

# Predicting Cost-effectiveness of Telehealthcare to Patients with COPD: A Feasibility Study Based on Data from the TeleCare North Cluster-randomized Trial

Flemming Witt Udsen, Ole Hejlesen

Department of Health Science & Technology, Aalborg University, Aalborg, Denmark

## Abstract

*International results have recently questioned the value of providing telehealthcare to all COPD patients. Results from the Danish TeleCare North trial nuanced the debate by concluding that telehealthcare would most likely only be cost-effective for patients in a subgroup of severe COPD. Machine-learning methods have been suggested as a strategy to target telehealthcare even better than clinical subgroups. Data from the TeleCare North trial was used to fit classification models in order to explore this feasibility. Three models were applied: a simple decision tree, logistic regression and a linear support vector machine. Results indicate that classification methods can be used to predict patient-level cost-effectiveness with a relatively high precision. With these methods, it is feasible to target telehealthcare even better in order to maximize survival and health-related quality of life while not overusing scarce health resources as argued by health economists and clinical advocates of rational medicine.*

## Keywords:

Pattern Recognition, Automated/classification; Cost-Benefit Analysis, Telemedicine; Pulmonary Disease, Chronic Obstructive, Denmark.

## Introduction

Chronic obstructive pulmonary disease (COPD) is a prevalent disease [1,2]. The World Health Organization projects that 64 million people have COPD worldwide, attributing to more than 3 million deaths annually [3]. COPD is a major health problem in the European Union with a member state COPD prevalence between 4-10% [4]. The Danish Lung Association estimates that 430,000 Danish citizens have COPD [5] with a scientific study setting the prevalence to 9% [6]. According to an analysis of the disease burden in Denmark, around 3,300 Danish citizens die due to COPD each year (6% of total number of deaths) [7] and 2,500 die each year of causes related to COPD [8].

Systematic reviews and a recent large-scale randomized trial have shown some effect of telehealthcare on health outcomes [9–11]. However, due to fiscal pressure in healthcare sectors, effective interventions might still not receive public funding,

if there is not a reasonable relationship between health outcomes and associated costs [12]. Although several pilot trials have suggested that telehealthcare could be cost-effective or cost-saving for COPD patients [13–17], recent larger-scale studies have found it hard to replicate similar results, suggesting that it might not be cost-effective to scale up telehealthcare solutions to all COPD patients assuming that “one-solution fit all” [18–21].

However, there is little knowledge on the implications of patient heterogeneity, i.e. that some patient groups might for some reason be more or less cost-effective than others. Recent results from the Danish TeleCare North trial is one of the first studies to indicate that, although telehealthcare is not cost-effective for all COPD patients (21), it might – on average - be highly cost-effective for patients in the severe COPD subgroup [22].

But at the same time, there is a growing demand for targeting telehealthcare even better than by clinical subgroups in order to identify the contexts and characteristics of ideal individual patients that should receive telehealthcare, i.e. enabling assessment that answers under which circumstances telehealthcare would be most likely to be cost-effective [23–25]. Machine-learning methods used to predict outcomes for individual patients eligible for telehealthcare has recently been suggested as one way forward [26].

The aim of this project is to investigate if it is possible to use predictive algorithms to help stratify telehealthcare for COPD patients in a way that maximizes the patient-level cost-effectiveness ratio.

## Materials and Methods

### Data

The study design, data collection procedure [27] as well as the results of TeleCare North has been published elsewhere [21,22]. In addition to usual care, patients in the intervention group received a set of telehealthcare equipment, disease-specific education and were monitored by a municipality-based healthcare team. The control group received usual care. 1,225 patients were included with 578 patients

receiving telehealthcare and 647 usual care. The duration of the study was 12 months.

For the purposes of this study, the variables collected as part of the study is of particular interest and can be divided into the cost-effectiveness ratio to be predicted and a set of features used to predict this outcome.

### **Outcome**

The primary outcome is the total costs per quality-adjusted life year (QALY) gained at the individual level, which represent an individualized cost-effectiveness ratio.

A QALY is a composite measure of survival and health-related quality-of-life (HRQoL) [28]. Information on mortality was taken from the Danish Register of Causes of Death [29]. HRQoL stemmed from generic EQ-5D-3L questionnaires distributed to patients at baseline and 12-months follow-up. The EQ-5D scores HRQoL on a scale from 0-1, where "1" indicates perfect HRQoL and "0" meaning dead. Negative scores are possible indicating health states considered worse than death [30]. QALYs were calculated by linear interpolation of EQ5D-3L scores with Danish societal weights [31,32].

Total costs included intervention costs, healthcare costs (hospital-, medicine-, and primary sector costs), and social sector costs (cost to practical help and care at home, home-based nursing care, and rehabilitation). Within-trial healthcare costs were all collected from national registers by applying patients' unique social security number. Hospital contacts were collected from the Danish National Patient Register [33]; contacts between patients and the primary care sector from the National Health Insurance Service Register [34]; and medication use was taken from The Danish Register of Medicinal Product Statistics [35]. Social sector costs was estimated from electronic care systems in each of the 10 included municipalities. Intervention costs were assessed from prices paid during the TeleCare North implementation and included costs of hardware, installation, maintenance and support as well as training costs, monitoring costs, and project management costs. All costs were reported in 2014 prices and were obtained in Danish kroner (DKK) and thereafter exchanged to € using the average 2014 exchange rate (1€=7.4547 DKK).

### **Features**

Different baseline features were collected as part of the trial. All features originated from a combination of questionnaires, physical measurements conducted by general practitioners and from national registers. Three baseline HRQoL summary scores (baseline EQ-5D, PCS and MCS scores) was available from two generic questionnaires distributed to all patients: the EQ-5D-3L [30] and the SF-36 [36]. Six physiological parameters was measured by the patients' general practitioner and consisted of systolic- and diastolic blood pressure, pulse, body mass index (BMI), spirometry measures (percentage of expected forced expiratory volume in one second (FEV1 (%)), percentage of expected forced vital capacity (FVC (%)). Six socio-demographics (marital status, highest education, duration of COPD, job status, number of persons

in household and smoking status) and the presence of five comorbidities (diabetes, musculoskeletal disease, cancer, mental illness and heart disease) was ascertained from questionnaires that were filled out by patients. Age and gender was identified from the patients' social security number and their residing municipality was collected from the Danish Civil Registration System [37]. Nine features associated with the historical activity for the included patients was incorporated to account for organizational differences and variation in visitation practices. Therefore, the number and duration of hospital admissions and the number of outpatient visits one year prior to randomization was included as were the costs associated with these contacts. Costs due to medicine, primary care and practical help and care at home, home-based nursing care, and rehabilitation was also collected for the one year leading up to the evaluation period. All features was collected from the same national registers as described above.

In total, the dataset consist of 32 features. In addition, six predictors were derived from other features: In the original cost-effectiveness analysis, three characteristics were important in distinguishing between subgroups that were more cost-effective than others [22]. The cost-effectiveness of telehealthcare depended on COPD severity group that can be calculated from FEV1(%) [38]. Furthermore, existing social sector costs was a driver of cost-effectiveness; as were baseline cost (accumulated costs one year prior to evaluation). Both variables were calculated from the individual baseline cost-categories that were collected as part of the trial. Three other features were generated that indicate whether a patient suffer from hypertension, multimorbidities or tachycardia and they were calculated from blood pressure, individual comorbidities and pulse, respectively. These were included because of their clinical meaningfulness.

### **Missing data**

Of the 1,225 patients originally included in the TeleCare North trial, complete data for both total costs (i.e. all cost categories), baseline EQ-5D score and EQ-5D score at follow-up were available for 728 patients (59%; 302 in telehealthcare group; 426 in control group). Missing data for the EQ-5D summary score were present for 8% of the participants at baseline (48 in the telehealthcare group; 53 in the control group). Due to non-response or to incomplete registration of EQ-5D questionnaire items, 27% had missing data on the EQ-5D summary score at follow-up (199 in the telehealthcare group; 133 in the control group). Two municipalities where unable to extract rehabilitation costs (79 in the telehealthcare group; 73 in the control group).

Only the 302 participants with complete outcome data in the telehealthcare group were included in this analysis.

Missing data in any features from those 302 participants were replaced by multiple imputation. Missing feature values were assumed missing at random (MAR) at replaced with the *mi impute chained* command in STATA12.1 and 30 complete datasets were created an averaged to one value per missing. Continuous variables were imputed by predictive mean

matching and categorical variables by multinomial logistic or logistic regression. Imputation models included predictors for the outcomes at both time points and predictors for missing observations in the individual variables. The imputation models were estimated separately by treatment group and included the clustering variable and the measures of disease status, presence of comorbidities and sociodemographic variables described previously.

The characteristics of the included patients are presented in Table 1.

Table 1- Baseline characteristics of the sample (n=302)

Age (years)	69.57 (8.87)
Men (%)	50.99 (n=154)
Marital status	
Married/in a relationship	66.56 (n=201)
Single	17.55 (n=53)
Widow/widower	15.89 (n=48)
Smoking status	
Non-smokers	65.89 (n=199)
Smokers	34.11 (n=103)
Highest education	
Elementary school	46.03 (n=139)
Secondary school	4.30 (n=13)
Vocational education	34.77 (n=105)
Short tertiary school (2-3 years)	7.95 (n=24)
Bachelor or equivalent (3-5 years)	6.95 (n=21)
Master or equivalent (>5 years)	0.00 (n=0)
Job status	
Full-time	3.64 (n=11)
Part-time	7.28 (n=22)
None	89.07 (n=269)
Blood pressure, diastolic	76.64 (10.51)
Blood pressure, systolic	130.62 (16.97)
Pulse	79.12 (13.31)
BMI	26.22 (5.01)
Duration of COPD (years)	7.97 (6.10)
FEV1 (%)	47.57 (16.29)
FVC (%)	70.85 (16.94)
Comorbidities	
Diabetes	10.69 (n=32)
Heart disease	32.12 (n=97)
Mental health problem	3.31 (n=10)
Musculoskeletal disease	25.83 (n=78)
Cancer	5.63 (n=17)
Baseline total cost (€)	4,863.37 (7,874.23)
Baseline HRQoL	
EQ-5D score	0.727 (0.20)
PCS (SF-36 physical score)	38.30 (8.68)

MCS (SF-36 mental score)	48.98 (10.97)
--------------------------	---------------

Data are mean (SD) or proportion (number of patients)  
COPD: Chronic obstructive pulmonary disease; FEV1(%): forced expiratory volume in one second of predicted normal; FVC(%): forced vital capacity

### Model and features

To investigate some basic methods of machine-learning, a relative naïve approach was chosen for this study. The prediction of cost-effectiveness was addressed as a simple pattern classification problem to test different methods.

First, all 38 features regardless of multicollinearity were used in subsequent analyses.

Second, the individualized cost-effectiveness ratio was categorized into two groups (Class 1: Total cost/QALY  $\leq$  €5,000; Class 2: Total cost/QALY  $>$  €5,000). This was done due to simplicity and because the cost-effectiveness ratio was highly skewed even after normalization. The boundary was exploratively chosen based on the distribution of the individualized cost-effectiveness ratio.

Third, MATLAB Release 2017a was used to train and evaluate all 38 predictors using a five-fold cross-validation. A simple decision tree, logistic regression and linear support vector machines were fitted to the data.

### Evaluation

All models were evaluated based on accuracy and the area under the receiver operating characteristics curve (AUC). Accuracy is the proportion of correct predictions made for the two cost-effectiveness categories among the total number of observations. The AUC curve quantifies the overall ability of the classification model to discriminate between observations that have a total cost/QALY  $\leq$  €5,000 or  $>$  €5,000 across classification thresholds. A model with no better precision than chance has an AUC of 0.5.

### Results

Table 2 presents model performance for each of the three models that were fitted. The linear SVM performed best with an accuracy of 79.1% and an area under the curve of 0.89. The classification model was therefore able to classify 79% of the observations into the correct cost-effectiveness categories with a relatively high ability to distinguish which observations should be predicted to have an individualized cost-effectiveness larger or smaller than €5,000 across classification levels.

Moreover, all three models had relatively high performance (accuracy between 76.7-79.1% and area under the curves between 0.81-0.89).

Table 2- Performance evaluation of the included models

	Accuracy (%)	AUC
Simple tree	76.7	0.81

Logistic regression	77.4	0.84
Linear SVM	79.1	0.89

AUC: Area under the curve; SVM: Support vector machine

Figure 1 and Figure 2 presents the receiver operation characteristics of both classes (Class 1: Total cost/QALY  $\leq$  €5,000; Class 2: Total cost/QALY  $>$  €5,000) for the best performing classification model (linear SVM).

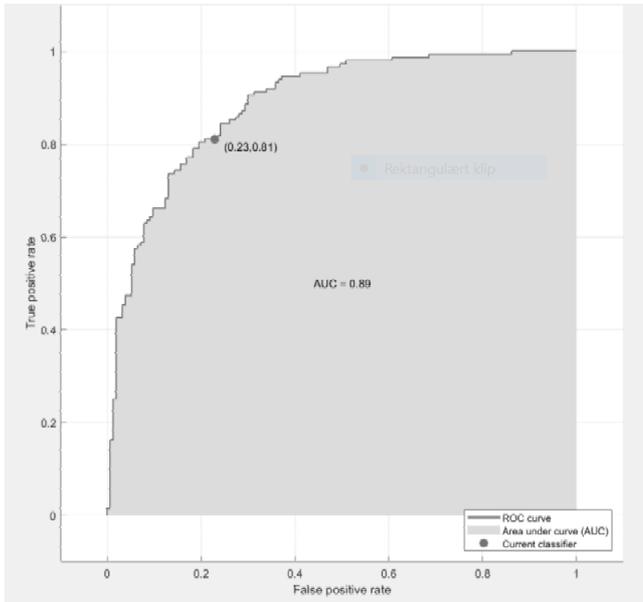


Figure 1- Receiver operation characteristics based on a linear support vector machine (Class 1)

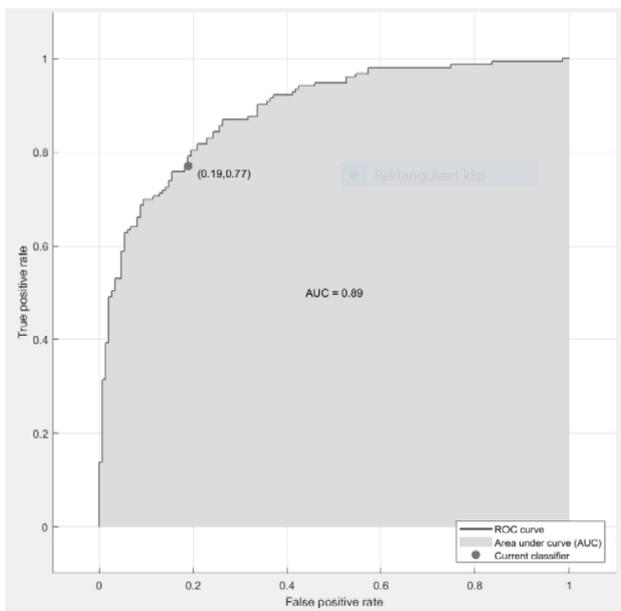


Figure 2- Receiver operation characteristics based on a linear support vector machine (Class 2)

## Discussion

The best performing model was the linear SVM, but all three compared models had relatively high accuracy (from 76.7-79.1) and large area under the ROC-curves (from 0.81-0.89). This seems to suggest that machine-learning methods can be used to predict the individualized cost-effectiveness of telehealth to COPD patients with a relatively high precision.

### Strength/weaknesses

This study applied data from a large clinical trial that rigorously collected patient-level data on both health outcomes, socio-demographics and resource patterns across all relevant stakeholders involved in managing COPD patients. All cost data are based on unambiguous register-data complete eliminating any self-report biases.

A limitation is that no feature selection was conducted prior to model training. Using 38 features to predict a two-category outcome resulting in a relatively high precision might be due to overfitting of the data, which means that the models accurately classifies the training data, but would perform poorly on future data. A five-fold cross validation was used to counter some of the problems of overfitting the data, but in general, much fewer features could make the same classification. From the original cost-effectiveness analysis, it is known that especially baseline HRQoL and total baseline costs (the sum of costs accumulated from all cost-categories 12 months prior to evaluation start) had a major impact on the cost-effectiveness estimate [21]. For this reason baseline EQ-5D is required as a variable in any models estimating cost-effectiveness in health economic research [39].

Another limitation is that no techniques for dealing with outliers were applied. It could rightfully be stated that very large outliers are attributed to patients having other serious disease alongside COPD. However, models did account for major comorbidities and baseline total costs. And health economists will insist on asking what threshold levels constitutes an outlier given smooth cost-effectiveness distributions and argue that outliers are crucial for estimating impact on healthcare budgets, which is why mean values instead of medians are always reported [40,41].

Finally, large proportions of missing data can be a particular problem in machine-learning. In clinical trials, good practice implies handling missing data with multiple imputation [42,43]; but this procedure could artificially boost the performance of prediction models, since multiple imputation entails using other variables to predict missing values. This study therefore sought to minimize the missing data challenge by focusing on the telehealthcare group with complete cost-effectiveness ratio. But these participants still had missing data on especially socio-demographics and physical measurements used to predict cost-effectiveness.

### Comparison with other research

While some studies have applied machine-learning techniques in medicine to predict outcomes related to health eco-

nomic evaluation, i.e. mortality in heart disease [44], HRQoL in COPD [45] and even the cost of treatment in liver disease [46]; no other studies have been found that sought directly to predict the cost-effectiveness ratio of health technologies.

The study published by Stausholm and colleagues [44] come closest to the design of this study. It applied data from the TeleCare North trial to predict increasing or decreasing HRQoL and healthcare sector costs for 553 COPD patients receiving both telehealthcare and usual care. The design entailed application of four different logistic regression models with 39 features, which were evaluated by accuracy and root mean square error (RMSE). Accuracy of models ranged from 61-65% for models predicting HRQoL and 74-75% for models predicting healthcare sector costs, while the RMSE was 5.265 scores for HRQoL and \$5430 for healthcare costs. The study concluded that predictive analytics could be used to stratify COPD patients for telehealthcare. This current study has a slightly higher accuracy (76.7-79.1%) with roughly the same number of features. However, this study applied a different HRQoL summary score (the EQ-5D as opposed to SF-36 in Stausholm) and accounted for a much broader cost-perspective. Especially costs in municipalities (due to daily practical help and home nursing care) led to larger proportion of COPD patients to have very large costs, that were in part accounted for by including a feature for having resource use in municipalities prior to data collection.

### Implications

Based on the cost-effectiveness results from the TeleCare North trial, the Danish Government decided to fund telehealthcare to the subgroup of patients with severe COPD throughout Denmark [47]. Although, this subgroup was likely to be cost-effective on average, there is both a scientific and practical need to select COPD patients from an even more limited set of potential candidates for telehealthcare [26]. This would result in a quicker and less expensive implementation of telehealthcare that would facilitate better health outcomes while also reducing overall health- and social care costs.

Machine-learning methods already has an established role in healthcare and biomedicine [48] in areas such as disease discovery, e.g. [49] and diagnosis [50]. Results from this study seems to indicate that it would be possible to predict individual cost-effectiveness of telehealthcare for future patients. This would enable more personalized COPD management, i.e. a COPD management strategy that are based on individual data paired with a predictive toolbox that evaluates which combinations of individuals and technologies, that are eligible for receiving public funding.

### Future research

More studies on the type of models used to predict cost-effectiveness should be conducted. The individualized cost-effectiveness ratio is a very left-skewed ratio-scaled variable ranging, in principle, from  $-\infty$  to  $+\infty$ , but with mostly positive values. This study applied a pattern recognition strategy by dividing cost-effectiveness into two categories, but models that are used to predict the continuous cost-effectiveness ra-

tio or even more cost-effectiveness categories could have been applied instead.

New predictive models might also need to be developed and incorporated into existing software that can account for the clustered nature of health- and cost data: due to variation in practice, it is plausible that mortality, health-related quality of life and costs are more similar within e.g. geographical areas or the patients' organizational affiliation than across these areas. Treating this information merely as fixed variables in modelling outcomes can lead to biased model coefficients and –uncertainty [51,52].

### Acknowledgements

Being completely new to the field of health informatics and pattern recognition, the author would like to thank the Department of Health Science & Technology at Aalborg University for funding and the opportunity to work with machine learning methods on health economic data.

### Competing interest

None declared.

### References

- [1] Halbert RJ, Natoli JL, Gano A, Badamgarav E, Buist AS, Mannino DM. Global burden of COPD: systematic review and meta-analysis. *Eur Respir J* [Internet]. 2006;28(3):523–32. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16611654>
- [2] Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. *Lancet* [Internet]. 2007;370(9589):765–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17765526>
- [3] WHO. Chronic obstructive pulmonary disease (COPD) [Internet]. Fact sheet 315. 2012. Available from: <http://www.who.int/mediacentre/factsheets/fs315/en/index.html>
- [4] European COPD Coalition. Prevalence of COPD in EU [Internet]. 2014. Available from: <http://www.copdcoalition.eu/about-copd/prevalence>
- [5] Danish Lung Association. COPD in Denmark [Internet]. 2012. Available from: [www.lunge.dk](http://www.lunge.dk)
- [6] Hansen JG, Pedersen L, Overvad K, Omland Ø, Jensen HK, Sørensen HT. The Prevalence of chronic obstructive pulmonary disease among Danes aged 45-84 years: population-based study. *COPD* [Internet]. 2008;5(6):347–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19353348>
- [7] Flachs EM, Eriksen L, Koch MB. The disease burden in Denmark. 2015.
- [8] National Institute of Public Health. The Public Health Report, Denmark. 2007.
- [9] McLean S, Nurmatov U, Liu JLY, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease: Cochrane Review and meta-analysis. *Br J Gen Pract* [Internet]. 2012;62(604):e739–49.

- Available from:  
<http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L365948923>
- [10] Steventon A., Bardsley M, Billings J, Dixon J, Doll H, Hirani S, et al. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *Bmj* [Internet]. 2012 Jun 21;344(jun21 3):e3874–e3874. Available from: <http://www.bmj.com/cgi/doi/10.1136/bmj.e3874>
- [11] Wootton R. Twenty years of telemedicine in chronic disease management—an evidence synthesis. *J Telemed Telecare* [Internet]. 2012;18(4):211–20. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3366107&tool=pmcentrez&rendertype=abstract>
- [12] Drummond M, Sculpher J. M, Glaxton K, Stoddart L. G, Torrance W. G. *Methods for the Economic Evaluation of Health Care Programmes*. 4th edition. Oxford University Press; 2015.
- [13] Haesum LKE, Soerensen N, Dinesen B, Nielsen C, Grann O, Hejlesen O, et al. Cost-utility analysis of a telerehabilitation program: A case study of COPD patients. *Telemed e-Health* [Internet]. 2012;18(9):688–92. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84869057985&partnerID=40&md5=87200bfb992018c29819b0f75c0d60cf>
- [14] de Toledo P, Jiménez S, del Pozo F, Roca J, Alonso A, Hernandez C. Telemedicine experience for chronic care in COPD. *IEEE Trans Inf Technol Biomed* [Internet]. 2006 Jul;10(3):567–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16871726>
- [15] Vitacca M, Assoni G, Pizzocaro P, Guerra A, Marchina L, Scalvini S, et al. A pilot study of nurse-led, home monitoring for patients with chronic respiratory failure and with mechanical ventilation assistance. *J Telemed Telecare* [Internet]. 2006;12(7):337–42. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-39049188481&partnerID=40&md5=cb0d21f36d8f7c730d9c5d2ff6b08297>
- [16] Koff PB, Jones RH, Cashman JM, Voelkel NF, Vandivier RW. Proactive integrated care improves quality of life in patients with COPD (Provisional abstract). *Eur Respir J* [Internet]. 2009 May [cited 2013 Sep 6];33(5):1031–8. Available from: <http://onlinelibrary.wiley.com/o/cochrane/celed/articles/NHSEED-22009102080/frame.html>
- [17] Pare G, Poba-Nzaou P, Sicotte C, Beaupre A, Lefrancois E, Nault D, et al. Comparing the costs of home telemonitoring and usual care of chronic obstructive pulmonary disease patients: A randomized controlled trial. *Eur Res Telemed* [Internet]. 2013;2(2):35–47. Available from: <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L52635611>
- [18] Henderson C, Knapp M, Fernández J, Beecham J, Hirani S, Cartwright M, et al. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. *BMJ*. 2013;346.
- [19] Stoddart A, van der Pol M, Pinnock H, Hanley J, McCloughan L, Todd A, et al. Telemonitoring for chronic obstructive pulmonary disease: a cost and cost-utility analysis of a randomised controlled trial. *J Telemed Telecare* [Internet]. 2015;21(2):108–18. Available from: <http://jtt.sagepub.com/lookup/doi/10.1177/1357633X14566574>
- [20] 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. *J Telemed Telecare* [Internet]. 2015;21(2):80–7. Available from: <http://jtt.sagepub.com.ez.statsbiblioteket.dk:2048/content/21/2/80.full>
- [21] Udsen F, Lilholt P, Hejlesen O, Ehlers L. Cost-effectiveness of telehealthcare to patients with chronic obstructive pulmonary disease: Results from the Danish “TeleCare North” cluster-randomized trial. *BMJ Open*. 2017;7(5).
- [22] Udsen FW, Lilholt PH, Hejlesen OK, Ehlers L. Subgroup analysis of telehealthcare to patients with chronic obstructive pulmonary disease: The case of the Danish “TeleCare North” cluster-randomized trial. *Clin Outcomes Res*. 2017;9:391–401.
- [23] Dixon P, Hollinghurst S, Edwards L, Thoms C, Foster A, Davies B, et al. Cost-effectiveness of telehealth for patients with depression: evidence from the Healthlines randomised controlled trial. *BJPsych Open* [Internet]. 2016;2:262–269. Available from: <http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2016-012355>
- [24] Dixon P, Hollinghurst S, Edwards L, Thomas C, Gaunt D, Foster A, et al. Cost-effectiveness of telehealth for patients with raised cardiovascular disease risk: evidence from the Healthlines randomised controlled trial. *BMJ Open* [Internet]. 2016;6(9):e012355. Available from: <http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2016-012355>
- [25] Udsen F. *Health Economic Evaluation of Telehealthcare: Can we include “why” and “under what circumstances” telehealthcare is cost-effective in health economic evaluation?* [Internet]. Aalborg: Aalborg University Press; 2016. Available from: <http://aauforlag.dk/Shop/phd-udgivelse/phd-by-flemming-witt-udsen.aspx>
- [26] Vitacca M, Montini A, Comini L. How will telemedicine change clinical practice in chronic obstructive pulmonary disease? *Ther Adv Respir Dis* [Internet]. 2018;12:175346581875477. Available from: <http://journals.sagepub.com/doi/10.1177/1753465818754778>
- [27] Udsen F, Lilholt P, Hejlesen O, Ehlers L. Effectiveness and cost-effectiveness of telehealthcare for chronic obstructive pulmonary disease: study protocol for the Danish “TeleCare North” pragmatic cluster-randomized trial. *Trials*. 2014;15(178).
- [28] Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Valuation*. 1st ed. New York: Oxford University Press;

2009. 344 p.
- [29] The State Serum Institute. The Danish Register of Causes of Death. 2014.
- [30] Reenen M van, Oppe M. EQ-5D-3L User Guide. EuroQol Res Found. 2015;(April):22.
- [31] Sørensen J, Davidsen M, Gudex C, Pedersen KM, Brønnum-Hansen H. Danish EQ-5D population norms. *Scand J Public Health*. 2009;37:467–74.
- [32] Wittrup-Jensen K, Lauridsen J, Gudex C, Pedersen K. Generation of a Danish TTO value set for EQ-5D health states. *Scand J Public Health*. 2009;37:459–66.
- [33] Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011;39(7 Suppl):30–3.
- [34] The State Serum Institute. National Health Insurance Service Register [Internet]. 2014. Available from: [http://www.ssi.dk/Sundhedsdataogit/Registre og kliniske databaser/De nationale sundhedsregistre/Sundhedsokonomi finansiering/Sygesikringsregister.aspx](http://www.ssi.dk/Sundhedsdataogit/Registre%20og%20kliniske%20databaser/De%20nationale%20sundhedsregistre/Sundhedsokonomi%20finansiering/Sygesikringsregister.aspx)
- [35] Wallach Kildemoes H, Toft Sorensen H, Hallas J. The Danish National Prescription Registry. *Scand J Public Health*. 2011;39(7 Suppl):38–41.
- [36] Ware J, Sherbourne C. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473–83.
- [37] Pedersen CB, Gøtzsche H, Møller JO, Mortensen PB. The Danish Civil Registration System. A cohort of eight million persons. *Dan Med Bull*. 2006;53(4):441–9.
- [38] Guide P, Copd TO. Global Initiative for Chronic Obstructive Lung A Guide for Health Care Professionals Global Initiative for Chronic Obstructive Disease. *Glob Initiat chronic Obstr lung Dis*. 2010;22(4):1–30.
- [39] Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: The importance of controlling for baseline utility. *Health Econ*. 2005;14(5):487–96.
- [40] Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. *Methods for the Economic Evaluation of Health Care Programmes*. Fourth ed. Oxford University Press; 2015.
- [41] Glick H, Doshi J, Sonnad S, Polsky D. *Economic Evaluation in Clinical Trials*. Oxford University Press; 2007.
- [42] Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics* [Internet]. 2014;32(12):1157–70. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4244574&tool=pmcentrez&rendertype=abstract>
- [43] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. 2009;339(July):157–60.
- [44] Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. *Eur Heart J*. 2017;38(7):500–7.
- [45] Stausholm MN, Egmosse A, Dahl SC, Lilholt PH, Cichosz SL, Hejlesen OK. Stratification of telehealthcare for patients with chronic obstructive pulmonary disease using a predictive algorithm as decision support: A pilot study. *J Telemed Telecare* [Internet]. 2017;23(3):410–5. Available from: <http://journals.sagepub.com/doi/10.1177/1357633X16644129>
- [46] Bertocco L, Haddad DP. Machine Learning in the Prediction of Costs Open for Liver Transplantation. 2017;25:335–43.
- [47] Finansministeriet. Aftaler om den kommunale og regionale økonomi for 2016 [Internet]. 2015. 90 p. Available from: [http://www.fm.dk/publikationer/2007/aftaler-om-den-kommunale-og-regionale-oekonomi-for-2008/download/~media/Files/Publikationer/2008/Download/Aftale\\_om\\_den\\_kommunale\\_og\\_regionale\\_ekonomi\\_2008\\_endelig\\_version.ashx](http://www.fm.dk/publikationer/2007/aftaler-om-den-kommunale-og-regionale-oekonomi-for-2008/download/~media/Files/Publikationer/2008/Download/Aftale_om_den_kommunale_og_regionale_ekonomi_2008_endelig_version.ashx)
- [48] Yoo I, Alafairet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, et al. Data mining in healthcare and biomedicine: A survey of the literature. *J Med Syst*. 2012;36(4):2431–48.
- [49] Vijeta S, Ajai K, Lakshmi P, Ganesh K, Anuradha L. Malaria Outbreak Prediction Model Using Machine Learning. *Int J Adv Res Comput Eng Technol* [Internet]. 2015;4(12):4415–9. Available from: <http://ijarcet.org/wp-content/uploads/IJAR CET-VOL-4-ISSUE-12-4415-4419.pdf>
- [50] Gunčar G, Kukar M, Notar M, Brvar M, Černelč P, Notar M, et al. An application of machine learning to haematological diagnosis. *Sci Rep*. 2018;8(1):1–12.
- [51] Gomes M, Ng ES-W, Grieve R, Nixon R, Carpenter J, Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Med Decis Making* [Internet]. 2012 [cited 2014 Oct 16];32(2):350–61.
- [52] Rabe-Hesketh S, Skrondal A. *Multilevel and Longitudinal Modeling Using Stata*. Volume I: Continuous Responses. Third edit. Texas: Stata Press; 2012.

#### Address for correspondence

Flemming Witt Udsen, Department of Health Science & Technology, Aalborg University, Fredrik Bajers Vej 7C, DK-9220 Aalborg Ø, Denmark.  
E-mail: [fwu@hst.aau.dk](mailto:fwu@hst.aau.dk)