. Now I’ve reported this to Fitbit but it’s a bit like talking to the wall, you know.” - *participant 3*

These concerns were only expressed by participants with a history of sensor use and their concerns regarding previous inaccuracies.

The focus group also delved into the topic of health anxiety. One participant expressed concerns about health anxiety, particularly in light of their preexisting heart-related issues and the visualisation of these issues through a sensor:

“I would be a bit wary to look at that one. Because at the moment well, I’m having issues with blips in my heartbeat. That one, I’d just be a little bit wary about because I think it probably frightens me a little bit more than anything else.” - *participant 5*

Despite these concerns, the participants did not have any strong objections to any particular sensor but it highlights the need for good accuracy of the sensors so that they can be trusted to avoid unnecessary worry.

6.3.3 Theme 3: Perceived inaccuracy and usability issues may hinder digital spirometry adherence

The most prominent obstacle to adherence became evident in the context of spirometer usage. Every participant encountered some form of difficulty while using the spirometer. Initially, during the focus group discussions, the spirometer garnered significant interest, and participants were enthusiastic about its usage. However, as the study progressed and during the midpoint and final interviews, a noticeable shift in perception became apparent. The device was no longer viewed favourably. A primary issue revolved around the challenge of getting the spirometer to accurately record their attempts.

“Not pleased with the blower it took me 4 attempts. I feel like there is a knack to it, if I give it the hardest blow, it usually comes up with not accepted. It’s pretty frustrating but apart from that I enjoy it.” - *participant 1*

“The spirometry one, occasionally I get an unacceptable reading. It’s usually a hesitant start but I’m not worried about it. I’m doing them as I would do normally on a proper spirometer.” - *participant 4*

One participant, in particular, was notably concerned about this, as they believed that the more inaccuracies they encountered, the more challenging it became to ensure accurate recording of the results:

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS
STUDY II

"I've been having a real problem with getting the spirometer going. Sometimes it takes 7 or 8 goes and of course, the more you get it wrong the more you get it wrong. It's self-perpetuating. But I got there in the end." - *participant 3*

This primarily stemmed from the physically demanding nature of the spirometry test itself, which necessitates participants to repeatedly exhale all the air from their lungs. As the testing sessions progressed, participants experienced increasing breathlessness, making it progressively more challenging to accurately measure their lung function. Many participants described the test as a physically taxing experience:

"I don't know with the spirometer I'm not 100% certain because I'm not sure if it's getting my readings right and it's quite exhausting especially if you get a test wrong and have to do all 3 again." - *participant 1*

"I don't like blowing into the thing. It doesn't blow properly. I'm sure it's wrong and I'm right [laughs]. Just like continuously blow, blow, blow, blow. You just get a quick blast and that's it. Whereas before it used to be a lot longer process. So, I think the tubes are too large in my own opinion, not that it counts for anything" - *participant 2*

Participants in the focus group expressed concerns about the accuracy of the sensors. These concerns came to fruition with regard to the results from the spirometer:

"The spirometer could be really good, but it doesn't think it correctly captures how I am breathing." - *participant 2*

"I have been concerned with the readings from the spirometer. The readings have been lower than hospital readings. If I had believed them I think god I'm going to die tomorrow." - *participant 3*

The frustration associated with administering the test and the perception of its poor accuracy may have been exacerbated by receiving suboptimal results. This potential connection aligns with earlier concerns expressed in the focus group regarding the importance of accuracy and the impact of health anxiety:

"I'm barely alive looking at the results. you know you google things and you see stuff like that and it will scare the life out of you." - *participant 1*

One participant remarked that seeing how much his lung function had declined was a significant deterrent to daily use and a clear barrier to adherence:

"It is saying my FEV has deteriorated by more than 10% in the last 6 months. I think if it wasn't for this trial, I probably wouldn't use it as much." - *participant 3*

Similarly, another participant expressed that the unalterable nature of the results was demotivating them to use the spirometer:

"But the spirometry is a bit disappointing to me, you know, the results. There's not any motivation to use it because it's not going to change anything. It's just giving you the information about what state your lungs are in." - *participant 3*

It may be that the frustration with registering tests is directly linked to what the participants perceived as poor results:

"As far as the spirometer, I wasn't satisfied after I had done it, that I'd got the best outcome. I just didn't feel I blew hard enough long enough; we'll have to see what tomorrow does." - *participant 4*

"I watched the guide videos, and it didn't make a difference, it seemed to make it worse!" - *participant 2*

"And you know I get those results from the spirometry and it's a bit disheartening. It's not as good as I think it should be." - *participant 3*

6.3.4 Theme 4: Adoption of increasing physical activity and calorie tracking

Two of the five participants reported not noticing any changes in their behaviour since starting to use the sensors on a regular basis. one of these two suggested their behaviour had not changed due to the length of time they have had COPD and its severity:

"I've gotta say not really. My COPD is at such a level, a low level, I don't need to manage it that much. I just need to make sure I don't go out in a fog and just wait for an exacerbation to hit." - *participant 2*

The other participant indicated he had already begun changing his behaviour prior to the use of the sensors to try and lose weight and improve overall health but felt that using the sensors further encouraged their positive behaviour change:

"Pre-Christmas I was getting continual exacerbations and haven't exacerbated recently. But I

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS STUDY II

think it's because I've been losing weight and getting healthier. I started this before the sensors but using these sensors I have been able to make better targets." - *participant 2*

On the other hand, as per the accounts of the remaining three participants, they believed that throughout the study, they had been making choices conducive to their well-being, such as increasing physical activity by walking more and adopting healthier eating habits. These participants frequently mentioned that they had been actively striving to achieve daily goals, such as accumulating 10,000 steps a day:

"I'm trying to get healthier and fitter. So I'm walking more and I think I'm getting better as my oximeter is going up. I'm averaging 94 and 95." - *participant 1*

"I'm doing more exercise but not necessarily pulmonary rehab. It's good for strengthening but not the same as exercise. I've been doing more walking and more steps but I'm feeling much better. hitting those 10,000 steps. I've been more active throughout the day." - *participant 4*

Even participant 2, who previously attributed his lack of behaviour change to the severity/stage of his COPD, expressed a strong desire to engage in more physical activity:

"Both types of things, like the spirometer, it self-challenges you to do better. If there's nobody challenging, you... it's just in your make-up, well it's in my makeup you know." - *participant 2*

Similar to physical activity, participants described aiming for weight loss goals through caloric tracking and restriction:

"I'll just go into the app, hang on. The section here breaks down my calories, and how I'm doing. Then it breaks down my walk, how many calories I do on that and the cardio, how many calories are burned on that. So that's interesting too, to work out the exercises I need to do to get the best results. I'm still overweight. I need to lose weight, so that's really good because weight is more of a problem when you're dealing with COPD." - *participant 4*

"I added a calorie tracking thing with the Fitbit yesterday. But I mean, that's a personal thing. I don't think that's part of your study but it's just to see what it was." - *participant 1*

6.3.5 Quantitative use rates of the myCOPD app and sensing technologies

myCOPD adherence during the study was 100% for three of the participants. One participant missed 4 days of use. One participant only used the myCOPD app twice. The use rate patterns

6.3. RESULTS

for myCOPD for each participant can be seen in Figure 6.2.

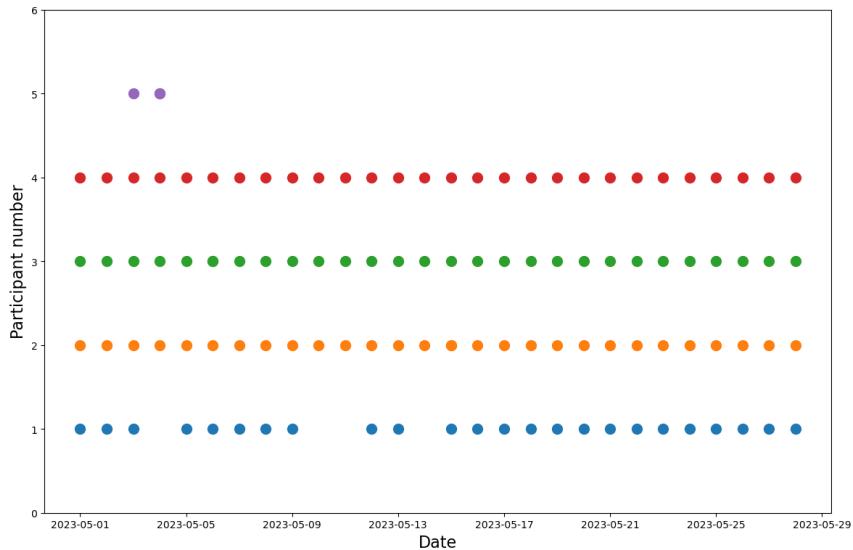


Figure 6.2: myCOPD use rates. Each dot represents an engagement with the technology for that day. Blue is participant 1, orange is participant 2, green is participant 3, red is participant 4, purple is participant 5

Pulse oximeter adherence during the study was 100% for only one of the participants. Two of the participants had adherence over 90%, one participant had adherence over 80% and one had nearly 80% adherence. The use rate patterns for the pulse oximeter for each participant can be seen in Figure 6.3.

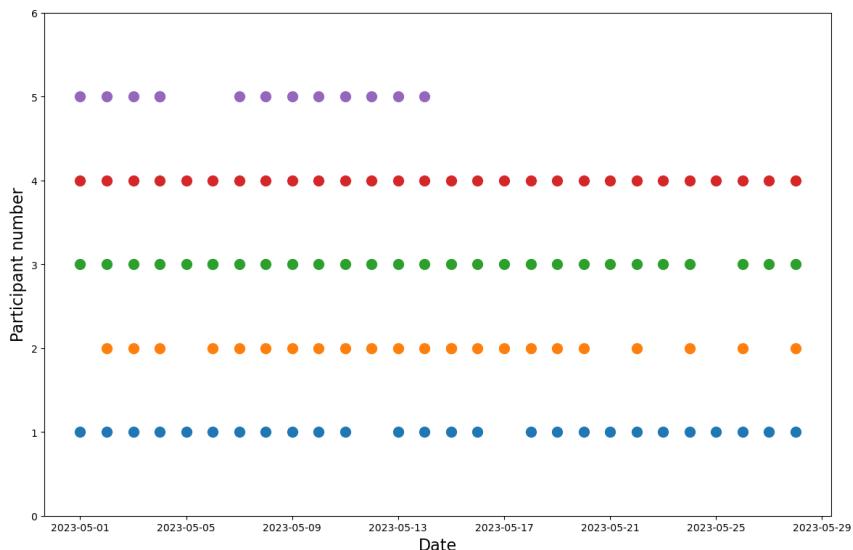


Figure 6.3: Pulse oximeter use rates. Each dot represents an engagement with the technology for that day. Blue is participant 1, orange is participant 2, green is participant 3, red is participant 4, purple is participant 5

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS
STUDY II

Spirometer adherence during the study was 100% for three of the participants and over 90% for the other two. The use rate patterns for the spirometer for each participant can be seen in Figure 6.4 where each dot represents an engagement with the technology for that day.

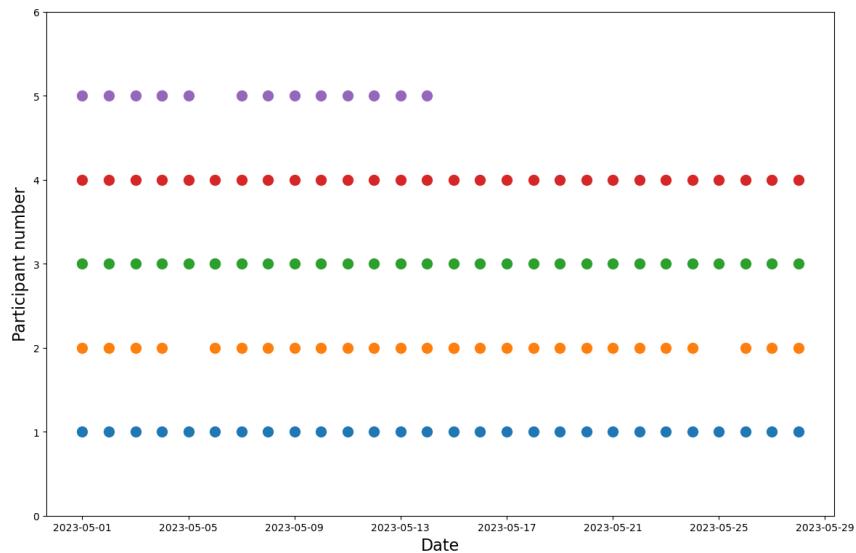


Figure 6.4: Spirometer use rates. Each dot represents an engagement with the technology for that day. Blue is participant 1, orange is participant 2, green is participant 3, red is participant 4, purple is participant 5

Wrist-worn wearable adherence during the study was 100% for one participant, over 90% for another, and over 85% for two participants. The use rate patterns for the wrist-worn wearable for each participant can be seen in Figure 6.5 where each dot represents an engagement with the technology for that day.

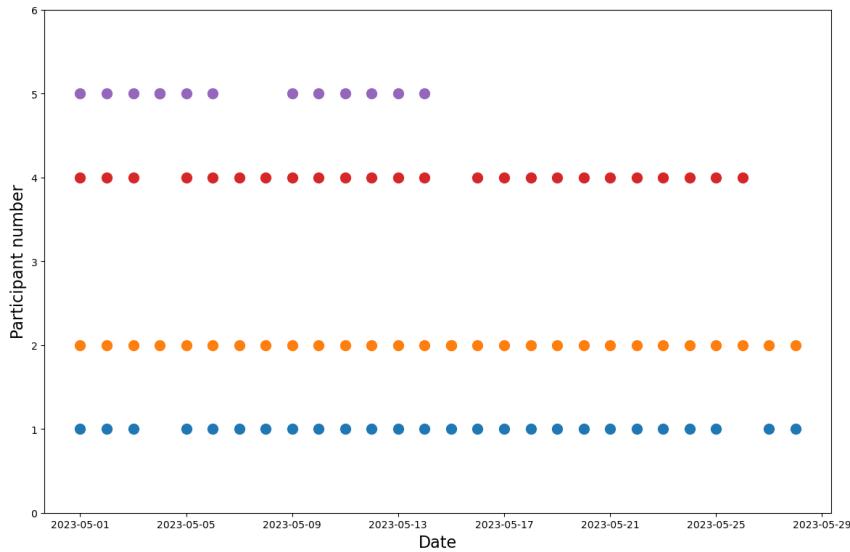


Figure 6.5: Wrist-worn wearable use rates. Each dot represents an engagement with the technology for that day. Blue is participant 1, orange is participant 2, green is participant 3, red is participant 4, purple is participant 5

6.3.6 Sensor data quality and exacerbation prediction potential

From the pulse oximeter, The mean oxygen and mean pulse rate are compared to symptom scores. For the spirometer, peak flow and FEV₁ are compared to symptom scores. For the wearable, RHR, HRV, sleep SpO₂, sleep RR, sleep temperature, steps, and sedentary minutes are compared to symptom scores.

6.3.6.1 Pulse oximeter

The pulse oximeter captured participants' SpO₂ and pulse rate when they clipped the sensor to their index finger and opened the app to record. The values represented in the visualisation are the daily mean of each of the means of measurements captured every second by the pulse oximeter during a recording. Figure 6.6 shows the pulse rate of participant 2 over one month remaining fairly consistent with a self-reported symptom score of 1 "normal for me".

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS
STUDY II

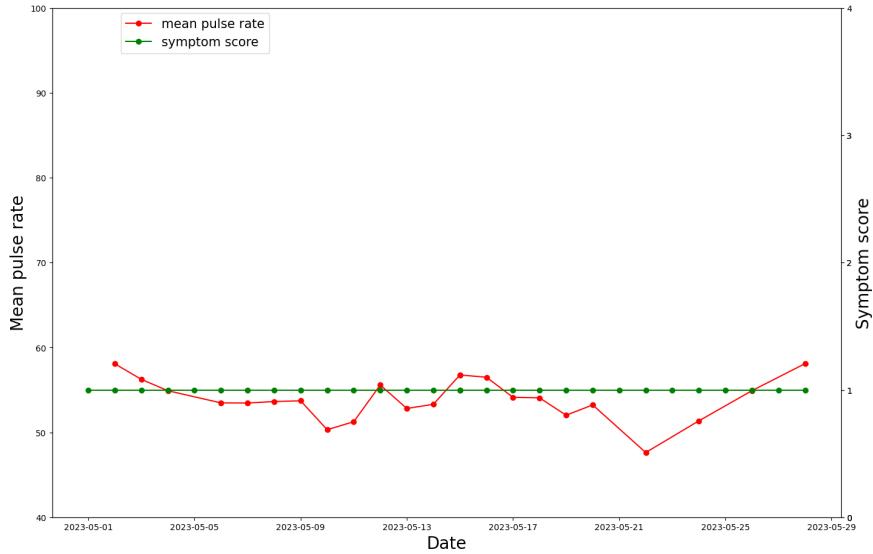


Figure 6.6: Participant 2 pulse oximeter mean pulse rate during one month.

Figure 6.7 is the mean pulse rate for participant 3. Participant 3 logged a symptom score of 2 “mild deterioration” on 2023-05-06, 2023-05-08, and 2023-05-17 and the rest as 1. The pulse rate is significantly more variable for this participant which may be because he uses the pulse oximeter when he exercises. The pulse rate did not appear to differ significantly when the participant first reported a symptom score of 2. However, the participant’s mean pulse rate had the largest decrease before the final report of a symptom score of 2.

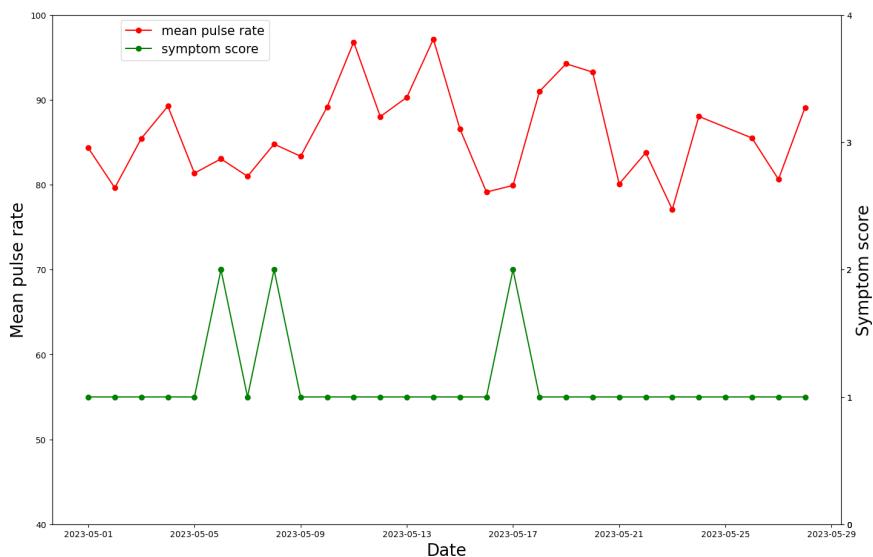


Figure 6.7: Participant 3 pulse oximeter mean pulse rate during one month.

The mean SpO₂ for participant 2 was fairly variable (nearly 8% difference between the highest and lowest) despite continued reporting of normal health. This variation can be seen in Figure

6.3. RESULTS

6.8.

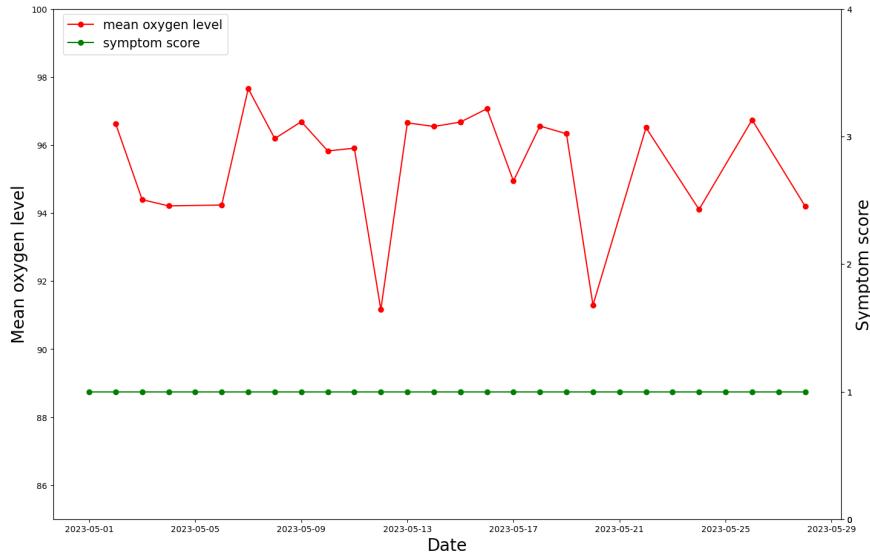


Figure 6.8: Participant 2 pulse oximeter mean SpO₂ during one month.

Participant 3 had a similar variation with the most significant decline occurring around the first two reports of a symptom score of 2, however, the participant reported a symptom score of 1 on the lowest recording of SpO₂. There is also a decline in mean SpO₂ around the last report of a symptom score of 2 as seen in Figure 6.9.

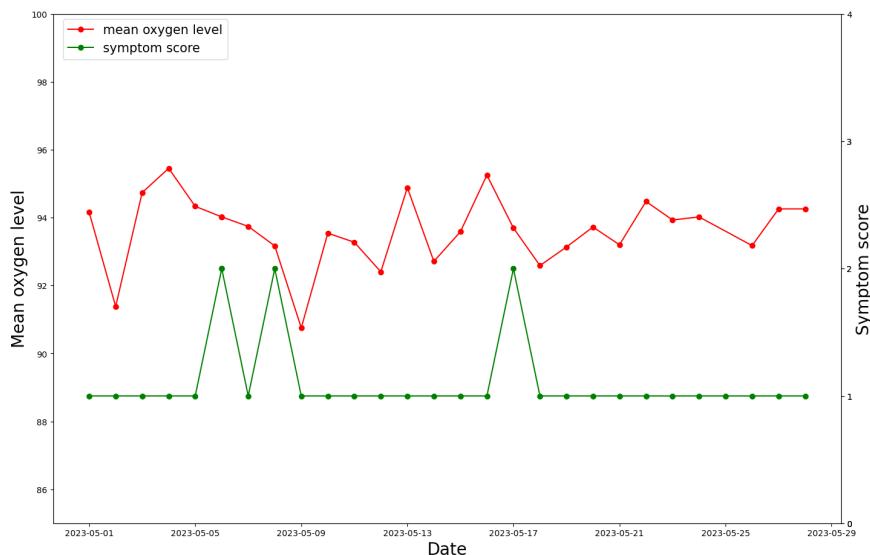


Figure 6.9: Participant 3 pulse oximeter mean SpO₂ during one month.

6.3.6.2 Spirometer

Participant 2 had variable peak flow across the month but experienced a significant decline from 2023-05-09 to 2023-05-10 of over 100 L per minute as seen in Figure 6.10. Despite this sudden drop, the participant reported a symptom score of 1 “normal for me”.

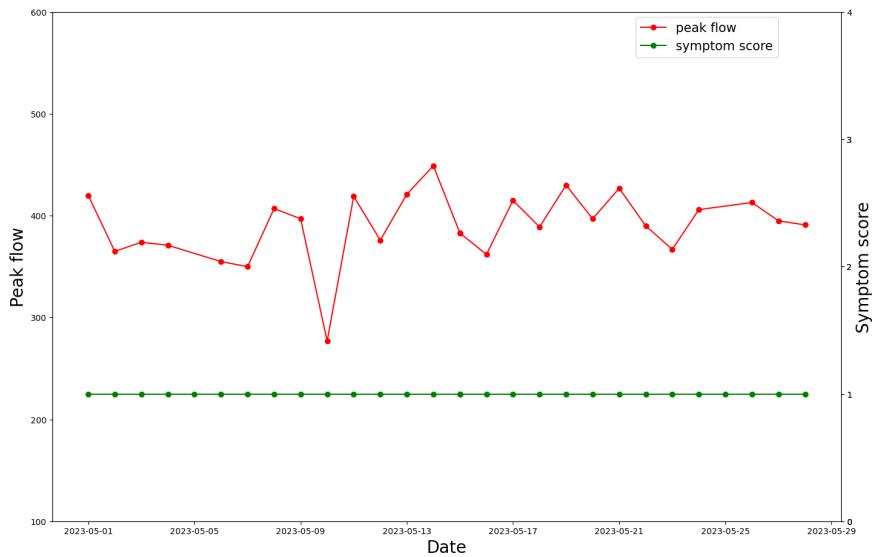


Figure 6.10: Participant 2 spirometer peak flow during one month.

Figure 6.11 shows participant 3 experienced a decline in peak flow over the course of a few days surrounding the first two reports of “mild deterioration.” However, the participant reported a symptom score of 1 after and had a declining peak flow for 3 days. Moreover, during the third report of a symptom score of 2, the participant has a small increase in peak flow.

FEV₁ for participant 2 is variable; decreasing or increasing by over a litre in the course of a few days as seen in Figure 6.12.

6.3. RESULTS

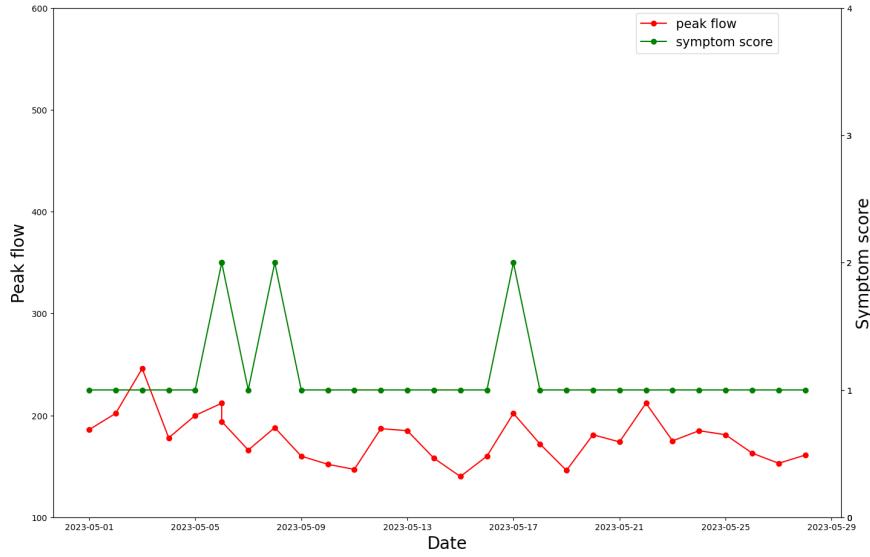


Figure 6.11: Participant 3 spirometer peak flow during one month.

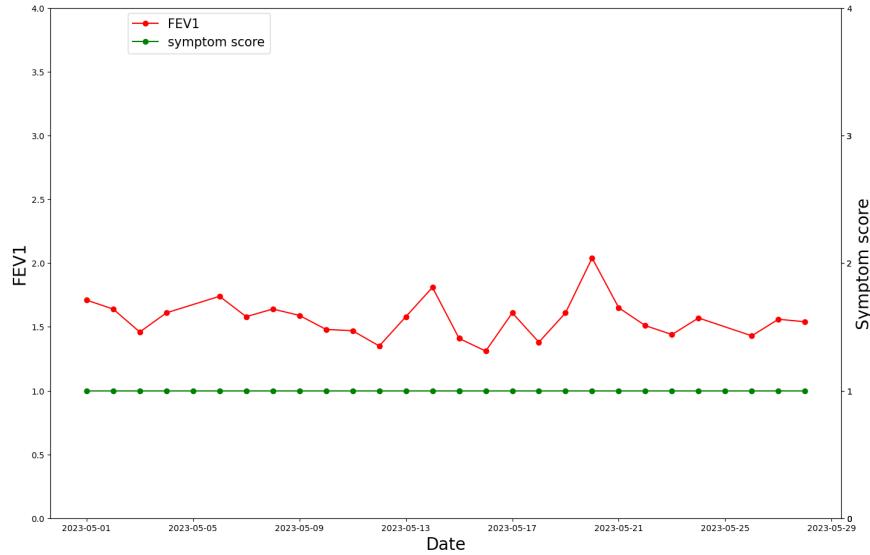


Figure 6.12: Participant 2 spirometer FEV₁ during one month.

The FEV₁ stayed consistent for participant 3 remaining around 0.5-0.8L for the duration of the month even though they reported 3 instances of a “mild deterioration”. This visualisation can be seen in Figure 6.13.

6.3.6.3 Wearable

Figure 6.14 shows participant 2 had a slight decline in resting heart rate over the course of the month without significant day-to-day variations.

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS
STUDY II

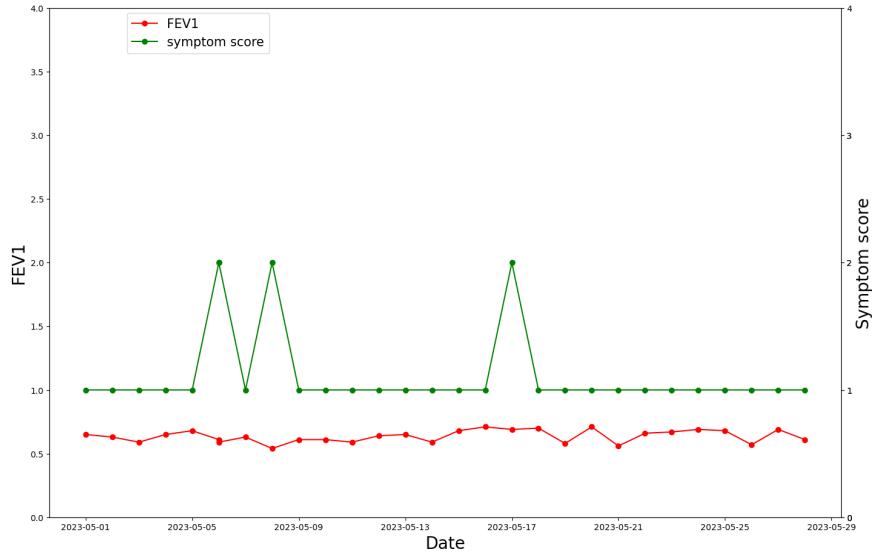


Figure 6.13: Participant 3 spirometer FEV₁ during one month.

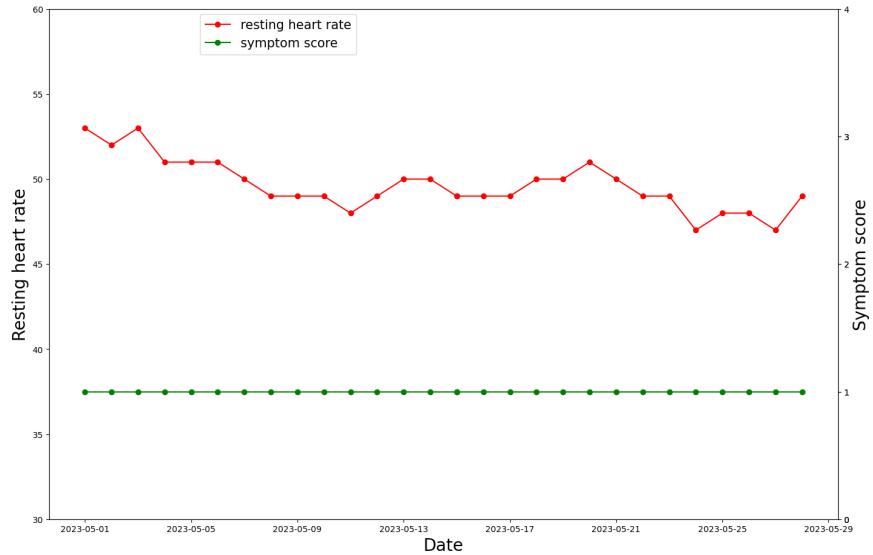


Figure 6.14: Participant 2 wrist-worn wearable resting heart rate during one month.

Figure 6.15 shows participant 5 had a very slight increase in resting heart rate during the first week of monitoring which remained roughly the same for the duration of the monitoring period. Participant 5 only used the myCOPD app for two days out of the 14 but used them to log he had been feeling worse, specifically symptom score 2 a “mild detereoration”. The days that are not reported would be a symptom score of 1 as the participant reported he felt symptomatically normal for the rest of the study.

Participant 2 maintained a relatively stable Root Mean Square of the Successive Differences (RMSSD) measurement for HRV throughout the entire month, ranging consistently between 20

6.3. RESULTS

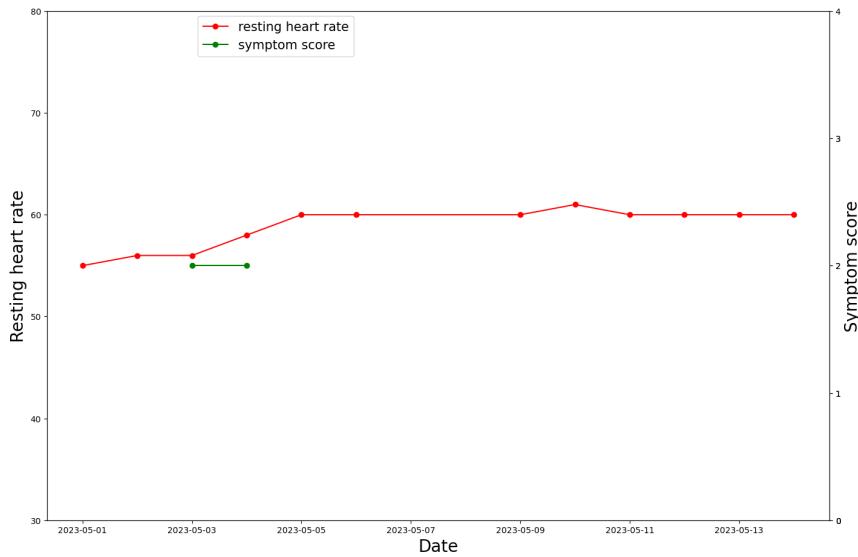


Figure 6.15: Participant 5 wrist-worn wearable resting heart rate during the two weeks.

and 30 milliseconds. The participant experienced a drop in their HRV by nearly 10 milliseconds in RMSSD between the last two records and between 2023-05-16 and 2023-05-17, as seen in Figure 6.16. A sudden drop in HRV may indicate stress or potential illness, yet participant 2 reports his health was “normal for me”. Conversely, it may be normal to experience this amount of change in RMSSD but there is currently a lack of evidence to suggest a clinically meaningful difference.

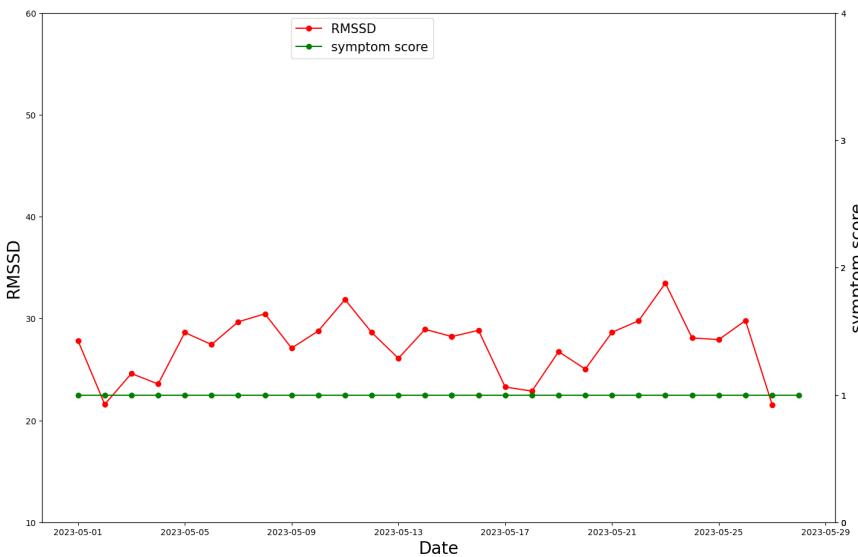


Figure 6.16: Participant 2 wrist-worn wearable heart rate variability as measured by the root mean square of successive differences (RMSSD) between normal heartbeats during one month.

Participant 5 has significantly more variation with a 60-millisecond difference between two

consecutive readings as seen in Figure 6.17. Nevertheless, the HRV data for this participant exhibited significant irregularities, with half of the entries missing and two days registering as 0. It's possible that this substantial variation could be attributed to computational errors, particularly when taking into account the participant's report of feeling worse on the day with the highest HRV, which would typically suggest good health. Alternatively, the participant mentioned experiencing heart-related fluctuations, and it's plausible that these fluctuations were responsible for the pronounced variation observed in the data.

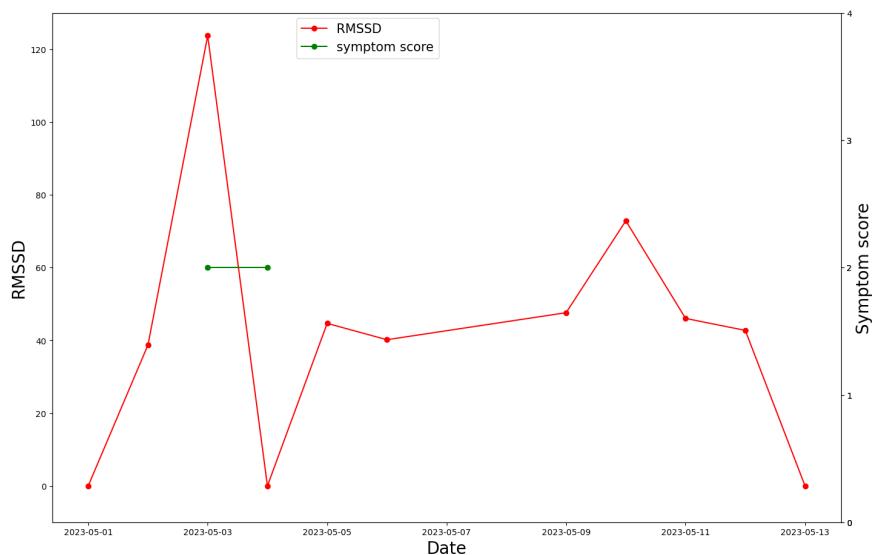


Figure 6.17: Participant 5 wrist-worn wearable heart rate variability as measured by the root mean square of successive differences between normal heartbeats during two weeks.

Sleep SpO₂ and sleep respiratory rate for participant 2 is shown in Figures 6.18 and 6.19. The participant had variations in sleep SpO₂ increasing by 6% and then decreasing by the same amount again in the course of one week, between 2023-05-12 and 2023-05-19. Despite these fluctuations in sleep SpO₂ the participant reported feeling normal. The participant's sleep respiratory rate was mostly consistent at around 20 breaths per minute. However, there were a few significant drops that spiked low by nearly 10 breaths per minute on 5 separate occasions. Moreover, there was a large spike after a drop followed by a report of 0. These fluctuations are possibly due to computational error due to the consistency of the other records.

6.3. RESULTS

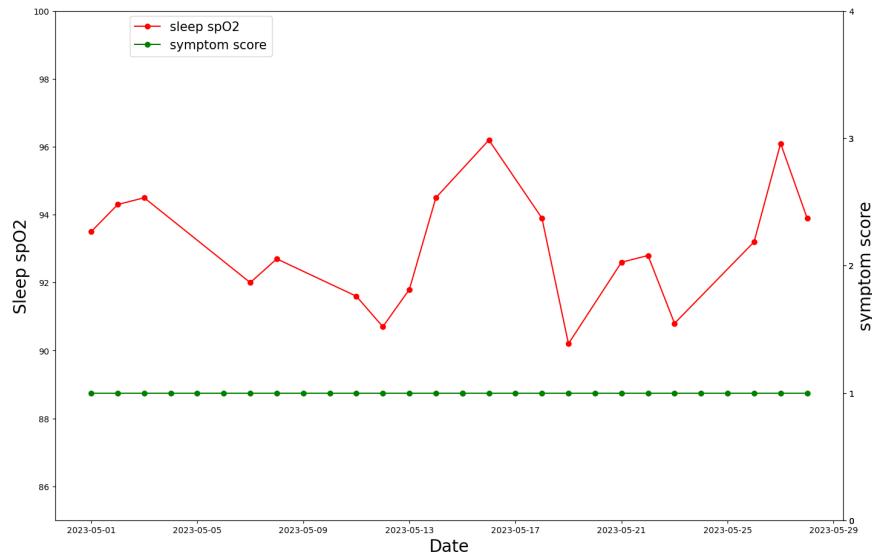


Figure 6.18: Participant 2 wrist-worn wearable sleep SpO₂ during one month.

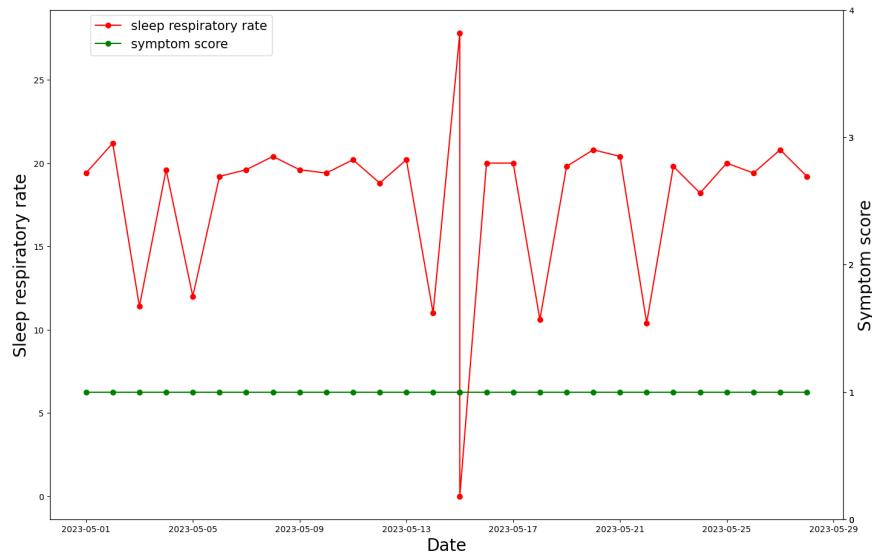


Figure 6.19: Participant 2 wrist-worn wearable sleep respiratory rate during one month.

Participant 5 had much more consistent readings for sleep SpO₂ and respiratory rate as seen in Figures 6.20 and 6.21. Interestingly, the measures stayed consistent over the two days the participant felt a “mild deterioration”. A noticeable drop in sleep RR to 0 occurred on 2023-05-12. This anomaly could be attributed to either computational errors or the possibility of the smartwatch being improperly worn during sleep.

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS
STUDY II

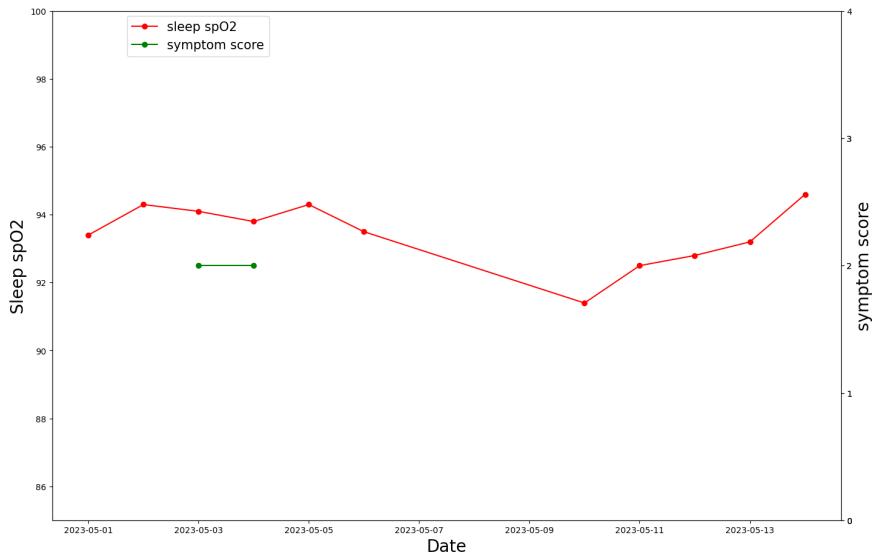


Figure 6.20: Participant 5 wrist-worn wearable sleep SpO₂ during two weeks.

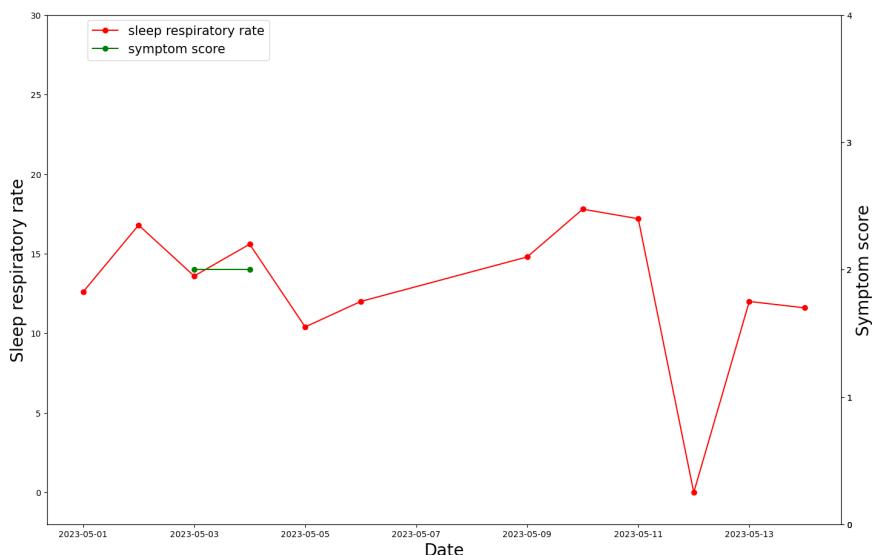


Figure 6.21: Participant 5 wrist-worn wearable sleep respiratory rate during two weeks.

Both participant 2 and 5 experienced normal day-to-day fluctuations in steps and sedentary minutes as shown in Figures 6.22, 6.23, 6.24, and 6.25. Both participants exhibit significant fluctuations in their daily step counts, with participant 2, in particular, demonstrating a striking pattern of alternating between 2,000 steps on one day and 10,000 steps on the next, all while consistently reporting normal symptoms.

6.3. RESULTS

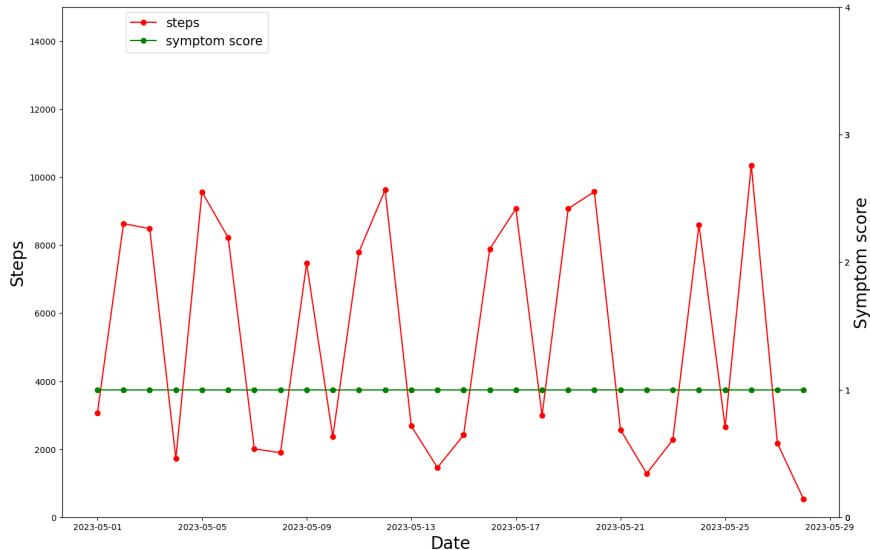


Figure 6.22: Participant 2 wrist-worn wearable steps during one month.

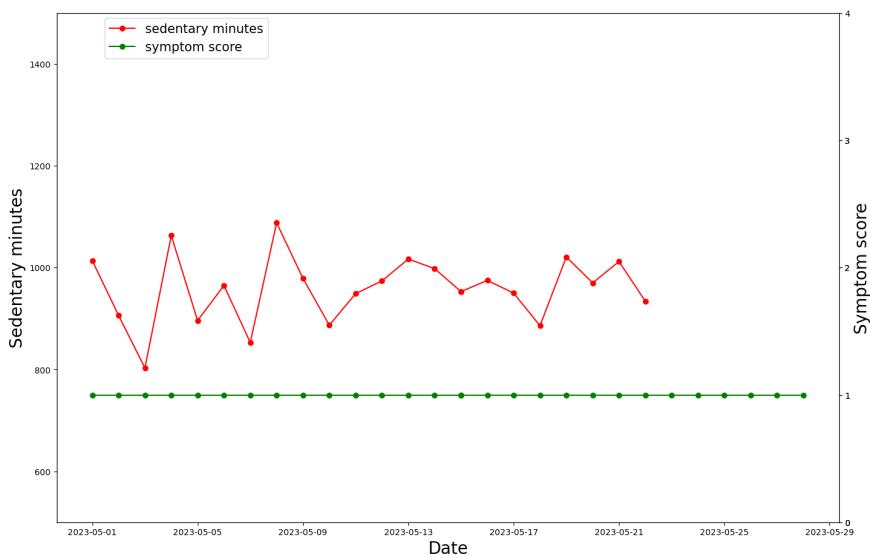


Figure 6.23: Participant 2 wrist-worn wearable sedentary minutes during one month.

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS
STUDY II

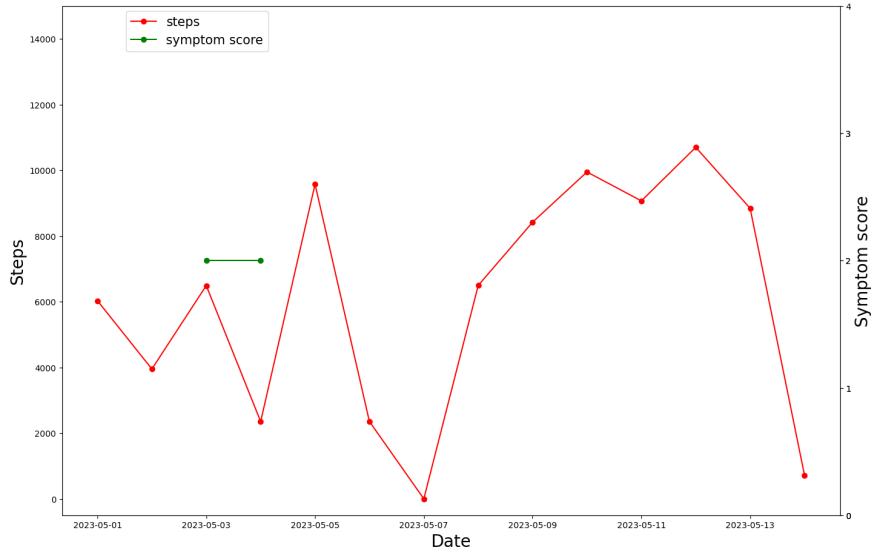


Figure 6.24: Participant 5 wrist-worn wearable steps during two weeks.

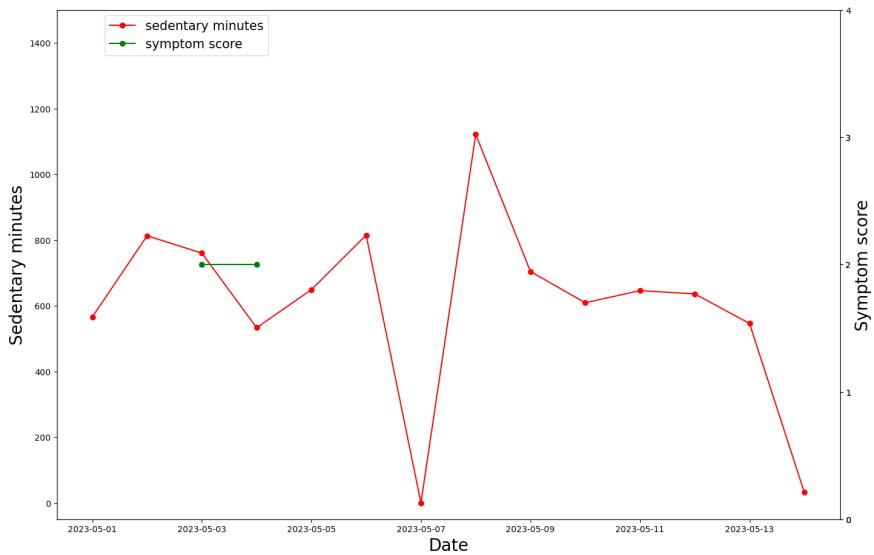


Figure 6.25: Participant 5 wrist-worn wearable sedentary minutes for two weeks.

On the two days that participant 5 felt a “mild deterioration” in their symptoms there was not a significant reduction in his step count or an increase in his sedentary minutes. On the day with a report of 2000 steps (2023-05-06), the participant was travelling for the day in the car. Excluding the day with 0 steps, when the participant wasn’t wearing the wrist-worn wearable, 2023-05-04 is the day with the lowest step count, which was approximately 2,000 steps lower than the next lowest count observed on 2023-05-02. This observation could imply a decrease in steps during periods of declining health. Notably, there was no corresponding increase in sedentary minutes on these specific days.

6.4 Discussion and conclusions

The purpose of this study was to identify the feasibility and acceptability of additional sensors for myCOPD users to support exacerbation prediction. The focus group identified that the participants were eager to engage with sensing technologies and that there could be seamless incorporation into their daily lives. However, concern about the accuracy of the sensors was identified at this time. This concern came to fruition with the digital spirometer with many of the participants perceiving the device as inaccurate and hard to use. Despite the issues with the spirometer, the participants described significant motivation to use the sensors due to the information they provide them to inform their health management and the potential to one day have access to an exacerbation prediction tool. The establishment of a routine enabled the seamless incorporation of all the sensors despite concerns and difficulties of use. This is exemplified in the use analysis that shows almost daily use of all the sensors. Whilst incorporation into the daily lives of myCOPD users shows promise, there is a lack of evidence to support sensor measures altered during intermittent days of mild deterioration. However, as none of the participants experienced a 2-week period of symptom worsening prodromal to exacerbations, no firm conclusions can be made on the potential of the sensors to support exacerbation prediction.

From the outset of the focus group, it was clear there was enthusiasm to engage with the sensing technologies. The participants conveyed they felt the sensors could be seamlessly incorporated into their daily lives. This positive perception indicates promising prospects for the future implementation of an exacerbation prediction intervention. Nevertheless, the limited sample size in this study poses a challenge in conclusively determining whether a broader population of myCOPD users would exhibit a similar willingness to accept and trial such interventions.

The interviews during the study elucidated the rationales behind seamless incorporation, a significant factor being establishing a routine. Each participant used the sensors roughly at the same time each day. Participants described establishing new routines using the sensors when they first woke up or setting an alarm in the evening. Others incorporated the sensors into their current routines like when taking medication or when using the myCOPD app. The participants felt strongly that this habitual use of their sensors was a key factor in their seamless incorporation. Within the myCOPD app, a sensor integration feature could be designed to seamlessly connect compatible wearables with the app. This sensor integration module could offer users optional settings to configure push notifications, taking the form of messages, to serve as reminders for their monitoring tasks. Another significant motivator for use is the feedback from the sensors. Positive affirmation of good health, useful information to inform daily health management, informing behaviour change, and tracking long-term health were highlighted as

CHAPTER 6. SENSING, PREDICTIONS, AND ALERTS IN COPD EXACERBATIONS STUDY II

the most important uses for sensor feedback. Similarly to routine, I believe there is potential to use this idea to encourage daily use in future interventions. One participant noted:

“The whole thing is a good support system for me. Visual representations are of real interest to me to understand trends in my health. And the system motivates me. I would wear this watch from now on.” - *participant 4*

If the sensor data were presented within the myCOPD app through visually informative graphs and figures, accompanied by contextual information that could establish a connection between the sensor data and the user’s health status, there exists a significant opportunity to enhance user retention, thereby ensuring the continued and frequent utilisation of sensor technology. Finally, all participants cited a key reason for continued daily use was the completion of the study to support research towards exacerbation prediction. It follows that having access to an exacerbation support/prediction tool with the knowledge that daily use of sensing technologies increases its accuracy or is a requirement to have access to the tool could incentivise daily use.

The participants did not scrutinise the wearable and pulse oximeter, while the digital spirometer raised some concerns. In the focus group, when questioned about potential problems using the sensors, the discussion centered around accuracy and mistrust of said accuracy. Whilst participants did not explicitly say what degree of accuracy is acceptable or more importantly to what degree of inaccuracy would result in non-use, it was clear that perceived inaccuracy could be a barrier to daily sensor incorporation. These concerns came to fruition with regards to the use of the spirometer with almost all of the participants believing the spirometer did not correctly capture their lung function. The scepticism about the spirometer’s accuracy could stem, in part, from participants’ dissatisfaction with their spirometer readings. Many participants expressed feelings of discouragement when confronted with their readings, which in turn led to concerns about their health or a sense of frustration due to their perceived inability to alter or enhance the readings. These challenges were compounded by the frequent need for participants to repeat tests, as they were often labelled as being not completed correctly, which some participants found to be a tiring and frustrating process. All the participants used the same spirometer (MIR SmartOne®). The accuracy for the spirometer on the technical data sheet states the volume accuracy is $\pm 2.5\%$ or 0.05 L and the peak flow accuracy is $\pm 10\%$ or 0.33 L/s. One participant undertook an outpatient spirometry test with a clinician and reported on the difference between the two:

“My FEV₁ percentage is now 25 which is very disappointing. But the spirometer is 2 to 3 points lower than what it actually is, which is even more disheartening than the actual result.” - *participant 2*

While the spirometer's inaccuracy range in this case is narrower than the range specified in the technical datasheet, the participant expressed dissatisfaction with this outcome. This observation suggests that when an individual is displeased with a result, they may have minimal tolerance for the margin of error, especially if the error is in a negative direction. Regarding the potential use of a spirometer in future exacerbation prediction, it is advisable to explore alternative spirometry devices to determine which one is the most user-friendly. To address concerns related to distressing results and inaccuracies, users should be clearly informed that there is always a margin of error associated with all sensors. Moreover, offering participants the option to hide or view results upon request may help mitigate their distress when confronted with poor results. Ultimately, the concerns expressed by participants regarding the spirometer may raise doubts about its suitability for exacerbation prediction, as long-term adherence to its usage might prove impractical.

Importantly, four out of five participants reported modifying their behaviour after using the sensors. Among these four individuals, one had already begun making these changes before participating in the study, with the goal of losing weight to reduce the frequency of exacerbations. Nevertheless, they believed that the sensors provided valuable support in achieving their pre-existing goals. This observed behaviour change has the potential to yield positive effects on outcomes within the intervention group in an RCT. Chapter 2 contains a narrative synthesis of all the RCT and machine learning studies using RPM in COPD exacerbations. Crucially, this potential effect is not discussed in these studies, nor is it included in the trial design. This study has highlighted that patient users of sensor technologies are likely to engage in behaviour change resulting in increasing physical activity, calorie tracking, and weight loss targets. However, it's crucial to note that this behaviour change was self-reported by the participants and not objectively monitored. Nonetheless, this information underscores significant implications. Further research is needed to comprehensively understand the extent of behaviour change in individuals who have initiated engagement with sensor usage and to determine the duration of this change's maintenance. It is well-reported in the literature that self-monitoring can cause behaviour change. A systematic review and meta-analysis by Compernolle *et al.* found that interventions using self-monitoring as a behaviour change technique can reduce sedentary behaviour [128]. Similarly, Baker and Kirschenbaum found in a study of 56 participants in a long-term cognitive behavioural weight-loss program that self-monitoring was positively correlated with weight change and not monitoring was negatively associated with weight change [129]. If further research confirms that self-monitoring with sensors leads to significant and enduring behaviour change, this has profound implications for both past and future RPM studies. If participants commence behavioural improvements, such as increased physical activity and achieving and sustaining a healthy weight, over the course of the study, several positive outcomes can be anticipated. These include likely enhancements in

the person's HRQL, a potential reduction in the frequency and severity of exacerbations, and a probable decrease in the associated costs of COPD management. Moving forward, future RCTs should consider incorporating three distinct groups for comparison: the control or usual care group, a sensor group with no clinical oversight or exacerbation prediction models, and the intervention group equipped with sensors and clinical oversight or machine learning AECOPD prediction models. This approach would facilitate the delineation of the benefit of behaviour change that might occur with sensing and the benefit of timely intervention in the course of exacerbations.

The evidence from the use rate pattern analysis demonstrates that overall use remains high across both the myCOPD app and the three sensors utilised in this study, with use rates consistently ranging from 80% to 100%. However, it is worth noting that Participant 5 did not engage with the myCOPD app. The scarcity of data entries in this context presents potential challenges and raises concerns regarding previous work related to the creation of exacerbation labels. This concern is especially pertinent when classifying events like this as stable, particularly when there is no available 14-day historical record. If similarly sparse data entries are similarly labelled as stable, there's a risk that some instances of stable labels may actually represent exacerbations but lack reports from the preceding or subsequent days. In such cases, our labelling system would categorise them as periods of stable health erroneously. To gain a better understanding of how sparse data impacts exacerbation labelling within the myCOPD dataset and the potential for misclassifying stable health periods, further investigation is warranted. Moving forward, the modelling strategies I will discuss in more depth in Chapter 7 will likely necessitate increased utilisation of the myCOPD app. It will be crucial to emphasise to patients that the effectiveness of exacerbation prediction will heavily depend on consistent usage of the myCOPD app and regular input of symptom scores and CAT scores. Making this information clear to users could potentially lead to heightened app engagement and, consequently, overall enhancements in model performance.

Comparing visualisations of the results from the sensors and myCOPD symptom score it is unclear what the indices' potential for AECOPD prediction is. In a handful of instances, it appears there is a decline in indices surrounding days of reported mild deterioration the most striking of which is seen in the decline of mean SpO₂ in participant 3 between the 2023-05-04 and the 2023-05-09 as measured by the pulse oximeter in Figure 6.9. Importantly, on the day the lowest mean oxygen was recorded by the participant, they reported "my symptoms are normal for me" within the myCOPD app. Moreover, many of the indices show considerable day-to-day variation despite reporting normal health. This evidence suggests the physiological and functional variables may have an insignificant impact on AECOPD prediction as normal variation may limit the predictive capability of indices. Conversely, none of the participants

6.4. DISCUSSION AND CONCLUSIONS

experienced an exacerbation which is preceded by 14 days of clinical worsening. It may be the gradual and persistent decline of physiological and functional variables prodromal to exacerbations results in clearly distinguishable patterns from normal day-to-day variations and isolated bad days that can be captured in a feature and used to train exacerbation prediction models. Importantly, a few instances were identified where computational errors or incorrect usage led to occurrences of 0 or missing variables. In such cases, imputation can be employed as a corrective measure to address these errors and missing values. Imputation techniques are utilised to substitute missing or erroneous data points with estimated or predicted values.

In coding the transcripts from SPACE Study I and II, two options were considered: semantic coding and latent coding. Semantic coding focuses on surface-level content, looking at explicit and observable elements in the text. It categorises words, phrases, and concepts without delving into deeper interpretations. Latent coding, on the other hand, uncovers underlying meanings and implicit concepts embedded in the text. It goes beyond surface content to reveal concealed insights and patterns inferred through careful analysis. In this work, semantic coding involved identifying and categorising direct references to the myCOPD app, medication management during exacerbations, actions in response to exacerbation predictions, and language used in warnings. In contrast, latent coding aimed to uncover implicit themes like personal responses to a RED ALERT, acceptance of exacerbation prediction, concerns about NHS service access, and the importance of accuracy when using a sensor. For this PhD work, a combination of both approaches was used. Integrating semantic and latent coding provided a comprehensive understanding of the data, encompassing explicit content as well as nuanced, underlying layers of meaning and context.

Rigour in qualitative research sparks extensive debate among researchers with diverse viewpoints. This debate hinges on established criteria, methodological flexibility, and the unique context of qualitative research. Rigour in research is closely linked to the research paradigms at play. The quantitative approach prioritises markers of rigour such as objective reality, empirical observation, quantifiable measurements, generalisability, and standardised procedures. In contrast, qualitative research places importance on context, reflexivity, and deep engagement with the research subject. Traditional criteria like reliability and validity are contested in qualitative research due to their fluid and subjective nature, leading to alternative criteria like trustworthiness, transferability, confirmability, and dependability that align better with interpretive paradigms. In the case of SPACE Studies I and II, qualitative research played a vital role in understanding patients' perspectives on upcoming technologies. These insights guide the design of these technologies to enhance their chances of success. Subsequent stages will complement these findings with quantitative and qualitative research, focusing on usage rates and patient reports. Reflexivity, a cornerstone of qualitative research rigour, revolves around acknowledging and addressing the

researcher's influence on the research process. It's a topic of ongoing discussion, with some highlighting its benefits in terms of transparency and validity, while others raise concerns about subjectivity and its potential to hinder data interpretation. Balancing the researcher's subjectivity with methodological rigour remains a significant challenge. The discourse on rigour in qualitative research is dynamic, reflecting the evolving landscape of research paradigms. Ensuring rigour in qualitative research requires an ongoing dialogue that embraces the diversity of perspectives and paradigms within the research community. Acknowledging my own perspective and bias is crucial in this work due to my close involvement in the ongoing project development.

6.4.1 Limitations

This study was limited by sample size and diversity. All 5 participants had similar demographics and dispositions. The original thoughts and ideas surrounding engagement with sensor use may not be transferable to other myCOPD users. Similarly, the consistent use rate patterns may not be similar to that of other myCOPD users with different genders, backgrounds, and mentalities. While the sample size was small and participants shared similar demographics, this feasibility study serves as a starting point. Future research can expand to include a more diverse group. Although findings may not directly apply to all myCOPD users, they provide valuable preliminary data to build upon in larger, more diverse studies. The study was also relatively short, the duration being one month long. Whilst use rates were fairly consistent from start to finish there is potential that a longer feasibility study could've seen signs of reduction in use. Despite the short duration, consistent use rates indicate initial engagement. To furnish more robust evidence, larger and long-term studies must be undertaken before drawing definitive conclusions regarding the potential longevity of a sensor-based exacerbation intervention.

6.4.2 Conclusions

Despite the limitations of this study and reported issues related to daily sensor use, the study's findings suggest that integrating sensors seamlessly into the daily routines of myCOPD users is feasible. Valuable insights from participants shed light on their perspectives, enabling RPM to be employed in a manner that enhances the likelihood of smooth integration. This involves offering options for establishing a routine with alarms and push notification reminders, providing useful sensor-derived information and feedback to support health management, allowing users to control the flow of sensor data through view and hide options, and providing assistance for behaviour change. The high use rates observed during the one-month study period reinforce the potential for seamless sensor incorporation. The visual representations of sensor data in correlation with symptom scores suggest a need for gathering additional evidence before determining their efficacy in predicting exacerbations. Future studies should involve a larger,

6.4. DISCUSSION AND CONCLUSIONS

more diverse participant cohort, extend over a longer duration for more comprehensive sensor data collection for modelling, and continue the PPI qualitative research to provide additional evidence regarding the feasibility of daily sensing among myCOPD users.

Chapter 7

Discussion

This thesis has adopted an interdisciplinary methodology to predict AECOPD, encompassing data science, clinical science, machine learning, as well as qualitative and mixed methods research. Through the application of predictive analytics and PPI research methodologies, I have initiated the process of generating evidence for an exacerbation prediction tool. The ultimate goal is to facilitate its widespread deployment among patient users engaged in a digital therapeutic for COPD self-management. Since each chapter contains a discussion reviewing its findings, this final chapter will give an overview of the main points from the separate chapters as a combined body of work and discuss the overall strengths, limitations, and implications. I first provide an overview of each chapter, then I give a commentary on three main themes from the thesis: remote patient monitoring and machine learning in AECOPD, modelling data from a widely deployed digital therapeutic, and patient factors in exacerbation prediction models, warnings and sensing technologies. Subsequently, I will delve into the future avenues for this particular research and present my conclusions.

Chapter 2 proposes that employing machine learning algorithms for predicting AECOPD could significantly enhance the accuracy of RPM. The inconsistency in digital health interventions for AECOPD poses challenges for interpretation and comparison, potentially explaining variations in success. Limited research addresses patient factors related to the burden of self-monitoring and long-term adherence, particularly outside RCTs. Standardisation of exacerbation labels, study design, and data collection is imperative. The clinical utility of machine learning approaches requires validation through RCTs. Further research on the long-term adoption of RPM, especially within patient communities, is warranted to understand their needs.

Chapter 3 outlines the processing and structuring of myCOPD patient data for data analytics and machine learning. A 14-day exacerbation-free window facilitates the generation of exacerbation and stable labels aligned with clinical definitions. The validity of these labels is discussed in more depth in Section 7.1. A subset of myCOPD users with both labels, demonstrating more regular app use and severe disease were identified. Change in CAT score from a stable baseline emerges as a potent predictor of AECOPD. The methodology discussed in Chapter 3 transforms

patient-entered app data for analysis, suggesting a standardised approach using 14 days without exacerbations adhering to clinical definitions. Targeted machine learning approaches are likely necessary due to differences in the COPD population.

Chapter 4 utilises 5 years of data from 506 patients, encompassing 55,066 stable records and 1,263 exacerbations, to develop and validate AECOPD prediction models. The EasyEnsemble Classifier and AdaBoost model exhibit some accuracy for a 1-8 day prediction horizon. Real-world data from a widely deployed digital therapeutic has some modest potential in predicting AECOPD. However, improvements in both accuracy and inter-patient generalisation require novel features and a larger patient cohort.

Chapter 5, in a group of myCOPD patient users there was dissatisfaction with the current performance of prediction models, emphasizing the need for trust-building through trial and error. Participants provide insights into language preferences, automation, percentage risk incorporation, optional selections, and severity of warnings. Confidence in self-management and concerns about access to NHS services were shared. Incorporating additional sensing measures is crucial for improving prediction performance to achieve acceptability. Further research is warranted exploring an exacerbation warning system by looking closely at the language used, incorporating automation, expressing risk through numbers, and customisable warning types. Further investigation is necessary to understand patient responses to warnings of varying intensity.

Chapter 6 highlights the enthusiasm for engaging in sensing technologies in a small group of myCOPD users and barriers to adherence. Daily sensor use is encouraged through routine, feedback, and the potential for exacerbation prediction access. Challenges such as inaccuracies, operational difficulties, and poor health-related results may deter use. While sensor retention is high, evidence for their potential to enhance exacerbation prediction remains limited. Feasibility and acceptability of daily are warranted through larger, longer studies, although the spirometer's barriers to adherence should be addressed. Future sensor deployment should emphasise establishing routines, managing expectations regarding accuracy, and offering options to view or hide feedback. Longer-term studies with larger cohorts are both possible and necessary.

Conducting PhD research in collaboration with a limited company like my mhealth presents both unique opportunities and challenges. On the one hand, this partnership has provided access to valuable resources, funding, and real-world data that can significantly enrich the research process. Additionally, working closely with a company has offered practical insights and guidance on translating academic findings into tangible solutions for healthcare challenges.

7.1. REMOTE PATIENT MONITORING AND MACHINE LEARNING IN ACUTE EXACERBATIONS OF COPD

However, it's important to navigate potential conflicts of interest and maintain academic rigour and independence throughout the research journey. Balancing the priorities of academia and the objectives of a company can be complex, but it offers a chance to bridge the gap between theory and practical application in the field of healthcare, potentially leading to innovations that benefit both academia and industry. Effective communication, ethical considerations, and a clear understanding of the boundaries are key to making such collaborations successful in the pursuit of meaningful contributions to the field.

7.1 Remote patient monitoring and machine learning in acute exacerbations of COPD

In our initial review of the literature, we encountered a diverse range of studies, encompassing pilot investigations, feasibility assessments, large-scale RCTs, as well as retrospective endeavours centred on model development and validation. The overarching objective of this collective body of work is to facilitate the prediction and timely intervention of AECOPD. Notably, these investigations employed a variety of RPM techniques in pursuit of this common goal. Upon closer examination, it became apparent that the literature could be categorised into two distinct segments. The first segment utilises RPM in conjunction with clinical oversight to identify early signs of health deterioration indicative of the prodromal phase preceding AECOPD. Subsequently, this approach encourages intervention, either through the administration of rescue packs or, potentially, clinical intervention within a hospital setting. In contrast, the second segment employs RPM, but rather than relying on clinical oversight, it harnesses machine learning models trained on patient data to predict AECOPD events, thereby facilitating timely intervention. This differentiation is grounded in the divergent strategies these studies employ and the respective advancements within each field. In the context of the RPM approach coupled with clinical oversight, the primary emphasis centres on enhancing patient outcomes. This segment of research has generated a substantial volume of studies, to the extent that it has progressed to the point of conducting numerous RCTs, aimed at assessing the safety and efficacy of various RPM approaches within the context of COPD. Conversely, the second segment focuses on achieving high predictive accuracy for AECOPD through machine learning models. These models draw upon retrospective patient data, yet there remains a paucity of insight into how these models translate into tangible improvements in patient outcomes. In our review, we identified only a limited number of studies that elucidate how their predictive models contribute to enhanced patient well-being. A comprehensive examination of both segments within this field was imperative. We aimed to underscore the existence of these distinct segments and to amalgamate the intricacies of their respective research efforts. This endeavour sought to generate new knowledge by unearthing fresh insights into the state-of-the-art, assessing the efficacy of these interventions, and offering a critical evaluation of the diverse methodologies

employed. Consequently, the rationale for conducting a dual review became evident, allowing for a narrative synthesis of the selected studies. This approach was deemed necessary to furnish recommendations for future research endeavours in this domain.

In the initial segment of the literature, which delves into RPM with clinical oversight, there is notable diversity not only in the approaches employed but also in the outcomes for patients. There's a discernible variation in the characteristics of the patient populations featured in RCTs, spanning different disease severities and exacerbation frequencies. While some commonalities emerge in the use of RPM indices, such as the prevalent inclusion of physiological measures, particularly SpO₂, in 23 out of 29 studies, the utility of these measures in the early identification of AECOPD remains uncertain. Advanced prediction algorithms may be necessary to harness the full potential of these physiological variables. In contrast, indices related to lung function are present in nearly half of the studies, yet their inclusion does not consistently yield significantly better outcomes. Moreover, the utilisation of PRO varies, with some studies employing custom-designed questionnaires while others opt for established and validated ones like the CAT. The key to moving forward lies in achieving consistency when selecting indices for RPM in COPD.

The emergence of wrist-worn wearables offers a promising avenue for monitoring physiological indices, encompassing heart rate, respiratory rate, SpO₂, temperature, and even physical activity. These wearables require minimal patient input and are relatively unobtrusive, save for periodic recharging. It is imperative to investigate the efficacy of these wrist-worn devices in predicting AECOPD through dedicated model validation and development studies. Various questionnaires are routinely administered, often through tablet or smartphone interfaces, to gauge PRO. Nonetheless, there is a lack of evidence affirming the accuracy of many of these questionnaires in identifying or predicting AECOPD. However, the findings elucidated in Chapter 3 of this doctoral study demonstrate that even a slight alteration in the CAT confers significant predictive capabilities for identifying and forecasting exacerbations. Consequently, it is prudent to standardise the use of CAT assessments via tablet or smartphone platforms within this field.

How these indices are employed for exacerbation intervention exhibits considerable variation. Researchers often formulate basic algorithms that trigger alerts or “red flags” when two or more patient symptoms deteriorate or when specific patient variables experience significant changes, like a 2-point drop in mean SpO₂. However, this method raises two critical issues. Firstly, these algorithms often lack empirical justification and exhibit limited evidence of accuracy in identifying AECOPD correctly. Secondly, the inconsistency in algorithm design hampers effective intervention comparison. Alternatively, the first literature search identified RCTs that

7.1. REMOTE PATIENT MONITORING AND MACHINE LEARNING IN ACUTE EXACERBATIONS OF COPD

have daily RPM indices reviewed by a clinical monitoring team, which promptly contacts patients at the first signs of health decline. While this meticulous oversight could enhance AECOPD detection, the practicality of deploying such an intervention on a large scale, given the extensive resources required, remains a challenge. Consequently, machine learning presents an appealing alternative, addressing issues of algorithmic inconsistency, accuracy, and resource burden.

The second literature search and narrative synthesis underscore the potential of machine learning algorithms to predict AECOPD days in advance with remarkable accuracy. Nevertheless, this segment of research lacks the robust RCTs seen in the first segment. This is understandable, as the field is still in its early stages, resulting in an absence of clinical investigations focused on improving patient outcomes. It is imperative to conduct research combining RPM and machine learning that have the potential for RCT scrutiny and real-world deployment. RCTs hold the key to generating the necessary evidence to ascertain whether integrating machine learning into RPM can transform the current inconsistencies into a solid evidence base demonstrating favourable outcomes in AECOPD. Moreover, the emphasis on ensuring the intervention's suitability for real-world implementation is paramount. If RCTs confirm the effectiveness of these interventions, they should be poised for scaling up to reach the hands of patients, thereby translating research findings into tangible benefits for the COPD population.

Both segments of the literature largely overlook patient-related factors. In the first segment, there is a conspicuous absence of data concerning patient perspectives on using various sensors, apps, and questionnaires daily to monitor their COPD, as well as the likelihood of long-term adherence. In the second segment, the response of patients to AECOPD prediction models remains uncertain and has the potential to impact intervention efficacy negatively. Therefore, there is an urgent need for PPI research to gain deeper insights into the burden of RPM on COPD patients and to facilitate the adoption and utilisation of AECOPD prediction models. This evidence can be used to design interventions that are more likely to be efficacious and have continued persistent engagement.

In Chapter 2, we introduce the concept of “exacerbation labelling” for the first time. Exacerbation labelling is a complex issue in AECOPD prediction. It involves defining what constitutes an exacerbation in the data, which can be quite challenging. There are different ways to define exacerbations, such as using patient-reported symptoms, clinician diagnosis, medication usage, hospitalisation, or a combination of these criteria. Each approach has its pros and cons. Patient-reported symptoms and medication use fall into the “symptom-based” category, while clinician diagnosis and hospitalisation are considered “event-based”. However, clinician diagnosis can also be symptom-based if the clinician verifies the patient’s symptoms and medication use

remotely. Patient-reported symptoms are advantageous because they capture all exacerbations, from mild to severe, and are resource-efficient. However, they lack oversight and verification with objective measures like SpO₂ or FEV₁. Clinician diagnosis is the gold standard but is resource-intensive and often not feasible for large datasets or in clinical trials involving many patients. Medication usage, particularly rescue packs, can reduce false positives due to patient hesitancy but may result in missing some exacerbations or delays in treatment initiation. Hospitalisation due to exacerbations offers the advantage of clinician diagnosis and may exhibit more pronounced fluctuations in RPM indices that could precede more severe exacerbation incidents necessitating hospitalisation. These pronounced fluctuations, in turn, have the potential to enhance the accuracy of prediction models. However, it may overlook milder events managed at home with rescue medication. The second aspect of the labelling challenge is how we distinguish between exacerbations and stable health when training machine learning models. This will be discussed further in the next section, considering the labels generated during this doctoral research.

In our analysis, we encountered limitations due to the limited number of studies available for review, which may restrict the generalisability of our findings to the broader COPD population. We primarily focused on RCTs for their robustness in assessing intervention effects, but this led to fewer studies. Similarly, our review of machine learning papers was limited due to the relatively narrow scope of research in this area, resulting in only 23 relevant papers during our literature search. Our strict inclusion and exclusion criteria further contributed to the limited study pool, although these criteria were necessary to maintain objectivity and rigour in our systematic review.

We acknowledge the potential for bias in our review, particularly in the context of prediction models. Although we didn't conduct a bias assessment, it's important to recognise that bias is prevalent in research developing prediction models. This bias can arise from various sources, including small study size, handling of missing data, and overfitting. However, our primary focus was on identifying different machine learning approaches for predicting AECOPD, rather than extensively assessing bias. Furthermore, we caution against relying too heavily on reported model accuracies, as their real-world performance may vary. It's crucial to subject these models to testing in RCTs to establish their clinical value. Regarding RCTs for RPM in COPD, while they are less susceptible to bias than other study designs, they are not entirely free from potential sources of bias. Despite not conducting a formal bias assessment, our conclusion remains that RPM alone may not significantly improve patient outcomes, as the evidence consistently leans towards non-significant results, as discussed in Chapter 2.

The conclusions drawn regarding RPM in COPD discussed in Chapter 2 were as follows; RPM

for COPD has inconclusive results, machine learning has the potential to enhance RPM and improve patient outcomes, RCT of RPM and machine learning is needed to demonstrate clinical utility, and this research needs standardisation across the field. The search strategy and inclusion and exclusion criteria are well-defined and add validity to these findings. However, a lack of interrogation into the biases that can affect the field is problematic. Although there is a possibility of bias, the research papers identified in the review selection process are of high quality with attempts to minimise or eliminate bias, ensuring integrity and reliability. Nevertheless, further studies attempting to identify bias in the field may be necessary to ensure the continuation of the rigour in this research.

7.2 Modelling data from a widely deployed digital therapeutic

Chapters 3 and 4 describe the data analytics and modelling that were undertaken during this doctoral study. Chapter 3 provided insight into the process of preparing the data for modelling through the development of a feature window, the creation of an exacerbation label, the combining of dynamic and baseline data into one DataFrame, and the analysis of the patient population and features within this DataFrame.

Organising the myCOPD data into feature windows has transformed it into a binary classification problem, this approach could be useful for similar sparse digital therapeutic datasets. An inherent limitation of the pre-processing algorithm is its computational expense. The existing method employs exhaustive searches throughout the entire dataset to locate data points within specified time windows. For the creation of 14-day time windows, the algorithm necessitates thorough searches for each data point, checking occurrences one day before, two days before, three days before, and so on, up to 14 days before and 14 days after each entry. This exhaustive search process is conducted for various variables such as CAT score, regular medication, and others. Consequently, the computational demand is substantial, and the processing time for each variable, including CAT score and regular medication, can extend over several days. Moreover, as user engagement with the app grows, the processing time is expected to increase proportionally, posing a scalability challenge for the algorithm. To handle larger datasets more efficiently, we can explore alternative pre-processing algorithms. One option is to group the data by patient keys and dates, allowing the algorithm to search within each patient key instead of scanning the entire dataset. Another approach could involve data imputation, where missing data is filled in using various techniques, including deep learning methods like autoencoders. This would transition the data from binary classification to time series modelling, but given the scarcity of data points, data imputation could introduce significant noise. Such an approach would necessitate a more consistent and frequent use of the digital therapeutic to be effective.

Classifying exacerbations and stable health in the data is crucial for training accurate AECOPD prediction models. This ensures we differentiate the prodromal period from temporary symptom relief in exacerbations and avoid mistaking moments of stable health within ongoing exacerbations. To ensure this occurred within our data clinicians and engineers participated in co-creation groups to develop exacerbation and stable labels through inspection of symptom score visualisations as discussed in Chapter 3. Ultimately it was decided that a 14-day window of no self-reports of “moderate deterioration” or “severe deterioration” before a report of “normal health” or “mild deterioration” was appropriate for a “stable health label” and before a report of “moderate deterioration” or “severe deterioration” for an “exacerbation label”. Table 7.1 describes the self-reports.

Table 7.1: The patient self-report options linked with the respective symptom score and self-report description

Patient report	Symptom score	Description
My symptoms are normal for me	1	Normal symptoms and will take the prescribed medication
Mild deterioration	2	More breathlessness than normal but no fever or change in sputum colour or volume, will take regular medication and use the reliever inhaler
Moderate deterioration	3	More breathlessness than normal and coughing up more sputum and/or sputum has changed in colour. Much more breathless despite using a reliever inhaler. Will use rescue medication containing oral corticosteroids and/or antibiotics
Severe deterioration	4	Breathing is much worse than normal despite treatment or chest pain or high fever. Will call the GP, 999, or get admitted to the hospital

Table 7.1 demonstrates our labels combine patient-reported symptoms and medication use. A notable strength of our label is its alignment with GOLD’s recent definition, which highlights symptom deterioration over 14 days before an exacerbation. Despite being created before this definition, our labels concur with it, demonstrating the label’s accuracy and relevance. Our label’s similarity to the definition, along with visualisations indicating no ongoing exacerbations were captured, suggests we likely accurately modelled the data leading up to the first exacerbation report. However, limitations include the lack of clinician verification and the inability to retrospectively adjust symptom reporting. Patients might initially describe “moderate deterioration” or “mild deterioration” but later realise it was just an isolated bad day. Hesitation poses a potential limitation, as patients may delay using rescue medication during an exacerbation, resulting in their reported exacerbation occurring a few days after the actual onset. Additionally, as discussed in Chapter 6, the lack of recent historical data can be problematic. For instance, in the SPACE Study II, a participant only used the myCOPD app twice to report a ”mild

deterioration” within a three-day window. This reporting would categorise it as a stable event, revealing a flaw in the current system. There is considerable uncertainty regarding whether a patient is experiencing an exacerbation during such a period. Therefore, it may be prudent to only label data with sufficient history to confidently determine whether a patient is stable or experiencing an exacerbation. This might entail considering only users classified as having “very high engagement” according to Cooper *et al*’s definition to ensure reliable labelling [130]. However, we must weigh this against the possibility of losing patient data.

We need to examine the factors leading to the significant reduction in the myCOPD user population, from almost 14,000 activated users to a final inclusion of only 506 in the model. This analysis can provide valuable insights for modelling data from other digital therapeutics. As explained in Chapter 3, the number of patient users that matched both our labels was much smaller than the general population. This is because they needed enough entries to correctly apply the labels and also use the app during exacerbations. This split the users into non-labelled and labelled groups. Importantly, the labelled group had a much higher proportion of users with very high engagement compared to the non-labelled group. However, there were still 448 highly engaged users in the non-labelled group. These users were excluded from the modelling because the final exacerbation prediction model is intended for individuals actively using the app during exacerbations. Therefore, any future intervention developed based on this model would likely be suitable only for this specific group. This decision has implications for future model development and clinical validation in RCTs. In future model development, we need a patient population that frequently experiences exacerbations and logs them in the myCOPD app for effective training. Similarly, in RCTs assessing intervention effectiveness through exacerbation-related outcomes, including individuals who do not experience exacerbations would necessitate significantly larger sample sizes and resources. Therefore, our focus is on modelling patients who use the myCOPD app during both AECOPD and stable periods.

We observed a further decrease in the myCOPD patient user count, declining from 1,758 to 506. This reduction occurred for two main reasons. Firstly, it resulted from the infrequent entry of the CAT score, a critical feature for our models. Users are prompted to input their CAT score every two weeks, and on the day of the prompt, they needed to complete the CAT before accessing the app and its features. The infrequent prompting has resulted in a decline in patient numbers because data points lacking a CAT score within 1-8 days before an AECOPD, which matched our label, could not be incorporated into the final model. The second reason for the reduction was missing data. Each data point included patient baseline characteristics such as GOLD group, age, pack year, hospitalisation history, etc. For a patient-user to be incorporated into the modelling, they had to possess all the baseline characteristics listed in the feature list stored in the myCOPD database; otherwise, they were excluded due to data incompleteness.

The models presented at the ERS Congress 2022 were trained on a dataset of 169 patient users [131]. At this stage of the modelling, we were using all the dynamic features: CAT score, symptom score, tile clicks, regular medication use, and reliever medication use. Similarly to the CAT score, this required the patient to engage in app use and use the medication diary to enable a data point to be included in modelling. Tile clicks only began being recorded at a later date limiting the length of time myCOPD data was available and reducing overall data availability. The small size of the dataset made overfitting likely and the generalisability of the model was limited. It was found that only using the CAT score and symptom score resulted in a dataset three times the size with better-performing models. Other research attempts to model digital therapeutic data with the potential to lose data due to incompleteness or sparse use should carefully consider the cost/benefit of incorporating features regarding how influential they can be to model accuracy vs. data loss for incorporation with other features.

In this doctoral study, we carefully considered our machine learning approach and chose ensemble learning, particularly AdaBoost, as the most promising method. Initially, we used KNN for its simplicity, but it showed limitations with scalability in larger datasets. Comparing AdaBoost and KNN, we found that AdaBoost excelled in handling complex data relationships and feature interactions. We selected ensemble learning because this approach leverages multiple base classifiers to improve predictions, especially beneficial for complex and noisy datasets. Ensemble learning reduces overfitting and bias and enhances model stability and adaptability by addressing individual classifier sensitivity to minor variations in training data.

After implementing AdaBoost models, we realised that model limitations might stem from the degree of separation between features for the two classes. To visualise this separation, we used dimensionality reduction techniques like t-SNE. Initially, we tried Principal Component Analysis (PCA), but it didn't provide informative results. We then turned to t-SNE, which is better at preserving local structures and capturing non-linear relationships. It became clear that our existing features might not fully distinguish between the classes. To enhance model performance, we'll need to introduce new features, possibly involving sensor data or innovative feature engineering and modelling methods for myCOPD data. I discuss this in detail in section 7.4.

Our dataset, analytics, and modelling represent an innovative approach for the unobtrusive monitoring of COPD patients to predict AECOPD. The insights and methodologies from this research are valuable for anyone working with data from widely deployed digital therapeutics, serving as a foundational reference for their efforts. Additionally, this analysis sets the stage for a potential intervention accessible to myCOPD users globally, advancing COPD management and offering practical applications to improve the lives of individuals with this condition worldwide.

7.3 Patient perspectives on exacerbation prediction models, warnings and sensing technologies

Chapters 5 and 6 consist of the findings from SPACE Study I and II. The focus of this research was to discover the patient factors regarding AECOPD prediction models, exacerbation warnings, and the use of sensor technologies in their daily lives. This was realised mainly through qualitative approaches: specifically thematic analysis. Qualitative research does not aim for the same type of statistical generalisability as quantitative research, however, it offers transferability and theoretical insights. The applicability of qualitative findings to different contexts depends on the depth of understanding, sample selection, and the extent to which the findings align with similar situations. The very small sample and lack of diversity of study participants may have impacted the transferability of the findings. However, the key concepts and shared meanings appeared frequently in different participants across different settings adding validity to the findings. Table 7.2 illustrates the themes and their implications from the qualitative research in both SPACE Study I and II.

Table 7.2: A summary of the findings and implications from qualitative research in this thesis

Theme	Implications
With further improvement, the exacerbation prediction models will be acceptable	The small focus groups supported the development of the machine learning models but highlighted the need for improvement in their accuracy. Users may worry about how exacerbation prediction integrates with NHS services, necessitating careful consideration from both patients and clinicians. Nevertheless, the concern about healthcare services' reliability is motivating a desire for access to exacerbation support tools. Exacerbation support will need to be developed and deployed with an AECOPD prediction model. Users with similar engagement and experience in managing exacerbations may have coping strategies that, when combined with exacerbation prediction, could enhance their overall health.
The exacerbation prediction models aren't useful in the real world	The current accuracy of the models is understood to be unusable to myCOPD users and is preventative of the initiation of user onboarding.

Seeking understanding and trust in exacerbation prediction models	Trust is not guaranteed and inaccurate predictions may result in non-use, to manage expectations model accuracy needs to be conveyed carefully reminding patients to rely on their intuition of how they feel. The search for understanding the intervention may be important to myCOPD users, clear explanations should be provided to users of prediction models so they can better understand how they work.
Insights and desires in the context of an exacerbation warning	The sub-themes suggest an exacerbation warning should be “you are at X% risk of having a flare-up”, this has implications for modelling with feedback on percentage certainty of model decisions. The interest in an automated warning implies a preference for background monitoring. myCOPD users could find value in having both a passive warning system that operates in the background and an active warning/support tool that requires their active involvement in inputting symptoms and sensor data. However, the severity of warnings remains uncertain and requires additional research. It’s crucial to strike a balance between a warning that is taken seriously and one that doesn’t unnecessarily cause concern.
An exacerbation prediction would trigger preparation and caution	Some myCOPD users may already have good insight into how to respond to an exacerbation prediction, and these insights could be reiterated in the future to mitigate any uncertainty.
Seamless incorporation of daily sensing	The small cohort of myCOPD users with similar characteristics displayed enthusiasm for daily sensing; however, it’s important to note that this positive engagement might not universally guarantee seamless onboarding for all users initially. When deploying sensors there should be encouragement to establish a digital routine using alarms and notifications. Feedback from the sensors was important to the users, educating users on sensor information or developing visualisations to display the information may improve long-term adherence. As the potential for exacerbation prediction motivated use, access to exacerbation prediction should motivate use.

7.3. PATIENT PERSPECTIVES ON EXACERBATION PREDICTION MODELS, WARNINGS AND SENSING TECHNOLOGIES

Adoption of increasing physical activity and calorie tracking	Using sensors as part of an intervention may encourage behaviour change, this requires further research and has implications for outcomes in RCT.
Accuracy is important to adherence	Sensors with good accuracy are likely important to adoption, identifying accurate sensors and conveying this information to users will be necessary.
Perceived inaccuracy and usability issues may hinder digital spirometry adherence	The concerns voiced by the participants about digital spirometry cast doubt on its effectiveness for predicting exacerbations, given that consistent, long-term adherence to its use could be challenging.

The use rate analysis revealed consistent sensor usage, despite concerns about the spirometer. However, conflicting findings between qualitative and quantitative research make it challenging to provide clear recommendations. The qualitative evidence suggests long-term spirometer adherence is unlikely, while quantitative data hints at potential consistency. Furthermore, the quantitative evidence on the spirometer's support for exacerbation prediction remains inconclusive. To fully understand its predictive potential, larger, longer-term studies capturing exacerbations are needed, combined with machine learning. Despite its issues, continuing with the spirometer in the SPACE Studies may be necessary due to its predictive power, which could enhance model acceptability. We can use insights from SPACE Study I and II to encourage sensor incorporation and reduce adherence barriers. Testing various digital spirometer options could help identify more user-friendly devices. Participants should be informed about potential error margins in their readings, acknowledging that values might be lower than actual levels. Integrating sensors into an app, like myCOPD, could enable an option to hide or view sensor data, potentially alleviating health-related anxiety.

To reflect on the overall qualitative approach, consider below Table 7.3 which shows a completed COREQ (Consolidated criteria for reporting qualitative research) to report important aspects of the research team, study methods, context of the studies, findings, analysis and interpretations.

Table 7.3: Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist

No.	Item	Description	Answer
Domain 1: Research team and reflexivity			
Personal characteristics			

CHAPTER 7. DISCUSSION

1.	Interviewer/ facilitator	Which author/s conducted the interview or focus group?	Glyde led the focus group which was also conducted by Cliffe and Morgan and Glyde conducted the interviews alone
2.	Credentials	What were the researcher's credentials? E.g. PhD, MD	Glyde has a BSc, Cliffe has a BSc and an MSc in clinical psychology, Morgan has an MBChB and BSc
3.	Occupation	What was their occupation at the time of the study?	At the time of the study, Glyde was a doctoral student, Cliffe was a postdoctoral researcher, and Morgan was an airways clinical research fellow
4.	Gender	Was the researcher male or female?	Glyde is male, and Cliffe and Morgan are female
5.	Experience and training	What experience or training did the researcher have?	Glyde has previous training and research experience conducting focus groups, Cliffe has published research in which she conducted focus groups, and Morgan has training in supporting interviews
Relationship with participants			
6.	Relationship established	Was a relationship established before study commencement?	No relationship was established before the study
7.	Participant knowledge of the interviewer	What did the participants know about the researcher? e.g. personal goals and reasons for doing the research	Participants were informed at the start of each focus group that Glyde is a doctoral student working towards developing an exacerbation prediction tool for the myCOPD app
8.	Interviewer characteristics	What characteristics were reported about the interviewer/ facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	It was reported that Glyde has a keen interest in the topic as his current and foreseeable research lies in this topic and is likely biased towards acceptability
Domain 2: study design			

**7.3. PATIENT PERSPECTIVES ON EXACERBATION PREDICTION MODELS,
WARNINGS AND SENSING TECHNOLOGIES**

Theoretical framework			
9.	Methodological orientation and Theory	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, and content analysis	Thematic analysis underpinned the studies
Participant selection			
10.	Sampling	How were participants selected? e.g. purposive, convenience, consecutive, snowball	Purposive convenience sampling
11.	Method of approach	How were participants approached? e.g. face-to-face, telephone, mail, email	Email contacted for registering interest
12.	Sample size	How many participants were in the study?	Six in SPACE Study I and five in SPACE Study II
13.	Non-participation	How many people refused to participate or dropped out? Reasons?	One participant dropped out of SPACE Study II due to health concerns
Setting			
14.	Setting of data collection	Where was the data collected? e.g. home, clinic, workplace	Data was collected in a home setting
15.	Presence of non-participants	Was anyone else present besides the participants and researchers?	There was a clinician present in SPACE Study II to explain how to use the spirometer
16.	Description of sample	What are the important characteristics of the sample? e.g. demographic data, date	Important demographics including white population for both studies and all male for SPACE Study II
Data collection			

CHAPTER 7. DISCUSSION

17.	Interview guide	Were questions, prompts, or guides provided by the authors? Was it pilot-tested?	Questions were provided by the author and were discussed and reworked in two co-creation meetings with clinicians and engineers
18.	Repeat interviews	Were repeat interviews carried out? If yes, how many?	No repeat interviews
19.	Audio/visual recording	Did the research use audio or visual recording to collect the data?	Audio recording to collect data
20.	Field notes	Were field notes made during and/or after the interview or focus group?	Key ideas were noted by the lead after the focus groups
21.	Duration	What was the duration of the interviews or focus group?	The focus groups lasted an hour, and the interviews lasted from 10 to 30 minutes
22.	Data saturation	Was data saturation discussed?	Daily interviews were scheduled to last 7 days but were stopped at 5 due to belief in data saturation
23.	Transcripts returned	Were transcripts returned to participants for comment and/or correction?	No transcripts were not returned to participants for comment
Domain 3: analysis and findings			
Data analysis			
24.	Number of data coders	How many data coders coded the data?	Only Glyde coded the data
25.	Description of the coding tree	Did the authors provide a description of the coding tree?	A description of the coding tree was not provided but a thematic map was
26.	Derivation of themes	Were themes identified in advance or derived from the data?	Themes were derived from the data

**7.3. PATIENT PERSPECTIVES ON EXACERBATION PREDICTION MODELS,
WARNINGS AND SENSING TECHNOLOGIES**

27.	Software	What software, if applicable, was used to manage the data?	Nvivo was used to code and manage the data
28.	Participant checking	Did participants provide feedback on the findings?	Participants were not asked to provide feedback on the findings
Reporting			
29.	Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	Participants were quoted to illustrate themes using participant numbers
30.	Data and findings consistent	Was there consistency between the data presented and the findings?	Presented data and findings are consistent
31.	Clarity of major themes	Were major themes clearly presented in the findings?	Major themes are clearly presented in the findings, they are numbered and listed at the beginning of the results
32.	Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Minor themes are present with clear discussion

The responses to the COREQ questionnaire serve as a testament to the rigorous methodology employed in the qualitative research conducted during SPACE Study I and II. Despite certain limitations pertaining to the study participants, the meticulous approach taken, along with the deliberate thematic framework chosen as displayed in Table 7.3 adds credibility and validity to the research findings. Moving forward, the insights gleaned from this work will be actively incorporated into future endeavours. This integration is expected to serve as a robust foundation for accumulating further evidence and insights, which will be invaluable in developing an exacerbation support tool tailored to the needs of myCOPD users.

7.4 Directions for future research

Through review, research, and meaningful interactions with patients during this doctoral study, valuable insights have been garnered that hold the potential to enhance existing predictive algorithms. A novel approach to modelling exacerbations using myCOPD data has been conceived. To begin with, it is worth noting that at the time of writing it has been nearly two years since the last data extraction from myCOPD. The forthcoming extraction is expected to encompass a larger user base, including individuals with more years of data and an increased number of exacerbation instances recorded. This expanded dataset will significantly bolster the generalisability of predictive models. The current method of extracting and engineering features will need to evolve substantially to be more algorithmically efficient. As evidenced by the t-SNE analysis, the present set of available features employed for modelling has likely reached its capacity for effective discrimination. Hence, a new set of features will be required to further distinguish between the two classes. Furthermore, combining features, as discussed earlier, tends to diminish the dataset, reducing data points and the patient-user count, and potentially impacting model performance and generalisability. Notably, the CAT score and symptom score, two frequently collected data points, have demonstrated their predictive potential for AECOPD in the myCOPD dataset without diminishing the amount of data available for training. Thus, investing efforts in refining these two features through a new approach to feature engineering may not only mitigate data loss but also enhance model performance. Feature engineering will involve tracking the change in CAT scores from a stable baseline. Instead of using a fixed mean of all entries as the stable baseline, a running average will be employed to account for changing health status in COPD. Additionally, 14-day and 7-day means of CAT and symptom scores leading up to stable or exacerbation labels will be integrated. Changes in these variables from the most recent 14-day and 7-day mean CAT and symptom scores before a stable label will also be considered. Indeed, the running average of the CAT score itself could serve as a predictive feature for AECOPD. Subsequently, after the creation of this dynamic dataset, individualised baseline characteristics can potentially be integrated to assess the extent of data loss and explore avenues for performance improvement.

To further expand the dataset, you can apply data augmentation techniques or generate synthetic data, thereby increasing the available sample size for modelling. With a substantially larger dataset at hand, neural network architectures can be harnessed to fully exploit the accuracy of the newly generated data. Neural networks have the capability to provide class probabilities, allowing the presentation of the percentage certainty of exacerbation occurrences. This feature was recognised as important to myCOPD users in Chapter 5.

An alternative approach highlighted in the conclusion of Chapter 2 is the pursuit of a

personalised approach. Such methodologies as novelty detection, recurrent neural networks, or memory-augmented neural networks for modelling patient data possess the capacity to handle temporal data and adapt to changing patient characteristics. However, it's important to acknowledge that these approaches may not be feasible in instances where a patient lacks sufficient exacerbation data for modelling.

Upon the generation of a novel set of AECOPD prediction models, the results obtained from model validation will dictate the subsequent steps of the SPACE Studies. Should the models exhibit a level of accuracy deemed acceptable to myCOPD patient users, the deployment of the exacerbation prediction models and development of warnings within myCOPD will be initiated. Conversely, if the accuracy remains below the required threshold, SPACE Study III will be undertaken. SPACE Study III is an expansion of SPACE Study II, encompassing a larger population of myCOPD users over an extended duration. This prospective observational study will leverage the insights gleaned from SPACE Study II to facilitate the seamless incorporation of daily sensing. Upon completion, use rate patterns will be identified, and predictive modelling using patient data will be employed to ascertain the necessity of a sensing approach for exacerbation prediction modelling. Based on the accuracy of the prediction model and its potential for generalisation, the exacerbation prediction system can be integrated into the myCOPD apps.

SPACE Study IV, an interventional study, would assess the safety and efficacy of the exacerbation prediction models and warnings integrated into the myCOPD app. This trial would commence by recruiting a diverse cohort of COPD patients, conducting a thorough recording of participant baseline characteristics, and randomly assigning them to one of four groups: an intervention group equipped with the myCOPD app, sensors for monitoring relevant physiological parameters, and an AECOPD prediction system; and three control groups, each with different combinations of the myCOPD app and sensor monitoring but without the AECOPD prediction system. Given the findings from SPACE Study II, which underscored the presence of behaviour change and the utility of sensor feedback, it is imperative to evaluate the model's efficacy against these potential benefits. The primary outcome of the trial offers two potential avenues. Firstly, it involves assessing the model's capacity to accurately predict impending AECOPD within the intervention group, contrasting it with the control group's ability to self-identify exacerbations. However, this outcome may be challenging to implement due to the absence of a universally accepted diagnostic gold standard and the requirement for rigorous clinical oversight to correctly identify exacerbations. Recognising that the primary goal of exacerbation prediction is to enhance patient outcomes, it may be prudent to consider alternative measures for improvement in patient health. Consequently, the number of hospital admissions and HRQL emerge as two commonly employed outcomes in RCTs for RPM in COPD. Opting for these as primary outcomes may

offer significant advantages over the quantification of exacerbation occurrences. Secondary outcomes could encompass the maintenance of lung function, a reduction in primary care contact, and a reduction in the overall cost of managing COPD. By adopting a meticulously designed RCT framework, this approach endeavours to furnish robust evidence regarding the effectiveness of machine learning and sensor technology in augmenting COPD management and mitigating exacerbation-related health risks.

In the foreseeable future, enhancing the accuracy of exacerbation diagnosis and tailoring targeted therapies through endotyping represents an achievable goal, thanks to the potential utility of sputum and blood biomarkers. A comprehensive systematic review conducted by Chen *et al.* sheds light on various promising blood biomarkers for diagnosing AECOPD [132]. These biomarkers include CRP, interleukin-6 (IL-6), and tumour necrosis factor-alpha (TNF- α). CRP consistently exhibited statistically significant elevations in AECOPD patients compared to control groups, suggesting its potential as a valuable candidate for aiding in AECOPD prediction. However, it's essential to acknowledge that blood-based diagnostics can be invasive, time-consuming, resource-intensive, and costly, potentially limiting their widespread adoption. Consequently, alternative approaches should be explored.

Research in the realm of salivary biomarkers has indicated that heightened levels of salivary CRP, procalcitonin, and neutrophil elastase (NE) could serve as predictors of AECOPD risk [133]. Nonetheless, a narrative review by Li *et al.* underscores the need for "rigorous and exhaustive diagnostic accuracy studies" to validate the clinical utility of these biomarkers [134]. Notably, Stockley *et al.* have drawn attention to the generally weak associations and lack of reproducibility in biomarker data across large patient cohorts [135]. They recommend a shift toward endotyping, a strategy that employs biological mechanisms to define distinct patient subgroups. This shift could potentially yield more substantial associations, fostering advancements in AECOPD prediction, precision medicine, and targeted therapies. However, this approach of endotyping would lead to a division of the COPD population, resulting in smaller sample sizes for training data. Widespread adoption of biomarkers would necessitate the development of cost-effective diagnostic assays, such as lateral flow technology, to ensure accessibility and uptake. Alternatively, recent advancements in capnography offer promising avenues for COPD diagnosis [136]. Capnography is a medical monitoring technique that measures the concentration of carbon dioxide in a patient's exhaled breath. Although evidence regarding its potential for predicting AECOPD is currently limited, it represents a noteworthy trajectory with the potential to further improve AECOPD prediction.

7.5 Conclusions

This thesis commenced with an extensive review of the literature, aiming to harness the potential of RPM to enhance outcomes in AECOPD through timely intervention. This comprehensive review underscored the existence of significant gaps in the literature, crucial to bridging in order to realise the ultimate goal of early prediction and intervention in AECOPD. These pivotal research areas encompass leveraging machine learning and RPM for the development of widely deployable AECOPD prediction models and gaining a comprehensive understanding of the patient-related factors that influence these interventions.

In response to these research gaps, our study embarked on an exploration of data generated by a widely deployed digital therapeutic, to identify its potential for predictive modelling of AECOPD. A preliminary analysis hinted at the promise of AECOPD prediction but with low PPV that limits clinical utility. Subsequently, SPACE Studies I and II were conducted to delve deeper into understanding the requirements and preferences of patient users. These studies provided insights into the creation of an exacerbation prediction tool that aligns with patient expectations, the seamless incorporation of daily sensing, and the identification of barriers to adherence. However, the small sample sizes and lack of diversity warrant further building on this foundation of evidence to provide more robust conclusions.

The PPV of the AECOPD prediction models currently stands at less than 10%, signifying a notably low accuracy. This result implies a high false positive rate, a concern that has been accentuated by feedback from patients during qualitative research. Despite my efforts, there has been difficulty in effectively communicating the intricacies of these models to patients. Moving forward, there is a clear imperative for further research and refinement aimed at enhancing the PPV of these models. Additionally, there is a crucial need for improved strategies in translating complex technical concepts to patients, ensuring a clearer understanding of the predictive capabilities and limitations of the models, and addressing patient concerns more comprehensively.

Given the evidence gathered during my doctoral study, there appears to be a potential opportunity to contribute to the development of an exacerbation prediction tool. However, it is essential to acknowledge the limitations, including the low accuracy of the models, small sample sizes in research groups, and a limited understanding of sensor technology performance. These challenges must be addressed and understood thoroughly before advancing further in this direction.

Appendix A

Search methods and strategy

The search strategy for a) SCOPUS remote patient monitoring, b) Web of Science remote patient monitoring, c) SCOPUS remote patient monitoring and modelling, d) Web of Science remote patient monitoring and modelling:

- a) TITLE-ABS-KEY ((copd OR "chronic obstructive pulmonary disease" OR "chronic obstructive lung disease") AND (telemedicine OR telemonitoring OR "remote patient monitoring" OR "remote monitoring" OR "continuous monitoring" OR "real-time monitoring" OR telehealth OR "mobile health" OR mhealth OR "digital health") AND ("RCT" OR "randomised controlled trial" OR "randomised trial" OR "randomised clinical trial" OR "clinical trial") AND (admission* OR readmission* OR exacerbation* OR "quality of life") AND NOT (feasibility OR pilot OR review OR rehab*))
- b) TS = (copd OR chronic obstructive pulmonary disease OR chronic obstructive lung disease) AND TS = (telemedicine OR telemonitoring OR remote patient monitoring OR remote monitoring OR continuous monitoring OR real-time monitoring OR telehealth OR mobile health OR mhealth OR digital health) AND TS = (RCT OR randomised controlled trial OR randomised trial OR randomised clinical trial OR clinical trial) AND TS = (admission* OR readmission* OR exacerbation* OR quality of life) NOT TS = (feasibility OR pilot OR review OR rehab*)
- c) TITLE-ABS-KEY ((copd OR "chronic obstructive pulmonary disease" OR "chronic obstructive lung disease") AND (telemedicine OR telemonitoring OR "remote patient monitoring" OR "remote monitoring" OR "continuous monitoring" OR "real-time monitoring" OR telehealth OR "mobile health" OR mhealth OR "digital health") AND (prediction OR algorithm OR "machine learning" OR "deep learning"))
- d) TS = (copd OR chronic obstructive pulmonary disease OR chronic obstructive lung disease) AND TS = (telecare OR telemedicine OR telemonitoring OR remote patient monitoring OR remote monitoring OR continuous monitoring OR real-time monitoring OR telehealth OR mobile health OR mhealth OR digital health) AND TS = (prediction OR algorithm OR machine learning OR deep learning)

Appendix B

Narrative syntheses

Table B.1: Narrative synthesis of randomised controlled trials of remote patient monitoring in COPD

Authors	Participants	Measures and study protocol	Results
Toledo <i>et al</i> [24]	Intervention n=67: age=71±8 years $FEV_1=42\pm20\%$ female=2% Control n=90: age=72±8 years $FEV_1=42\pm15\%$ female=3%	One-lead electrocardiogram (ECG), lung function, oxygen saturation (SpO ₂), blood pressure (BP), heart rate (HR) and PRO. This study was a researcher-blind, multicentre, randomised controlled trial with clinical specialists directly monitoring patient data.	There was no difference in the number of days to admission, in the mean number of COPD admissions, and no significant effect on health status related to SGRQ.
Koff <i>et al</i> [25]	Intervention n=20: age=66.6±9.1 $FEV_1=33.6\pm9.1\%$ female=50% Control n=20: age=65.0±8.2 $FEV_1=31.1\pm10.2\%$ female=55%	Health Buddy System: PRO (dyspnea, cough, sputum, fever) and self-report (activity, depression, Forced expiratory volume (FEV) ₁ , 6-minute walking test (6MWT), and SpO ₂). This was a 3-month, single-centre, RCT with a registered respiratory therapist monitoring data who contacted patients if red flags were raised on changes in measures.	The SGRQ improved (decreased) by 10.3 in the intervention vs. 0.6 in the control group, and healthcare costs decreased by US\$1,401 in the intervention group vs. an increase of US\$1,709 from pre-study costs of US\$9,248±18,897 (not statistically significant).

APPENDIX B. NARRATIVE SYNTHESES

Vitacca <i>et al</i> [26]	Intervention n=57: age=61.2±17.6 FEV ₁ =39±23% female=36% Control n=44: age=61.1±17.4 FEV ₁ =34±16% female=28%	Respicard: PRO (dyspnea, cough, sputum, sputum colour, weight/ankle oedema, temperature, neurological status, ventilator interaction (if used as part of care), and walk) and self-report (SpO ₂ and HR). At baseline, a nurse administered the Respicard. If there was a score variation greater than 3 points from baseline, the nurse contacted a pulmonologist for consultation.	The intervention group experienced fewer hospitalisations ($p=0.018$), were less likely to have AECOPD ($p=0.0001$), to urgent call their General Practitioner (GP) after the first ($p=0.013$) and be admitted to the emergency room ($p=0.0003$) compared to the controls. The mortality rate did not differ between groups ($p=0.148$).
Halpin <i>et al</i> [27]	intervention n=40: age=68.5±1.5 FEV ₁ =48±4% female=26% Control n=39: age=70.2±1.6 FEV ₁ =54±3% female=26%	PRO (breathlessness, cough, congestion and fatigue). It was a 4-month RCT where an increase of >2 points in the mean of the summed raw daily scores on the last two days compared to the previous five days resulted in flagging the patient data and subsequently an alert call to the patient would be made.	Exacerbation frequency was greater on predicted high-risk vs non-high-risk days (0.086 ± 0.010 v 0.055 ± 0.010). However, there was no statistically significant difference in exacerbation frequency, duration, or severity and no significant differences in SGRQ scores.
Dinesen <i>et al</i> [28]	Intervention n=57: age=68 (45:82) FEV ₁ =0.90 (0.26:2.09)L Control n=48: age=68 (46;89) FEV ₁ =0.93 (0.33;2.13)L	BP, pulse rate (PR), weight, SpO ₂ , and lung function. An RCT in which a healthcare professional such as a GP, nurse, or doctor at a healthcare centre or hospital could assess and monitor the patient's data and provide advice to the patient.	The intervention group had a reduced mean hospital admission rate compared to the control (0.49 vs 1.17 ($p=0.041$)). However, there was no statistically significant difference in the length or cost of admission.

De San Miguel <i>et al</i> [29]	Intervention n=36: age=71(54-88) female=61% Control n=35: age=74(57-87) female=43%	HealthHUB™: BP, weight, temperature, PR, SpO ₂ and general state of health, daily. This was an RCT where the data was monitored daily by a nurse. Any deviations outside an individual's normal parameters (specified by their GP or specialist) trigger an alert. The nurse would contact the participant and provide advice or recommend an appointment to visit the GP.	In the intervention group, there was a non-statistically significant reduction in emergency department presentations, hospital admissions and length of stay compared to the control group. However, the reduction in health service use resulted in annual cost savings of \$2,931 per person in the intervention group compared to the control.
Jehn <i>et al</i> [30]	Intervention n=32: age=64.1±10.9 FEV ₁ =50.2±15.0% female=19% Control n=30: age=69.1±9.2 FEV ₁ =52.6±17.5% female=27%	Daily COPD Assessment Test (CAT), daily lung function, and weekly 6MWT (accelerometry). This was a 9-month, RCT where patient data was transmitted directly to the study centre which was reviewed by a physician daily.	The intervention compared to controls had fewer exacerbations (7 vs. 22; <i>p</i> =0.012), shorter hospital stays (34 vs. 97 days), and improvement in CAT score (-2.9±4.5 vs +4.4±5.7; <i>p</i> =0.001). However, there was no significant difference in visits to primary care, lung function or 6MWT.

APPENDIX B. NARRATIVE SYNTHESES

Pedone <i>et al</i> [31]	Intervention n=50: age=74.1±6.4 FEV ₁ =52.5±14.9% female=28% Control n=49: age=75.4±6.7 FEV ₁ =55.4±15.8% female=57%	Wristband monitor (HR, PA, temperature, and galvanic skin response), and SpO ₂ . This was a single-centre, unmasked, RCT with 9 months of follow-up. Each parameter was measured every three hours and was evaluated every day by a respiratory physician. Alerts were generated when a measurement was outside the predefined range causing the physician to contact the patient for appropriate management.	The intervention group experienced fewer single events and a reduction in the incidence of multiple events with an incidence rate ratio (IRR) of 0.67 (0.32 to 1.36, 95% confidence interval (CI)) but with an average length of stay of 9.7 days in the intervention group and 6.9 days in the control group.
Pinnock <i>et al</i> [32]	Intervention n=128: age=75±8.8 FEV ₁ =44.9±18.8% female=59% Control n=128: age=68.4±8.4 FEV ₁ =40.0±17.0% female=51%	PRO (dyspnea, sputum purulence, sputum volume, cough, wheeze) medication self-report, SpO ₂ , and upper respiratory infection and fever identification questionnaires. This was a 12-month, researcher-blinded, RCT. Patient data was monitored daily by a supporting clinical team (specialist respiratory team, a nurse specialist, or a trained call handler) and alerts were generated for no submission or a high symptom score resulting in action by the monitoring clinician.	The number of days to admission, the mean number of admissions, and the duration of admissions were not statistically different between groups. There was no significant effect on SGRQ scores between groups or on other questionnaire outcomes.

Sorknaes <i>et al</i> [33]	Intervention n=132: age=71±10 FEV ₁ =33±13% female=60% Control n=134: age=72±9 FEV ₁ =37±14% female=62%	PR, SpO ₂ , and lung function. This was over a year-long, two-site RCT. Patients were monitored for 7 days following admission and had daily remote consultations by video calls with a nurse who collected the patient's measures. The nurse could organise rapid treatment in consultation with a respiratory physician, the patient's general practitioner and/or the home care system if needed.	There was no significant difference in the total mean number of hospital readmissions after 26 weeks and there was no statistically significant difference between the two groups in mortality, time to readmission, the mean number of total hospital readmissions, the mean number of readmissions with AECOPD, the mean number of total hospital readmission days or mean number of readmission days with AECOPD.
Calvo <i>et al</i> [34]	Intervention n=29: age=75.0±11.7 FEV ₁ =38.3±11.9% female=24% Control n=30: age=72.7±9.3 FEV ₁ =37.1±10.8% female=26%	BP, SpO ₂ , and HR and peak expiratory flow (PEF) three times a week. This was a cluster assignment, controlled trial study design where the patient data was received, monitored, assessed and followed up by the clinical monitoring centre through a traffic light system application. Green (no further action), Yellow ("technical alert" - measurements had not been taken or had not been received), and Red ("clinical alert" - a measurement exceeded the predefined limits, verification initiates an escalation and clinical response).	After 7 months of monitoring and follow-up, there was a significant reduction ($p=<0.05$) in emergency department visits (20 in the intervention group vs 57 in the control group), hospitalisations (12 vs 33), duration of hospital stay (105 vs 276 days), and need for non-invasive mechanical ventilation (0 vs 8), and time to the first severe exacerbation (141 vs 77).

APPENDIX B. NARRATIVE SYNTHESES

McDowell <i>et al</i> [35]	Intervention n=55: age=69.8±7.1 $FEV_1=45.5\pm13.7\%$ female=58% Control n=55: age=70.2±7.4 $FEV_1=43.4\pm19.9\%$ female=55%	BP, HR, SpO ₂ and PRO (dyspnea, cough, sputum, tiredness). This was a 6-month two-centre, RCT. A trend report was created to set normal limits, after this, daily data was reviewed by a nurse who compared it to the set limits. An alert was generated if clinical and symptom observations were outside the limits. A nurse would then contact the patient, repeat the monitoring, and would escalate based on the results.	SGRQ scores improved significantly in the intervention group. However, there were no significant differences between EQ-5D scores, HADS depression scores, GP activity, emergency department visits, hospital admissions or exacerbations.
Ringbæk <i>et al</i> [36]	Intervention n=141: age=69.8±9.0 $FEV_1=34.9\pm13.3\%$ female=61% Control n=140: age=69.4±10.1 $FEV_1=33.8\pm12.0\%$ female=45%	Lung function, SpO ₂ , weight, and PRO (dyspnea, sputum colour, volume, and purulence). In this randomised controlled trial, patient data was transferred to a call centre at each participant's local hospital and was automatically categorised and prioritised. The call centres were staffed by a trained respiratory nurse who could confer the patient's data with a specialist in respiratory medicine at the hospital if values were alarming.	No statistically significant difference was found in hospital admissions for COPD, time to first admission, or all-cause hospital admissions. However, there was a significant difference found in the severity of exacerbations - the intervention group had more moderate exacerbations (treated with oral corticosteroids and/or antibiotics), whereas the control group had more visits to outpatient clinics.

Cordova <i>et al</i> [37]	Intervention n=34: age=64±6 FEV ₁ =31±13% female=50% Control n=33: age=63±8 FEV ₁ =32±15% female=73%	Daily PEF and PRO (dyspnea, and sputum quantity, colour, and consistency, cough, wheeze, sore throat, nasal congestion, and high temperature). A randomised, unblinded, parallel-group trial. A symptom algorithm detected changes from initial participant symptom values. If the symptom score generated by the algorithm reached or exceeded the predetermined threshold participants contacted a number that was available 24h/day, 7 days a week, staffed by nurses and pulmonologists. Exacerbations were treated according to the GOLD guidelines.	There were no differences in hospitalisation rates, hospital duration, or mortality in the intervention vs. control (35 vs. 44 ($p=0.63$)), (392±30 vs. 463±32 days ($p=0.8$)), (n=6 versus n=2 ($p=0.25$)). The intervention group's PEF, Borg dyspnea score, DUKE activity status index, and SF-36 significantly improved and were sustained for up to 24 months but were unchanged in the control.
------------------------------	---	--	---

APPENDIX B. NARRATIVE SYNTHESES

Ho <i>et al</i> [38]	Intervention n=53: age=81.4±7.8 $FEV_1=62\pm23\%$ female=19% Control n=53: age=79.0±9.6 $FEV_1=62\pm21\%$ female=28%	Self-report (SpO ₂ , temperature, BP) and PRO (disease-related symptoms). In this RCT, the data was processed according to a predefined algorithm, if the algorithm detected exacerbation signs a warning was generated and nurses and attending pulmonary physicians received a notification to respond to the situation. The patient's data was assessed which could result in referral to the clinic or emergency department.	At six months, the probability of COPD-related readmission was significantly lower in the intervention group (hazard ratio=0.42; 95% CI=0.19–0.92). Time to readmission was significantly increased in the intervention group compared to controls ($p=0.026$). There was also a significant reduction in the number of all-cause readmissions (0.23 v. 0.68 per patient; $p=0.002$) and emergency room visits (0.36 vs 0.91 per patient; $p=0.006$).
-------------------------	---	---	---

Vianello <i>et al</i> [39]	Intervention n=230: age=75.96±6.54 FEV ₁ =41.9±8.6% female=29% Control n=104 age=76.48±6.16 FEV ₁ =41.9±8.30% female=27%	Self-report every other day (HR and SpO ₂) and a single measurement of SpO ₂ once a day during periods of clinical worsening. This was a pragmatic, unblinded, parallel-group, two-arm, 12-month RCT. Operators reviewed the patient data and if values were outside of the patient's "normal" range, the patient was contacted for a second measurement. If also outside the normal range clinical staff were alerted and would contact the patient to check for worsening and possible intervention (medication modification, nurse home visit, pulmonary appointment, or emergency department admission).	SF36 Physical and Mental Component Summary scores, HADS, and hospital admission rates did not significantly differ between the intervention and control groups. However, the readmission rate was significantly lower in the intervention group (IRR=0.43 (95% CI 0.19–0.98); <i>p</i> =0.01 and 0.46 (95% CI 0.24–0.89); <i>p</i> =0.01).
-------------------------------	---	---	--

APPENDIX B. NARRATIVE SYNTHESES

<p>Farmer <i>et al</i> [40]</p>	<p>Intervention n=110: age=69.8±9.1 FEV₁=47.4±15.6% female=38%</p> <p>Control n=56: age=69.8±10.6 FEV₁=50.1±16.9% female=39%</p>	<p>EDGE: PRO, SpO₂, and HR.</p> <p>This was a multicentre, RCT of 12-month duration. An initial 6-week period of monitoring created a threshold for the participant safety alert.</p> <p>Participants continued to input their data daily. A respiratory clinician (nurse, physiotherapist, or doctor) reviewed a summary of the SpO₂, HR, and symptom diary module data twice weekly to ensure that data transmission was taking place and to deal with safety alerts. If data were not received or there were safety alerts, the participant record was accessed for review. If there was a clinically important change in the data, then the patient was contacted.</p>	<p>The estimated difference in SGRQ at 12 months in the intervention vs. control was -1.7 (95% CI -6.6-3.2 ($p=0.49$)) and the relative risk of hospital admission for the intervention was 0.83 (0.56-1.24, $p=0.37$) compared with the control. However, generic health status (EQ-5D, EuroQol 5-Dimension Questionnaire) was significantly better for the intervention (0.076, 95% CI 0.008-0.14, $p=0.03$). The median number of visits to general practitioners and practice nurses was lower for intervention, 4 vs. 5.5 ($P=0.06$) and 1.5 vs. 2.5 ($p=0.03$), respectively.</p>
--	---	--	--

Lilholt <i>et al</i> [41]	Intervention n=258: age=68.2±8.8 FEV ₁ =48.9±18.3% female=46% Control n=316: age=69.5±9.1 FEV ₁ =50.3±19.8% female=55%	Telekit: PRO (questions related to COPD exacerbations and symptoms), BP, HR, weight and SpO ₂ . This was a large-scale, pragmatic, two-level, cluster RCT with 12 months of follow-up. Data was sent to healthcare personnel, if the data adversely deviated from normal threshold values the patients were contacted. Patients were also contacted if the measurements were not carried out as agreed or the measurements were not received as expected.	In an intention-to-treat analysis (n=1225, with imputed data), the raw mean difference in the physical component summary from baseline to 12-month follow-up was -2.6±12.4 in the intervention group and -2.8±11.9 in the control group, the raw mean difference in mental component summary scores was -4.7±16.5 and -5.3±15.5 for intervention and control groups, respectively.
Rixon <i>et al</i> [42]	Intervention n=275: age=71.11 (SE=0.524) female=45% Control n=172: age=72.08 (SE=0.721) female=43%	SpO ₂ , BP, weight, and questions about health. The system demonstrator evaluation is one of the world's largest pragmatic cluster RCTs. Data was reviewed by healthcare professionals at a monitoring centre. Review of data would result in no response, requesting another reading, contacting the participant or their carer, a home visit by their community nurse or referral to a health service.	There was a lack of statistical significance for improvement in dyspnea, fatigue, SF12, EQ5D, anxiety, and depression but there was a significant improvement identified in the intention to treat analysis in emotional functioning and mastery.

APPENDIX B. NARRATIVE SYNTHESES

Kessler <i>et al</i> [43]	Intervention n=157: age=67.3±8.9 $FEV_1=37.8\pm12.4\%$ female = 30.6% Control n=162: age=66.6±9.6 $FEV_1=36.4\pm12.3\%$ female=30.2%	<p>Lung function, SpO₂, and HR were monitored daily, participants on long-term oxygen therapy also had daily oxygen use and respiration rate.</p> <p>COMET was a 12-month, randomised, open-design, parallel-group trial conducted at 33 centres. Information was transmitted by patients using a telephone-based questionnaire once per week, and each day of symptom worsening. An e-health platform enabled patient follow-up by case managers for early detection of exacerbations. Information was transmitted to hospital physicians via the e-platform to intervene when necessary.</p>	<p>The number of participants with exacerbations, exacerbation frequencies, time to first exacerbation, and all-cause hospitalisation days per year were similar for both intervention and control. However, the intervention had significantly fewer acute care hospitalisation days per year ($p=0.047$), a lower body mass index-airway obstruction-dyspnea-exercise tolerance (BODE) index ($p=0.01$) and a lower mortality rate (1.9% vs 14.2%; $p<0.001$).</p>
------------------------------	---	---	--

Soriano <i>et al</i> [44]	Intervention n=115: age=71±8.0 FEV ₁ =34.2±9.1% female = 21.7% Control n=114: age=71.3±8.9 FEV ₁ =32.2±8.8% female=17.5%	SpO ₂ , BP, lung function, and respiratory rate (RR) and oxygen therapy compliance (VisionOx®). The PROMETE II study was a 12-month, multicentre, non-blinded, RCT. The first four days were used as reference values and configured for alerts. The monitoring centre received the information as Red: "clinical alert" – measurements exceeded reference values, Yellow: "Technical Alert" - measurements missing, or Green: The measurements had been made and found to be within the reference. The "clinical alerts" were confirmed by a clinical questionnaire which resulted in notifying the relevant pulmonologist/local coordinator of the AECOPD.	There was no statistically significant difference at one year between the intervention and control in the proportion of participants who had an AECOPD, the number of all-cause deaths was similar, and total resource utilisation cost was similar. There was a trend towards a shorter duration of hospitalisation and intensive care in the intervention (18.9±16.0 and 6.0±4.6 days) compared to the control (22.4±19.5 and 13.3±11.1 days), though this was non-significant.
------------------------------	---	---	---

APPENDIX B. NARRATIVE SYNTHESES

Tupper <i>et al</i> [45]	Intervention n=141: age=69.8±9.0 $FEV_1=34.9\pm13.3\%$ female = 61% Control n=140: age=69.4±10.1 $FEV_1=33.8\pm12.0\%$ female=45%	PRO (dyspnea, sputum colour, volume, and purulence), lung function, SpO ₂ , and weight. The NetKOL study was an RCT that recruited patients from 4 hospitals. Measurements were conducted once a week for the first 4 weeks and then every 4 weeks. Measurements were performed every week and were transferred to a call centre at each participant's local hospital and automatically categorised and prioritised (green, yellow, or red). Concerning measurements resulted in a phone call from a specialist nurse either by phone or video consultation and possibly conferred with a respiratory specialist.	At 6 months, the 15D significantly improved (0.016 $p=0.03$; minimal clinically important difference 0.015) in the intervention group (compared to baseline), whilst there was no improvement in the control group –0.003 ($p=0.68$). However, there was no statistically significant change in the CAT score.
Walker <i>et al</i> [46]	Intervention n=135: age= 70(44–86) $FEV_1=49.4\%$ $(37.1\text{--}59.2)\%$ female = 34% Control n=158: age=71.0(65–76) $FEV_1=50.4\%$ $(38.0\text{--}63.9)\%$ female=34%	CHROMED: Within-breath respiratory mechanical impedance using forced oscillation technique. This was a multicentre, randomised, unblinded, parallel-group clinical trial where an algorithm generated respiratory alerts if a trend of worsening was detected (increase of inspiratory resistance, difference between inspiratory and expiratory reactance, or decrease of inspiratory reactance). The alert triggered contact with the study nurse to determine intervention.	The intervention did not affect time-to-first-hospitalisation, EQ-5D score, antibiotic prescriptions, hospitalisation rate, or questionnaire scores. In an exploratory analysis, the intervention was associated with fewer repeat hospitalisations (–54%; $p=0.017$).

Mínguez Clemente <i>et al</i> [47]	Intervention n=58: age=68±8 FEV ₁ =50±17% female=55% Control n=58: age=70±8 FEV ₁ =52±17% female=65%	ECG (leads I, II and III), SpO ₂ , HR, BP, temperature, and RR. The TELEMED-COPD study was an RCT with 2 parallel early discharge groups. Patients transmitted data twice daily. Thresholds were set for the measures. Deviations from the thresholds generated an alert in the form of a text message sent to a physician's telephone. The physician could contact the participant for intervention.	No statistically significant differences in baseline characteristics, time until first exacerbation, the number of exacerbations or costs between intervention and control groups. There was a significant decrease in the number of visits observed in the intervention compared to the control, 3.8±1 vs 5.1±2($p=0.001$).
Sink <i>et al</i> [48]	Intervention n=83: age=59.89±1.09 FEV ₁ =64±0.02% female=54% Control n=85: age=61.94±1.07 FEV ₁ =63±0.02% female=53%	EpxCOPD: breathing (better, same, worse). An eight-month RCT. A daily message from the EpxCOPD system asks "Are you breathing better than, worse than, or the same as yesterday?". Responding "better" or "the same" resulted in no. Responding "worse" triggers an alert to an assigned resident clinic. The resident would contact the participant and then counsel on how to return to normal breathing or order an appropriate intervention.	The intervention group's time-to-hospitalisation was statistically significantly longer than the control group's with a hazard ratio of 2.36 (95% CI=1.02–5.45, $p=0.0443$) and there were significantly fewer COPD-related hospitalisations in the intervention group compared to the control group (6 vs 16, respectively). The absolute risk reduction was 11.6% and the relative risk reduction was 61.7%. The number needed to treat was 8.62.

APPENDIX B. NARRATIVE SYNTHESES

Koff <i>et al</i> [49]	Intervention n=352: age=68.3±8.7 $FEV_1=36.2\pm12.9\%$ female=42.0% Control n=159: age=68.4±9.5 $FEV_1=38.2\pm14.7\%$ female=33.3%	iCare: COPD and exacerbation education, and RPM Health Buddy system (PRO, SpO ₂ , lung function, 6MWT (pedometer), and post-exertion SpO ₂). A prospective, quasi-RCT. Measurements were taken every weekday and were analysed by predetermined algorithms, and categorised into green (stable), yellow (caution), and red (possible decline in health status). Coordinators contacted participants with red flags and resolved clinical problems directly or by reaching the participant's primary care provider.	The intervention group saw improvements in total SGRQ by 7–9 units, had increases in 6MWT distance by 40m ($p<0.001$), reduced annual COPD-related urgent office visits by 76 visits per 100 participants ($p<0.0001$), identified unreported exacerbations, and decreased smoking ($p=0.01$) and had a reduction in symptoms, BODE index and oxygen titration ($p<0.05$) when compared with the control group. However, mortality, hospitalisation, intensive care visits, emergency department visits, and length of stay were not significantly different between groups.
------------------------	---	--	--

Rassouli <i>et al</i> [97]	All n=168: age=67(61-73) FEV ₁ =51 (16–115)% female=35%	PRO (dyspnea, sputum volume, sputum colour, cough, fever, emergency medication). This was a multicentre crossover RCT, involving 6 centres. Participants answered questions every morning. Answers were coded with different colours: not answered (grey), had answered zero or one question with ‘yes’ (green), had answered two or more questions with ‘yes’, but was green the previous day (yellow), or had answered two or more questions with ‘yes’ and had a yellow or red the previous day (red). Red indicated a possible AECOPD and resulted in participants being contacted by a physician for intervention (self-medication, consult GP, or hospital/ED admission).	The mean increase in CAT score in the intervention group vs. control was 1.8 vs. 3.6 points/year ($p=0.0015$), respectively. However, there was not a significant difference in the intervention vs. control in the ED visit rate, hospitalisation rate, number of AECOPD, days in hospital due to AECOPD, and cost.
-------------------------------	---	---	--

APPENDIX B. NARRATIVE SYNTHESES

Anderson <i>et al</i> [51]	Intervention n=110: age=70 (64–76) $FEV_1=34.5$ (28–44)% female=65% Control n=112: age=71 (65–76) $FEV_1=38.0$ (19–47)% female=57%	SpO ₂ , HR, PEF, weight, PRO (dyspnea, cough, sputum volume, and colour). The study was a 6-month, non-blinded, RCT. The monitoring took place every weekday for the first month and then three times a week. Thresholds were set from individual starting measurements. An algorithm generated alerts if a trend of worsening was detected: green (stable), yellow (one or more PRO indicated clinical worsening), and red alert (one or more measurements were abnormal). yellow or red alerts, triggered contact to the participant by the respiratory nurse on the same day. Treatment, frequent monitoring, outpatient appointments, contact with the patient's GP or hospitalisation could be started.	33% of the control group and 28% of the intervention group had an AECOPD hospitalisation (odds ratio (OR)=1.26, CI=0.71–2.23, $p=0.4$). No significant difference was seen in time to first hospitalisation (hazard ratio=1.23, CI=0.77–1.99, $p=0.4$). The number of hospitalisations was 66 in the control group and 42 in the intervention, (IRR=1.42, CI=1.04–1.95, $p=0.03$).
-------------------------------	---	---	---

Køpfli et al [52]	Intervention n=101: age=69±8.3 FEV ₁ =38.7±14.1% female=65.3% Control n=97: age=70±7.8 FEV ₁ =40.2±15.1% female=56.7%	SpO ₂ , HR, lung function, weight, and PRO (dyspnea, cough, sputum volume and sputum colour). A 6-month RCT with 24-month follow-up. The measures were sent every weekday for the first month, and three times a week during the last 5. A threshold was set for each measurement in each participant. Based on the patient's reported measurements and symptoms there were three categories: green (stable), yellow (change in symptoms), and red (change in physiological measurement). The respiratory nurses evaluated the data and with a respiratory physician determined additional treatment or follow-up.	There was no significant difference in SGRQ, HADS-Anxiety, or HADS-depression. No difference was seen at 12–24 months follow-up.
-------------------	--	---	--

Table B.2: Narrative synthesis of remote patient monitoring and machine learning to predict AECOPD.

Authors	Participants	Measures and study protocol	Results
Jensen, et al [53]	Participants n=57: male(n=23): age=69.6 years, FEV ₁ =1.10L, female(n=34): age=67.2 years, FEV ₁ =0.750L	Weight, BP, SpO ₂ and, lung function. This was a retrospective cohort study to train and validate models generated using a linear discriminant analysis (LDA) algorithm on data taken from the TELEKAT trial.	Evaluating the model in a leave-one-out cross-validation, the model has a sensitivity of 70%, a specificity of 95%, and an area under the receiver operating characteristic curve (AUROC) of 0.73 for predicting AECOPD.

APPENDIX B. NARRATIVE SYNTHESES

Heijden, <i>et al</i> [54]	Participants n=5: n/a	FEV ₁ , SpO ₂ , PRO (dyspnea, sputum production, sputum colour, cough, wheezing, activities, malaise, fever). This was a retrospective model evaluation study. A Bayesian network model was used to predict the risk of exacerbation. Patient data were collected weekly if the patient is at low risk of exacerbation, the check-in becomes daily with increasing risk.	The model performance on a different data set includes an AUROC of 0.87, a sensitivity of 0.88, a false-positive rate of 0.2 and an accuracy of 0.81 for predicting AECOPD.
Bellos, <i>et al</i> [55]	Participants n=30: n/a	CHRONIOUS: ECG, RR, steps, standing and lying time, cough, snoring, temperature, BP, SpO ₂ , lung function, blood glucose, weight, PRO (dyspnea, cough, sputum), and demographics (lifestyle, mental status). This was a retrospective machine learning model development and validation study. Clinicians labelled the severity of the patient's situation to make the classes. A stratified tenfold cross-validation was used to randomise the records and split the available instances into training and testing datasets.	Using correlation-based feature subset selection and a hybrid classification system combining random forest (RF), support vector machine (SVM), and a rule-based system achieved an accuracy of 94% for current disease severity estimation, which can be used to recognise an abnormal health episode or an AECOPD.

Fernández-Granero, <i>et al</i> [56]	Participants n=16: age=70.2±6.6 GOLD=>I female=7%	PRO through the Automated Questionnaire for the early detection of COPD Exacerbations (AQCE) (general health status, cough, phlegm, dyspnea, sleep, cold symptoms, lung sounds and coordination test). A retrospective study where participants answered questions daily and a probabilistic neural network (PNN) was used for classification.	Ten-cross validation was used to evaluate the model which has an accuracy of 88.3%, a sensitivity of 80.5%, a specificity of 94.34%, a positive-predictive value (PPV) of 91.67%, and a negative-predictive value (NPV) of 86.21% and can detect AECOPD on average 4.8 ± 1.8 days to the event.
Shah, <i>et al</i> [58]	Participants n=18: age=71±9 SpO2=93.5±4.1 female=50%	Daily PRO (general health status, breathlessness, wheeze, cough, sputum colour, presence of cold, sleep breathing), SpO2, HR, RR, and photoplethysmography (PPG) waveform. This was a retrospective pilot study that developed a model through multivariate novelty detection using Parzen windows.	The model achieved an AUROC of 0.91 for AECOPD.
Fernández-Granero, <i>et al</i> [59]	Participants n=16: age=70.2±6.6 years GOLD=C/D female=71%	EXACT: PRO (breathlessness, cough, sputum, chest symptoms, difficulty bringing up sputum, fatigue, sleep disturbance, and health anxiety). Data were collected daily during a 6-month pilot.	A PNN classifier performed best with an accuracy of 88.3%, a sensitivity of 80.5% and a specificity of 94.3%. 100% of event-based AECOPD were detected 6.3±3.3 days before onset. 80.5% of symptoms-based AECOPD were predicted (33 out of 41) with 4.8±1.8 days before onset.

APPENDIX B. NARRATIVE SYNTHESES

Fernández-Granero, <i>et al</i> [60]	Participants n=16: age=70.2±6.6 years GOLD=C/D female=71%	Respiratory sounds through a respiratory sensor embedded in a self-tailored housing. Data were collected daily during a 6-month pilot.	This was a retrospective model development study. Principal component analysis (PCA) and an SVM classifier achieved a sensitivity of 73.76%, a specificity of 97.67%, a PPV of 84.66% and an NPV of 95.53% for predicting AECOPD. 75.8% exacerbations were detected early 5±1.9 days in advance of intervention for AECOPD.
Mohktar, <i>et al</i> [61]	Participants n=21: age=70.2±6.6 years COPD=moderate/severe female=71%	Lung function, SpO ₂ , RR, HR, temperature and weight. This was a retrospective cohort study where patient data was monitored daily over 11 months and was used to train a classification and regression tree (CART) to predict if patients are at high risk or low risk of exacerbation.	The CART model has an accuracy of 71.8%, specificity of 80.4%, and 61.1% sensitivity, due to the small dataset size its performance was evaluated using leave-one-out cross-validation for AECOPD.
Sanchez-Morillo, Fernandez-Granero, and León Jiménez [62]	Participants n=16: age=70.2±6.6 years GOLD=C/D female=71%	PRO via AQCE: general health status, cough, phlegm, dyspnea, sleep, cold symptoms, lung sounds, and coordination test. Participants completed the test daily in the evening. Healthcare contact or use of a rescue pack was labelled as an exacerbation. K-means clustering with five-fold cross-validation to select the R-value was used to generate an exacerbation prediction model.	Using tenfold cross-validation to evaluate, the model achieved an accuracy of 84.7%, a sensitivity of 74.6%, a specificity of 89.7%, PPV 78.5%, NPV 87.6% for AECOPD.

Christian Riis, <i>et al</i> [63]	Participants n=108: age=70.2±9.4 years FEV ₁ =0.76±0.33 L female=52%	PRO, BP, SpO ₂ , and PR. This a retrospective model development study where patients either did daily or weekly measurements depending on disease severity. K-nearest neighbours (KNN) was used to develop the prediction model.	A leave-one-out cross-validation was used to evaluate and the best model has a sensitivity of 73%, a specificity of 74%, a PPV of 69% and an NPV of 78% for AECOPD.
Shah, <i>et al</i> [64]	Participants n=110: age=69.8±9.1 years FEV ₁ =47.4±15.6% female=38%	PR, RR, and SpO ₂ . It was a retrospective cohort study using data from the EDGE clinical trial where participant data was used to train a logistic regression (LR) classifier.	tenfold cross-validation was repeated 10 times to evaluate the model which has an AUROC of 0.682% (95% CI 0.681-0.682), and an 80% sensitivity and 36% specificity, or 60% sensitivity and 68% specificity for AECOPD.
Fernandez-Granero, <i>et al</i> [65]	Participants n=16: age=70.2±6.6 years GOLD=C/D	Respiratory sounds were recorded daily with an electronic sensor, this was a retrospective cohort study to enable an automatic prediction of symptom-based exacerbations, recorded data were used to train and validate an RF classifier.	With a 4.4 days margin before the onset of AECOPD, the sensitivity of the model is 78.1%, the specificity is 95.9%, and the PPV is 94.1%
Kronborg, <i>et al</i> [66]	Participants n=57: age=69.23±7.84 years female=56.15%	PRO (symptoms, cough, mucus, shortness of breath, inhaler use, antibiotic use), SpO ₂ , PR, weight, and BP. This was a retrospective validation study using data collected from the TeleCare North trial using LR to generate exacerbation prediction models.	The best-performing model achieved an AUROC of 0.74 for AECOPD which was based on two-fold patient-dependent cross-validation.

APPENDIX B. NARRATIVE SYNTHESES

Nunavath <i>et al</i> [67]	Participants n=94: age= 69.2 ± 9.2 years $FEV_1 = 38.1 \pm 14.8\%$ female=49%	Daily measurements of a symptom-specific questionnaire, SpO ₂ , and results of automatically generated health status overview. This was a retrospective cohort study using Deep artificial neural networks; feed-forward neural networks (FFNN) for the classification of COPD patient's health category, and Long short-term memory (LSTM) for early prediction of COPD exacerbations and subsequent triage.	the FFNN model can reproduce diagnosed health conditions with an accuracy of 92.86% and the LSTM model was able to predict COPD patients' health conditions one-day with an accuracy of 84.12%.
Orchard, <i>et al</i> [68]	Participants n=135: age= 69.8 ± 9.1 years female=40%	Daily symptoms, physiological measures, and medication data, with baseline demography, COPD severity, quality of life, and hospital admissions linked with meteorological data. This was a retrospective cohort study using data from the Telescot COPD telemonitoring program to construct predictive models fitted to training sets of patient data using multitask neural networks (MTNN).	The MTNN based on 57,150 episodes has an AUROC of 0.74 (95% CI 0.67-0.80), adding weather data did not improve the predictive performance (AUROC 0.74, 95% CI 0.69-0.79).

Boer, <i>et al</i> [69, 70]	Intervention n=43: age= 69.3±8.8 years FEV ₁ =53.0±21.5% female=42% Control n=44: age= 65.9±8.9 years FEV ₁ =52.1±19.8% female=34%	12 yes-or-no PRO, SpO ₂ , FEV ₁ , and temperature (forehead thermometer). The most recent study was a multi-centre, 2-arm RCT for 12 months where participants used either the smart tool ACCESS which uses a Bayesian network model to predict AECOPD or a paper action plan.	In the initial validation study ACCESS has a sensitivity of 97.4% (95% CI 92.0–99.3), specificity of 65.6% (95% CI 63.5–67.6), and a PPV and negative predictive value (NPV) of 13.4% (95% CI 11.2–15.9) and 99.8% (95% CI 99.3–99.9), respectively. In the RCT there was not a statistically significant difference between the intervention group and the control group in exacerbation-free weeks (30.6±13.3 vs 28.0±14.8 weeks, respectively) or health status, self-efficacy, self-management behaviour, and health care utilisation
Jin, <i>et al</i> [71]	Participants n=22: age=76.14±11.60 years female=18.19%	Non-invasive ventilator: airflow, pressure and SpO ₂ . It was a prospective cohort study where there was continuous monitoring by the home ventilators which were transmitted to a data platform. SVM, RF and LDA were used to generate exacerbation prediction models.	To evaluate the models, they used five-fold cross-validation. The best-performing model is the LDA model with an accuracy of 74.5%, a sensitivity of 77.6%, and a specificity of 42.9%.

APPENDIX B. NARRATIVE SYNTHESES

Iadanza, <i>et al</i> : n/a [72]	Participants n=424	FEV ₁ , forced vital capacity (FVC), slow Vital Capacity (SVC), FEV ₁ /FVC ratio, FEV ₁ /SVC ratio, forced expired flow at 25–75%, PEF, FVC, total lung capacity (TLC), residual volume (RV), functional residual capacity (FRC), expiratory reserve volume (ERV), diffusing capacity of the lungs for carbon monoxide (DLCO), alveolar volume (VA), and DLCO/VA. A neural network, SVM and C5.0 were used to generate the AECOPD prediction models	C5.0 performed best across mild, moderate and severe COPD. The model performed best in severe COPD achieving a sensitivity of 98.9%, a specificity of 96.2%, and an accuracy of 97.4% for AECOPD.
Kronborg, <i>et al</i> [73]	Participants n=9: age=70.2±9.4 years Modified Medical Research Council (mMRC) dyspnea=3.8±0.8	SpO ₂ , PR, and BP. This was a retrospective validation study using data collected from a telehealth project in Denmark. A one-layer model was based on features extracted from SpO ₂ , PR, and BP. A two-layer model was based on features extracted from the model-estimated probability of exacerbation based on probabilities from the one-layer model. Subsequently, these features were used to train an exacerbation prediction model using LR, LDA, SVM, KNN, naive Bayes, CART, and RF.	SVM with radial basis function on the two-layer model performed the best with an AUROC of 0.95 and a sensitivity of 0.94.

Patel <i>et al</i> [74]	Participants n=90: age=69±8 years FEV ₁ =50±20% female=50%	Using COPDPredict™ App, PRO, lung function, CRP (finger-prick testing). This was a prospective observational study where patients performed a daily wellbeing assessment and weekly spirometry, a decision tree could prompt further testing of spirometry and/or CRP test, and the exacerbation prediction model was trained on a 2-week learning phase (baseline entries: daily PRO, every 3rd-day FEV ₁ , and day 1st and last day CRP).	Compared to the 6 months pre-COPDPredict™, hospitalisations were reduced by 98% (90 vs 2, <i>p</i> <0.001), the model has a sensitivity of 97.9% (95% CI 95.7–99.2), specificity of 84.0% (95% CI 82.6–85.3), PPV and NPV of 38.4% (95% CI 36.4–40.4) and 99.8% (95% CI 99.5–99.9), respectively.
Wu, <i>et al</i> [75]	Participants n=67: age=66.62±11.38 years FEV ₁ (30-79%)=66% female=12%	Environmental (temperature, humidity, particulates), physiological (HR, walking steps, calories consumption, deep sleep time, light sleep time, rapid eye movement time, awake time), PRO (CAT scores, mMRC dyspnea scale, quality of life questionnaire). This was a prospective cohort study using a decision tree, RF, KNN clustering, LDA, adaptive boosting and deep neural network (DNN) were used to generate the AECOPD prediction models.	The DNN model performed the best at predicting AECOPD within 7 days with an f1 of 0.923, an accuracy of 92.1%, a sensitivity of 90.4%, a specificity of 94% and an AUROC 0.964.

Chmiel, <i>et al</i> [76]	Participants n=2374: age=60-79 years	PRO (CAT score, symptom score) demographics (Age, gender, smoking years, smoking status). This was a retrospective study using patient-entered self-management app data from the myCOPD app to generate AECOPD prediction models. LR and RF machine learning algorithms were used.	The RF with Youden's J statistic as threshold achieved an AUROC of 0.727, a sensitivity of 0.755, and a specificity of 0.629.
Wu, <i>et al</i> [77]	Participants n=106: n/a	Air pollution, PA, PRO and PEF. This was a prospective cohort study with data collected over 24 months. Through the feature importance map and SHAP module, then applying backward elimination cost-effective features were selected and used to train an RF model.	The cost-effective RF model performed the best at predicting AECOPD within 7 days with an accuracy of 88.6%, a sensitivity of 77.8%, a specificity of 94.9% and an f1 of 0.833.

Appendix C

SPACE Study I interview schedules

C.1 Focus group - exacerbation model accuracy

How do you use myCOPD?

1. In what ways do you use myCOPD?
2. How do you manage exacerbations?
3. Do you use myCOPD for exacerbation management?

Key Questions - breakout rooms

1. What do you think about this accuracy?
2. How would you use this information?
3. Would this information be useful to you?
4. What would you do/do differently with this information?
5. Would you use a sensor if it could make the models better at predicting flare-ups?

Key Questions - main room

Share your thoughts?

Would you use a sensor if it could make the models better at predicting flare-ups?

C.2 Focus group - exacerbation warning

How do you use myCOPD?

1. In what ways do you use myCOPD?

APPENDIX C. SPACE STUDY I INTERVIEW SCHEDULES

2. How do you manage exacerbations?
3. Do you use myCOPD for exacerbation management?

Key Questions - breakout rooms

1. What does a flare-up warning/alert look like to you?
2. What would a warning/alert say?
3. When do you want a warning/alert?
4. How might a flare-up warning/alert affect your management of COPD?
5. How do you think a warning/alert might affect your mood?

Key Questions - main room

What ideas do we have for a flare-up warning/alert?

What are the views on these different ideas?

Key Questions

1. Do these alerts share any similarities or differences to your ideas?
2. Would you respond differently to any of these warnings/alerts?
3. Do you feel any of these alerts would be useful to you?
4. If you could make any changes what would that look like?

Appendix D

SPACE Study II interview schedules

D.1 Focus group - sensors

How do you use myCOPD?

1. In what ways do you use myCOPD?
2. How do you manage exacerbations?
3. Do you use myCOPD for exacerbation management?

Key Questions - breakout rooms

1. Would you use a sensor if you knew it could improve an early flare-up warning?
2. Can you see any problems with using the sensors?
3. Do you prefer any one of the sensors over the others? Is there a reason for that preference?
4. Is there a sensor you particularly don't like and why?
5. Can you see yourself using any one or all of these sensors regularly? (Why do you think this is?)
6. What would encourage you to use a sensor regularly? (daily? Weekly?)

Key Questions - main room

Share your thoughts?

D.2 Daily interviews

Did you use your sensor today?

If yes

APPENDIX D. SPACE STUDY II INTERVIEW SCHEDULES

Did your sensor cause your behaviour to change in any way in relation to managing your COPD?

Did the sensor negatively impact you in any way?

If no

Why did you not use your sensor today?

Did you log into myCOPD today?

If yes

What did you do when you logged on?

If no

Why did you not log in today?

Do you plan to use your sensor tomorrow?

If yes

What are the benefits and motivations for using your sensor?

If no

Why do you think you will not use it?

D.3 Midpoint and final interviews

Have you used your sensors every day since we last spoke?

If yes

How did you use the sensors?

Have you noticed any changes in your behaviour since you started using your sensors?

D.3. MIDPOINT AND FINAL INTERVIEWS

Did the sensors negatively impact you in any way or have you had any issues using the sensors?

If no?

Why did you not use your sensors every day?

Did you log into myCOPD every day since we last spoke?

If yes

What did you do when you logged on?

Have you used myCOPD differently since you started using these sensors?

If no

Why did you not log in every day?

Do you plan to continue using your sensor every day over the next two weeks/would you continue using your sensors over a longer period every day?

If yes

Why do you think you will continue to use the sensor on a daily basis?

If

Why do you think you will not continue to use the sensors regularly?

Bibliography

- [1] Soriano JB, Kendrick PJ, Paulson KR, Gupta V, Abrams EM, Adedoyin RA, et al. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Respiratory Medicine.* 2020;8(6):585-96.
- [2] Bennet J, Russell R, Echevarria C, Quigley E, Aliabadi-Oglesby T, Costello J, et al.. Acute exacerbation of chronic obstructive pulmonary disease;. Available from: <https://bestpractice.bmjjournals.com/topics/en-gb/3000086>.
- [3] Roberts C, Lowe D, Bucknall C, Ryland I, Kelly Y, Pearson M, et al. Clinical audit indicators of outcome following admission to hospital with acute exacerbation of chronic obstructive pulmonary disease. *Thorax.* 2002;57(2):137-41.
- [4] Suissa S, Dell'Aniello S, Ernst P. Long-term natural history of chronic obstructive pulmonary disease: severe exacerbations and mortality. *Thorax.* 2012;67(11):957-63.
- [5] Soler-Cataluna J, Martínez-García MÁ, Sánchez PR, Salcedo E, Navarro M, Ochando R. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax.* 2005;60(11):925-31.
- [6] McGhan R, Radcliff T, Fish R, Sutherland ER, Welsh C, Make B. Predictors of rehospitalization and death after a severe exacerbation of COPD. *Chest.* 2007;132(6):1748-55.
- [7] Whittaker H, Van Ganse E, Dalon F, Nolin M, Marant-Micallef C, Pison C, et al. Differences in severe exacerbations rates and healthcare utilisation in COPD populations in the UK and France.

BIBLIOGRAPHY

- BMJ open respiratory research. 2022;9(1):e001150.
- [8] Perera WR, Hurst JR, Wilkinson TM, Sapsford RJ, Müllerova H, Donaldson GC, et al. Inflammatory changes, recovery and recurrence at COPD exacerbation. European Respiratory Journal. 2007;29(3):527-34.
- [9] Anzueto A. Impact of exacerbations on COPD. European Respiratory Review. 2010;19(116):113-8.
- [10] Makris D, Moschandreas J, Damianaki A, Ntaoukakis E, Siafakas N, Emili JM, et al. Exacerbations and lung function decline in COPD: new insights in current and ex-smokers. Respiratory medicine. 2007;101(6):1305-12.
- [11] Miravitles M, Ferrer M, Pont A, Zalacain R, Alvarez-Sala J, Masa F, et al. Effect of exacerbations on quality of life in patients with chronic obstructive pulmonary disease: a 2 year follow up study. Thorax. 2004;59(5):387-95.
- [12] Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. New England Journal of Medicine. 2010;363(12):1128-38.
- [13] Celli BR, MacNee W, Agusti A, Anzueto A, Berg B, Buist AS, et al. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. European Respiratory Journal. 2004;23(6):932-46.
- [14] Iheanacho I, Zhang S, King D, Rizzo M, Ismaila AS. Economic burden of chronic obstructive pulmonary disease (COPD): a systematic literature review. International journal of chronic obstructive pulmonary disease. 2020;15:439.
- [15] Alqahtani JS, Aquilina J, Bafadhel M, Bolton CE, Burgoyne T, Holmes S, et al. Research priorities for exacerbations of COPD. The Lancet Respiratory Medicine. 2021;9(8):824-6.
- [16] Wilkinson TM, Donaldson GC, Hurst JR, Seemungal TA, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2004;169(12):1298-303.
- [17] Jang S, Kim Y, Cho WK.

- A systematic review and meta-analysis of telemonitoring interventions on severe COPD exacerbations.
International Journal of Environmental Research and Public Health. 2021;18(13):6757.
- [18] Kruse C, Pesek B, Anderson M, Brennan K, Comfort H, et al.
Telemonitoring to manage chronic obstructive pulmonary disease: systematic literature review.
JMIR medical informatics. 2019;7(1):e11496.
- [19] Malasinghe LP, Ramzan N, Dahal K.
Remote patient monitoring: a comprehensive study.
Journal of Ambient Intelligence and Humanized Computing. 2019;10:57-76.
- [20] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al.
Artificial intelligence in healthcare: past, present and future.
Stroke and vascular neurology. 2017;2(4).
- [21] Bourne S, DeVos R, North M, Chauhan A, Green B, Brown T, et al.
Online versus face-to-face pulmonary rehabilitation for patients with chronic obstructive pulmonary disease: randomised controlled trial.
BMJ open. 2017;7(7):e014580.
- [22] Agustí A, Celli BR, Criner GJ, Halpin D, Anzueto A, Barnes P, et al.
Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary.
American Journal of Respiratory and Critical Care Medicine. 2023;207(7):819-37.
- [23] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group* t.
Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.
Annals of internal medicine. 2009;151(4):264-9.
- [24] De Toledo P, Jiménez S, del Pozo F, Roca J, Alonso A, Hernandez C.
Telemedicine experience for chronic care in COPD.
IEEE transactions on information technology in biomedicine. 2006;10(3):567-73.
- [25] Koff P, Jones RH, Cashman JM, Voelkel NF, Vandivier R.
Proactive integrated care improves quality of life in patients with COPD.
European Respiratory Journal. 2009;33(5):1031-8.
- [26] Vitacca M, Bianchi L, Guerra A, Fracchia C, Spanevello A, Balbi B, et al.
Tele-assistance in chronic respiratory failure patients: a randomised clinical trial.
European Respiratory Journal. 2009;33(2):411-8.

BIBLIOGRAPHY

- [27] MG Halpin D, Laing-Morton T, Spedding S, L Levy M, Coyle P, Lewis J, et al. A randomised controlled trial of the effect of automated interactive calling combined with a health risk forecast on frequency and severity of exacerbations of COPD assessed clinically and using EXACT PRO. Primary Care Respiratory Journal. 2011;20(3):324-31.
- [28] Dinesen B, Haesum LK, Soerensen N, Nielsen C, Grann O, Hejlesen O, et al. Using preventive home monitoring to reduce hospital admission rates and reduce costs: a case study of telehealth among chronic obstructive pulmonary disease patients. Journal of telemedicine and telecare. 2012;18(4):221-5.
- [29] De San Miguel K, Smith J, Lewin G. Telehealth remote monitoring for community-dwelling older adults with chronic obstructive pulmonary disease. Telemedicine and e-Health. 2013;19(9):652-7.
- [30] Jehn M, Donaldson G, Kiran B, Liebers U, Mueller K, Scherer D, et al. Tele-monitoring reduces exacerbation of COPD in the context of climate change—a randomized controlled trial. Environmental Health. 2013;12:1-8.
- [31] Pedone C, Chiurco D, Scarlata S, Incalzi RA. Efficacy of multiparametric telemonitoring on respiratory outcomes in elderly people with COPD: a randomized controlled trial. BMC health services research. 2013;13(1):1-7.
- [32] Pinnoch H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. Bmj. 2013;347.
- [33] Sorknaes AD, Bech M, Madsen H, Titlestad IL, Hounsgaard L, Hansen-Nord M, et al. The effect of real-time teleconsultations between hospital-based nurses and patients with severe COPD discharged after an exacerbation. Journal of telemedicine and telecare. 2013;19(8):466-74.
- [34] Calvo GS, Gómez-Suárez C, Soriano J, Zamora E, González-Gamarra A, González-Béjar M, et al. A home telehealth program for patients with severe COPD: the PROMETE study. Respiratory medicine. 2014;108(3):453-62.

BIBLIOGRAPHY

- [35] McDowell JE, McClean S, FitzGibbon F, Tate S.
A randomised clinical trial of the effectiveness of home-based health care with telemonitoring in patients with COPD.
Journal of Telemedicine and Telecare. 2015;21(2):80-7.
- [36] Ringbæk T, Green A, Laursen LC, Frausing E, Brøndum E, Ulrik CS.
Effect of tele health care on exacerbations and hospital admissions in patients with chronic obstructive pulmonary disease: a randomized clinical trial.
International journal of chronic obstructive pulmonary disease. 2015;10:1801.
- [37] Cordova FC, Ciccolella D, Grabianowski C, Gaughan J, Brennan K, Goldstein F, et al.
A telemedicine-based intervention reduces the frequency and severity of COPD exacerbation symptoms: a randomized, controlled trial.
Telemedicine and e-Health. 2016;22(2):114-22.
- [38] Ho TW, Huang CT, Chiu HC, Ruan SY, Tsai YJ, Yu CJ, et al.
Effectiveness of telemonitoring in patients with chronic obstructive pulmonary disease in Taiwan-a randomized controlled trial.
Scientific Reports. 2016;6(1):1-8.
- [39] Vianello A, Fusello M, Gubian L, Rinaldo C, Dario C, Concas A, et al.
Home telemonitoring for patients with acute exacerbation of chronic obstructive pulmonary disease: a randomized controlled trial.
BMC Pulmonary Medicine. 2016;16(1):1-12.
- [40] Farmer A, Williams V, Velardo C, Shah SA, Yu LM, Rutter H, et al.
Self-management support using a digital health system compared with usual care for chronic obstructive pulmonary disease: randomized controlled trial.
Journal of medical Internet research. 2017;19(5):e7116.
- [41] Lilholt PH, Udsen FW, Ehlers L, Hejlesen OK.
Telehealthcare for patients suffering from chronic obstructive pulmonary disease: Effects on health-related quality of life: Results from the Danish ‘TeleCare North’cluster-randomised trial.
BMJ open. 2017;7(5):e014587.
- [42] Rixon L, Hirani SP, Cartwright M, Beynon M, Doll H, Steventon A, et al.
A RCT of telehealth for COPD patient’s quality of life: the whole system demonstrator evaluation.
The Clinical Respiratory Journal. 2017;11(4):459-69.
- [43] Kessler R, Casan-Clara P, Koehler D, Tognella S, Viejo JL, Dal Negro RW, et al.

BIBLIOGRAPHY

- COMET: a multicomponent home-based disease-management programme versus routine care in severe COPD.
European Respiratory Journal. 2018;51(1).
- [44] Soriano JB, García-Río F, Vázquez-Espinosa E, Conforto JI, Hernando-Sanz A, López-Yepes L, et al.
A multicentre, randomized controlled trial of telehealth for the management of COPD.
Respiratory medicine. 2018;144:74-81.
- [45] Tupper OD, Gregersen TL, Ringbaek T, Brøndum E, Frausing E, Green A, et al.
Effect of tele-health care on quality of life in patients with severe COPD: a randomized clinical trial.
International journal of chronic obstructive pulmonary disease. 2018;2657-62.
- [46] Walker PP, Pompilio PP, Zanaboni P, Bergmo TS, Prikk K, Malinovschi A, et al.
Telemonitoring in chronic obstructive pulmonary disease (CHROMED). A randomized clinical trial.
American journal of respiratory and critical care medicine. 2018;198(5):620-8.
- [47] Mínguez Clemente P, Pascual-Carrasco M, Mata Hernández C, Malo de Molina R, Arvelo L, Cadavid B, et al.
Follow-up with telemedicine in early discharge for COPD exacerbations: randomized clinical trial (TELEMEDCOPD-trial).
COPD: Journal of Chronic Obstructive Pulmonary Disease. 2020;18(1):62-9.
- [48] Sink E, Patel K, Groenendyk J, Peters R, Som A, Kim E, et al.
Effectiveness of a novel, automated telephone intervention on time to hospitalisation in patients with COPD: a randomised controlled trial.
Journal of Telemedicine and Telecare. 2020;26(3):132-9.
- [49] Koff PB, Min Sj, Freitag TJ, Diaz DL, James SS, Voelkel NF, et al.
Impact of proactive integrated care on chronic obstructive pulmonary disease.
Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation. 2021;8(1):100.
- [50] Rassouli F, Germann A, Baty F, Kohler M, Stolz D, Thurnheer R, et al.
Telehealth mitigates COPD disease progression compared to standard of care: a randomized controlled crossover trial.
Journal of Internal Medicine. 2021;289(3):404-10.
- [51] Andersen FD, Trolle C, Pedersen AR, Køpfli ML, Børgesen S, Jensen MS, et al.
Effect of telemonitoring on readmissions for acute exacerbation of chronic obstructive pulmonary disease: A randomized clinical trial.
Journal of Telemedicine and Telecare. 2023;1357633X221150279.

- [52] Køpfli ML, Børgeisen S, Jensen MS, Hyldgaard C, Bell C, Andersen FD. Effect of telemonitoring on quality of life for patients with chronic obstructive pulmonary disease-A randomized controlled trial. *Chronic Respiratory Disease*. 2023;20:14799731231157771.
- [53] Jensen MH, Cichosz SL, Dinesen B, Hejlesen OK. Moving prediction of exacerbation in chronic obstructive pulmonary disease for patients in telecare. *Journal of telemedicine and telecare*. 2012;18(2):99-103.
- [54] van der Heijden M, Lucas PJ, Lijnse B, Heijdra YF, Schermer TR. An autonomous mobile system for the management of COPD. *Journal of biomedical informatics*. 2013;46(3):458-69.
- [55] Bellos CC, Papadopoulos A, Rosso R, Fotiadis DI. Identification of COPD patients' health status using an intelligent system in the CHRONIOUS wearable platform. *IEEE journal of biomedical and health informatics*. 2013;18(3):731-8.
- [56] Fernández-Granero M, Sánchez-Morillo D, León-Jiménez A, Crespo L. Automatic prediction of chronic obstructive pulmonary disease exacerbations through home telemonitoring of symptoms. *Bio-medical materials and engineering*. 2014;24(6):3825-32.
- [57] Anthonisen N, Manfreda J, Warren C, Hershfield E, Harding G, Nelson N. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. *Annals of internal medicine*. 1987;106(2):196-204.
- [58] Shah SA, Velardo C, Gibson OJ, Rutter H, Farmer A, Tarassenko L. Personalized alerts for patients with COPD using pulse oximetry and symptom scores. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2014. p. 3164-7.
- [59] Fernandez-Granero MA, Sanchez-Morillo D, Lopez-Gordo MA, Leon A. A machine learning approach to prediction of exacerbations of chronic obstructive pulmonary disease. In: International Work-Conference on the Interplay Between Natural and Artificial Computation. Springer; 2015. p. 305-11.
- [60] Fernandez-Granero MA, Sanchez-Morillo D, Leon-Jimenez A. Computerised analysis of telemonitored respiratory sounds for predicting acute exacerbations of COPD. *Sensors*. 2015;15(10):26978-96.

BIBLIOGRAPHY

- [61] Mohktar MS, Redmond SJ, Antoniades NC, Rochford PD, Pretto JJ, Basilakis J, et al. Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data. *Artificial intelligence in medicine*. 2015;63(1):51-9.
- [62] Sanchez-Morillo D, Fernandez-Granero MA, Jiménez AL. Detecting COPD exacerbations early using daily telemonitoring of symptoms and k-means clustering: a pilot study. *Medical & biological engineering & computing*. 2015;53:441-51.
- [63] Christian Riis H, Jensen MH, Cichosz SL, Hejlesen OK. Prediction of exacerbation onset in chronic obstructive pulmonary disease patients. *Journal of Medical Engineering & Technology*. 2016;40(1):1-7.
- [64] Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. *Journal of medical Internet research*. 2017;19(3):e7207.
- [65] Fernandez-Granero MA, Sanchez-Morillo D, Leon-Jimenez A. An artificial intelligence approach to early predict symptom-based exacerbations of COPD. *Biotechnology & Biotechnological Equipment*. 2018;32(3):778-84.
- [66] Kronborg T, Mark L, Cichosz SL, Secher PH, Hejlesen O. Population exacerbation incidence contains predictive information of acute exacerbations in patients with chronic obstructive pulmonary disease in telecare. *International Journal of Medical Informatics*. 2018;111:72-6.
- [67] Nunavath V, Goodwin M, Fidje JT, Moe CE. Deep neural networks for prediction of exacerbations of patients with chronic obstructive pulmonary disease. In: International Conference on Engineering Applications of Neural Networks. Springer; 2018. p. 217-28.
- [68] Orchard P, Agakova A, Pinnock H, Burton CD, Sarran C, Agakov F, et al. Improving prediction of risk of hospital admission in chronic obstructive pulmonary disease: application of machine learning to telemonitoring data. *Journal of medical Internet research*. 2018;20(9):e9227.
- [69] Boer LM, van der Heijden M, van Kuijk NM, Lucas PJ, Vercoulen JH, Assendelft WJ, et al. Validation of ACCESS: an automated tool to support self-management of COPD exacerbations.

- International journal of chronic obstructive pulmonary disease. 2018;13:3255.
- [70] Boer L, Bischoff E, van der Heijden M, Lucas P, Akkermans R, Vercoulen J, et al. A smart mobile health tool versus a paper action plan to support self-management of chronic obstructive pulmonary disease exacerbations: randomized controlled trial. JMIR mHealth and uHealth. 2019;7(10):e14408.
- [71] Jin Y, Zhang T, Cao Z, Zhao N, Chen C, Wang D, et al. Prediction Indicators for Acute Exacerbations of Chronic Obstructive Pulmonary Disease by Combining Non-linear analyses and Machine. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2018. p. 2515-21.
- [72] Iadanza E, Mudura V, Melillo P, Gherardelli M. An automatic system supporting clinical decision for chronic obstructive pulmonary disease. Health and Technology. 2020;10:487-98.
- [73] Kronborg T, Hangaard S, Cichosz SL, Hejlesen O. A two-layer probabilistic model to predict COPD exacerbations for patients in telehealth. Computers in Biology and Medicine. 2021;128:104108.
- [74] Patel N, Kinmond K, Jones P, Birks P, Spiteri MA. Validation of COPDPredictTM: Unique Combination of Remote Monitoring and Exacerbation Prediction to Support Preventative Management of COPD Exacerbations. International journal of chronic obstructive pulmonary disease. 2021;16:1887.
- [75] Wu CT, Li GH, Huang CT, Cheng YC, Chen CH, Chien JY, et al. Acute exacerbation of a chronic obstructive pulmonary disease prediction system using wearable device data, machine learning, and deep learning: development and cohort study. JMIR mHealth and uHealth. 2021;9(5):e22591.
- [76] Chmiel FP, Burns DK, Pickering JB, Blythin A, Wilkinson TM, Boniface MJ. Prediction of chronic obstructive pulmonary disease exacerbation events by using patient self-reported data in a digital health app: statistical evaluation and machine learning approach. JMIR medical informatics. 2022;10(3):e26499.
- [77] Wu CT, Wang SM, Su YE, Hsieh TT, Chen PC, Cheng YC, et al. A Precision Health Service for Chronic Diseases: Development and Cohort Study Using Wearable Device, Machine Learning, and Deep Learning.

BIBLIOGRAPHY

- IEEE Journal of Translational Engineering in Health and Medicine. 2022;10:1-14.
- [78] NICE. Chronic obstructive pulmonary disease (acute exacerbation): antimicrobial prescribing. December. 2018;5:2018.
- [79] Jenkins C, Celli B, Anderson J, Ferguson G, Jones P, Vestbo J, et al. Seasonality and determinants of moderate and severe COPD exacerbations in the TORCH study. European Respiratory Journal. 2012;39(1):38-45.
- [80] Donaldson GC, Wedzicha JA. The causes and consequences of seasonal variation in COPD exacerbations. International journal of chronic obstructive pulmonary disease. 2014;9:1101.
- [81] Almagro P, Hernandez C, Martinez-Cambor P, Tresserras R, Escarrabill J. Seasonality, ambient temperatures and hospitalizations for acute exacerbation of COPD: a population-based study in a metropolitan area. International journal of chronic obstructive pulmonary disease. 2015;10:899.
- [82] Seemungal TA, Hurst JR, Wedzicha JA. Exacerbation rate, health status and mortality in COPD—a review of potential interventions. International journal of chronic obstructive pulmonary disease. 2009;4:203.
- [83] Lewis KE, Annandale JA, Warm DL, Rees SE, Hurlin C, Blyth H, et al. Does home Telemonitoring after pulmonary rehabilitation reduce healthcare use in optimized COPD?? A pilot randomized trial. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2011;7(1):44-50.
- [84] Karuppanan K, Vairasundaram AS, Sigamani M. A comprehensive machine learning approach to prognose pulmonary disease from home. In: Proceedings of the International Conference on Advances in Computing, Communications and Informatics; 2012. p. 518-23.
- [85] GOLD. GLOBAL STRATEGY FOR PREVENTION, DIAGNOSIS AND MANAGEMENT OF COPD: 2023 Report; 2023.
- [86] Buekers J, De Boever P, Vaes AW, Aerts JM, Wouters EF, Spruit MA, et al. Oxygen saturation measurements in telemonitoring of patients with COPD: a systematic review. Expert review of respiratory medicine. 2018;12(2):113-23.

- [87] Burton C, Pinnock H, McKinstry B.
Changes in telemonitored physiological variables and symptoms prior to exacerbations of chronic obstructive pulmonary disease.
Journal of telemedicine and telecare. 2015;21(1):29-36.
- [88] Miłkowska-Dymanowska J, Biały AJ, Obrebski W, Górska P, Piotrowski WJ.
A pilot study of daily telemonitoring to predict acute exacerbation in chronic obstructive pulmonary disease.
International Journal of Medical Informatics. 2018;116:46-51.
- [89] Xie Y, Redmond SJ, Mohktar MS, Shany T, Basilakis J, Hession M, et al.
Prediction of chronic obstructive pulmonary disease exacerbation using physiological time series patterns.
In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2013. p. 6784-7.
- [90] Shin SH, Kwon SO, Kim V, Silverman EK, Kim TH, Kim DK, et al.
Association of body mass index and COPD exacerbation among patients with chronic bronchitis.
Respiratory research. 2022;23(1):1-9.
- [91] Van der Heijden M, Velikova M, Lucas PJ.
Learning Bayesian networks for clinical time series analysis.
Journal of biomedical informatics. 2014;48:94-105.
- [92] Chau JPC, Lee DTF, Yu DSF, Chow AYM, Yu WC, Chair SY, et al.
A feasibility study to investigate the acceptability and potential effectiveness of a telecare service for older people with chronic obstructive pulmonary disease.
International journal of medical informatics. 2012;81(10):674-82.
- [93] Watz H, Tetzlaff K, Magnussen H, Mueller A, Rodriguez-Roisin R, Wouters EF, et al.
Spirometric changes during exacerbations of COPD: a post hoc analysis of the WISDOM trial.
Respiratory research. 2018;19(1):1-7.
- [94] Chawla H, Bulathsinghala C, Tejada JP, Wakefield D, ZuWallack R.
Physical activity as a predictor of thirty-day hospital readmission after a discharge for a clinical exacerbation of chronic obstructive pulmonary disease.
Annals of the American Thoracic Society. 2014;11(8):1203-9.
- [95] Jones P, Harding G, Berry P, Wiklund I, Chen W, Leidy NK.
Development and first validation of the COPD Assessment Test.
European Respiratory Journal. 2009;34(3):648-54.

BIBLIOGRAPHY

- [96] Min SJ, Lee YS, Park YB, Yum HK, Park MJ, Yoo JH. COPD Assessment Test (CAT) has a positive relationship with exacerbations in COPD patients. Eur Respiratory Soc; 2019.
- [97] Rassouli F, Baty F, Stolz D, Albrich WC, Tamm M, Widmer S, et al. Longitudinal change of COPD assessment test (CAT) in a telehealthcare cohort is associated with exacerbation risk. International Journal of Chronic Obstructive Pulmonary Disease. 2017;12:3103.
- [98] Tabak M, Brusse-Keizer M, van der Valk P, Hermens H, Vollenbroek-Hutten M. A telehealth program for self-management of COPD exacerbations and promotion of an active lifestyle: a pilot randomized controlled trial. International journal of chronic obstructive pulmonary disease. 2014;9:935.
- [99] Lacoma A, Prat C, Andreo F, Dominguez J. Biomarkers in the management of COPD. European Respiratory Review. 2009;18(112):96-104.
- [100] Hurst JR, Donaldson GC, Perera WR, Wilkinson TM, Bilello JA, Hagan GW, et al. Use of plasma biomarkers at exacerbation of chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2006;174(8):867-74.
- [101] Butler CC, Gillespie D, White P, Bates J, Lowe R, Thomas-Jones E, et al. C-reactive protein testing to guide antibiotic prescribing for COPD exacerbations. New England Journal of Medicine. 2019;381(2):111-20.
- [102] Koutsokera A, Kostikas K, Nicod LP, Fitting JW. Pulmonary biomarkers in COPD exacerbations: a systematic review. Respiratory research. 2013;14(1):1-12.
- [103] Lawal O, Ahmed WM, Nijssen TM, Goodacre R, Fowler SJ. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. Metabolomics. 2017;13:1-16.
- [104] Kamal F, Kumar S, Edwards MR, Veselkov K, Belluomo I, Kebadze T, et al. Virus-induced volatile organic compounds are detectable in exhaled breath during pulmonary infection. American Journal of Respiratory and Critical Care Medicine. 2021;204(9):1075-85.
- [105] Navarro CLA, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. bmj. 2021;375.

- [106] Lewis SC, Warlow C.
How to spot bias and other potential problems in randomised controlled trials.
Journal of Neurology, Neurosurgery & Psychiatry. 2004;75(2):181-7.
- [107] Cooper R, Giangreco A, Duffy M, Finlayson E, Hamilton S, Swanson M, et al.
Evaluation of myCOPD digital self-management technology in a remote and rural population: real-world feasibility study.
JMIR mHealth and uHealth. 2022;10(2):e30782.
- [108] Kon SS, Canavan JL, Jones SE, Nolan CM, Clark AL, Dickson MJ, et al.
Minimum clinically important difference for the COPD Assessment Test: a prospective analysis.
The lancet Respiratory medicine. 2014;2(3):195-203.
- [109] Lin L, Song Q, Cheng W, Liu C, Zhao YY, Duan JX, et al.
Comparation of predictive value of CAT and change in CAT in the short term for future exacerbation of chronic obstructive pulmonary disease.
Annals of Medicine. 2022;54(1):875-85.
- [110] Medicines Healthcare products Regulatory Agency. Oversight and monitoring activities; 2022.
Accessed on January 11, 2024.
Available from: <https://www.gov.uk/government/publications/oversight-and-monitoring-of-investigational-medical-product-trials/oversight-and-monitoring-activities>.
- [111] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al.
Scikit-learn: Machine Learning in Python.
Journal of Machine Learning Research. 2011;12:2825-30.
- [112] Lemaître G, Nogueira F, Aridas CK.
Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning.
Journal of Machine Learning Research. 2017;18(17):1-5.
Available from: <http://jmlr.org/papers/v18/16-365.html>.
- [113] Bishop CM, Nasrabadi NM.
Pattern recognition and machine learning. vol. 4.
Springer; 2006.
- [114] Kuhn M, Johnson K, et al.
Applied predictive modeling. vol. 26.
Springer; 2013.

BIBLIOGRAPHY

- [115] Liu XY, Wu J, Zhou ZH.
Exploratory undersampling for class-imbalance learning.
IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics).
2008;39(2):539-50.
- [116] Hastie T, Rosset S, Zhu J, Zou H.
Multi-class adaboost.
Statistics and its Interface. 2009;2(3):349-60.
- [117] Breiman L, Friedman JH, Olshen RA, Stone CJ.
Classification and regression trees. Belmont, CA: Wadsworth.
International Group. 1984;432(151-166):9.
- [118] Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F.
Learning from imbalanced data sets. vol. 10.
Springer; 2018.
- [119] Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G.
Learning from class-imbalanced data: Review of methods and applications.
Expert systems with applications. 2017;73:220-39.
- [120] Braun V, Clarke V.
Using thematic analysis in psychology.
Qualitative research in psychology. 2006;3(2):77-101.
- [121] Suzuki R, Yajima N, Sakurai K, Oguro N, Wakita T, Thom DH, et al.
Association of patients' past misdiagnosis experiences with trust in their current physician among Japanese adults.
Journal of General Internal Medicine. 2022;37(5):1115-21.
- [122] Kunneman M, Stiggebout AM, Pieterse AH.
Do clinicians convey what they intend? Lay interpretation of verbal risk labels used in decision encounters.
Patient education and counseling. 2020;103(2):418-22.
- [123] Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM.
Measuring numeracy without a math test: development of the Subjective Numeracy Scale.
Medical Decision Making. 2007;27(5):672-80.
- [124] Dodd J, Getov S, Jones PW.
Cognitive function in COPD.
European Respiratory Journal. 2010;35(4):913-22.

- [125] Laue J, Melbye H, Risør MB.
Self-treatment of acute exacerbations of chronic obstructive pulmonary disease requires more than symptom recognition—a qualitative study of COPD patients' perspectives on self-treatment.
BMC family practice. 2017;18:1-10.
- [126] Hadjistavropoulos HD, Craig KD, Hadjistavropoulos T.
Cognitive and behavioral responses to illness information: the role of health anxiety.
Behaviour Research and Therapy. 1998;36(2):149-64.
- [127] Gilkes A, Ashworth M, Schofield P, Harries TH, Durbaba S, Weston C, et al.
Does COPD risk vary by ethnicity? A retrospective cross-sectional study.
International journal of chronic obstructive pulmonary disease. 2016;739-46.
- [128] Compernolle S, DeSmet A, Poppe L, Crombez G, De Bourdeaudhuij I, Cardon G, et al.
Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis.
International Journal of Behavioral Nutrition and Physical Activity. 2019;16(1):1-16.
- [129] Baker RC, Kirschenbaum DS.
Self-monitoring may be necessary for successful weight control.
Behavior Therapy. 1993;24(3):377-94.
- [130] Cooper CB, Sirichana W, Neufeld EV, Taylor M, Wang X, Dolezal BA.
Statistical process control improves the feasibility of remote physiological monitoring in patients with chronic obstructive pulmonary disease.
International Journal of Chronic Obstructive Pulmonary Disease. 2019;2485-96.
- [131] Glyde H, Blythin A, Wilkinson T, Nabney I, Dodd J. Exacerbation predictive modelling using real-world data from the myCOPD app. Eur Respiratory Soc; 2022.
- [132] Chen YWR, Leung JM, Sin DD.
A systematic review of diagnostic biomarkers of COPD exacerbation.
PloS one. 2016;11(7):e0158843.
- [133] Patel N, Belcher J, Thorpe G, Forsyth NR, Spiteri MA.
Measurement of C-reactive protein, procalcitonin and neutrophil elastase in saliva of COPD patients and healthy controls: correlation to self-reported wellbeing parameters.
Respiratory research. 2015;16(1):1-15.
- [134] Li CX, Zhang L, Yan YR, Ding YJ, Lin YN, Zhou JP, et al.
A narrative review of exploring potential salivary biomarkers in respiratory diseases: Still on its way.

BIBLIOGRAPHY

- Journal of Thoracic Disease. 2021;13(7):4541.
- [135] Stockley RA, Halpin DM, Celli BR, Singh D.
Chronic obstructive pulmonary disease biomarkers and their interpretation.
American journal of respiratory and critical care medicine. 2019;199(10):1195-204.
- [136] Talker L, Neville D, Wiffen L, Selim AB, Haines M, Carter JC, et al.
Machine diagnosis of chronic obstructive pulmonary disease using a novel fast-response
capnometer.
Respiratory Research. 2023;24(1):1-11.