COPD in primary care: Towards simple prediction of quality of life, exacerbations and mortality

Siebeling, L.

Citation for published version (APA):
CHAPTER 8

GENERAL DISCUSSION
Introduction

Most patients with Chronic Obstructive Pulmonary Disease (COPD) are managed almost completely in primary care. Health-related quality of life (HRQL) is one of the most important outcomes from a patient’s perspective. At least up to 2008, there were no valid models predicting HRQL. These three observations were taken into consideration in the development process of ICE COLD ERIC, an ongoing prospective cohort study on 409 primary care COPD patients from the Netherlands and Switzerland (follow up from 2008 until 2014, still ongoing at the time of writing). The primary aim was to develop models to predict the course of four separate domains of HRQL, dyspnoea, fatigue, emotional function and mastery, as distinguished by the Chronic Respiratory Questionnaire (CRQ). Secondary aims were to develop models to predict exacerbations and mortality.

Currently, in 2014, only one single prediction model exists for prediction of COPD-specific HRQL in COPD patients. Many existing prediction models focus on mortality as outcome and were developed in secondary or even tertiary COPD populations. The only model that predicts HRQL has limitations. First, this Health-Activity-Dyspnoea-Obstruction (HADO) model was derived in men and its applicability in women is uncertain. Second, HADO discriminated only moderately between patients with different levels of HRQL, as measured by the CRQ ($R^2=0.21$). Third, it was not corrected for overoptimism nor has it been externally validated yet. External validation has only been done for this HADO model with regard to mortality outcomes, not to HRQL outcomes.

To enhance the use of prediction models in primary care settings, models must be practical. First, the predictors must be available in these settings and second, the collection of the predictors must be easy and fast, taking into consideration that a GP or nurse practitioner will have around 10 to 30 minutes per patient. Measurement of potentially important predictors and/or outcomes such as exercise capacity, physical activity and exacerbations are challenging and usually difficult or impossible in primary care settings. The six-minute-walk-test (6MWT), for example, is a familiar test to measure exercise capacity, however, infeasible in most primary care settings. Simpler test have been developed that showed high correlations with the 6MWT, such as the one-minute sit-to-stand-test (STST) and the handgrip test. However, these simpler tests had not yet been tested on their predictive properties. With regard to physical activity, precise measurement is expensive and time-consuming and therefore unsuitable for primary care settings. Several questionnaires have been developed for measuring physical activity but the majority have important limitations. Finally, COPD exacerbations are often ascertained through patient self-reports although little is known about the accuracy of this method. These were important issues to explore in order to eventually develop simple prediction models for use in primary care settings.
Main findings

We found that COPD-specific domain-specific HRQL after 6 and 24 months in primary care COPD patients could be reasonably well predicted by the corresponding domain-specific scores at baseline. These models could be improved by adding between one to six other predictors to the strongest predictor, such as the Hospital Anxiety and Depression Scale (HADS), the feeling thermometer (FT) and the other domain-specific CRQ scores. All information necessary for the predictor variables can be easily collected in current primary care settings. The predictions were close to the average observed values and within limits demarcated by the minimal important difference of 0.5. This indicates good calibration. Explained variances were high and above 0.4 with the exception of mastery whereas, for example, the HADO score showed an explained variance of 0.21.

We also found that simple tests for measuring exercise capacity, such as the one-minute STST and the handgrip strength test, are strongly associated with mortality, significantly associated with HRQL and not associated with exacerbations.

With regard to measuring physical activity, we found that the practical Longitudinal Ageing Study Amsterdam Physical Activity (PA) Questionnaire (LAPAQ) was unsuitable for exact measurement of PA in older adults in comparison with a triaxial accelerometer. However, it may be used to determine if the PA level of a certain person is above the recommended level.

With regard to measuring exacerbations, we found that the common way of measuring exacerbations by patient’s self-report is only moderately accurate, when compared to central event adjudication by a committee. 48% of the patients correctly reported on the total number of exacerbations. More patients over-reported than under-reported their exacerbations, 34% versus 18%. Single expert adjudication, meaning that one expert adjudicates the exacerbation, also turned out to be more accurate than patients’ self-report.

Methodological considerations

Generalizability

Our population consists mostly of patients with moderate COPD; the majority had a GOLD stage 2 (66%) or GOLD stage A (41%) at baseline, had mild to moderate dyspnoea (65% has MRC score ≤ 2) and had no exacerbation in the year before inclusion (67%). Most other COPD cohorts, such as -the ECLIPSE- (Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints) and -the INSPIRE- (Investigating New Standards for Prophylaxis In Reducing Exacerbations) cohorts consist of more severe COPD patients. In both these cohorts, the majority had a GOLD stage 3 and had one or more exacerbations in the previous year. These studies included more severe COPD patients than the ICE COLD ERIC study, which can be explained by differences in setting and inclusion criteria. By developing our models in a population derived from primary
care, we hope to avoid validation issues for use of the models in primary care. Often, prediction models perform very well in the dataset in which they have been developed. But when applied to new patients or other populations, predictive performance may be less. It is possible to (partly) correct for this so-called overoptimism by, for example, using repeated bootstrap resampling. In every bootstrap sample, a new dataset from the original dataset is constructed in which a new model is built following the same procedure as for the original model. In our analysis, we computed an estimate of overoptimism as the average difference between the performance in the bootstrap samples and the performance in the original sample. This overoptimism was averaged over ten data sets (constructed to deal with missing values) and subtracted from the original explained variance to obtain an overoptimism-corrected performance measure 50.

**Primary care focus**

The primary aim of our study was to develop models to predict HRQL in primary care practice. To enhance the use in primary care, the models must consist of predictors easily collectable in primary care settings, unlike, for example, the 6-minute-walk-test for exercise capacity testing. Our candidate predictors were specifically selected on their availability and practicality in current primary care settings. For exercise capacity testing, for example, we chose to use the one-minute sit-to-stand-test, which only requires a watch and a chair. It is a simple test that any GP is able to use and it only takes 2 minutes (1 minute of explaining and 1 minute of actual testing). In the ICE COLD ERIC cohort, we found that this test as a measure of exercise capacity was strongly associated with mortality and significantly associated with HRQL 31. We believe that this test is an attractive alternative to the 6MWT, which requires both more time and space.

With regard to predictor selection and model fitting, we used penalized linear regression (“least absolute shrinkage and selection operator” (lasso)) 32. This technique enabled us to shrink regression coefficients in order to obtain more parsimonious models. In this context parsimony means: the fewer predictors the better, while maintaining good predictive performance. Again for practicality reasons, we simplified the potential application of the prediction models by creating nomograms, although, arguably, the use of internet or smartphone applications would still be better.

**Missing data**

With regard to missing data and loss-to-follow-up percentages, our cohort also differs from other COPD cohorts from the literature. High percentages of loss-to-follow-up are frequently seen in COPD cohort studies. In the ICE COLD ERIC cohort, the percentage of loss-to-follow-up was low compared to other cohorts; 4.6% compared to for example 26% in the INSPIRE cohort, both after 2 years of follow-up. Our study nurses tried to keep close contact with the patients and they did their absolute best to reach the patients for the follow-up
measurements, for example, calling in the evening time in case patients had full
time jobs. Patients also received Christmas cards every year and regular updates of
the study. In our study, the loss-to-follow up percentage in Switzerland was es-
pecially low after 2 years (1% versus 5% for the Netherlands) probably because all
measurements were performed by the same study nurse from the beginning of the
study until the end. In the Netherlands, several study nurses were needed to col-
lect all measurements. Even now, in 2014, after almost 5 years of follow-up, the
loss-to-follow-up percentages are still relatively low: 7% for Switzerland versus
30% for the Netherlands. For the total cohort, the loss-to-follow-up percentage
after 5 years is 21%.

Implications for clinical practice
Prediction models
Our models can be used in clinical practice to inform patients about their expected
HRQL. The expected change of overall and domain-specific HRQL (4 domains:
dyspnoea, fatigue, emotional function and mastery) within the next 6 or 24
months can be predicted. Since all predictors are available in primary care, any
GP or nurse practitioner can use the models. Depending on the outcome per
HRQL-domain, GPs and patients can discuss and prioritize different possible
treatment actions in a shared decision-making context where feasible.

![Nomogram for CRQ dyspnoea outcome at 6 months](image)

**Figure 8.1.** Nomogram for CRQ dyspnoea outcome at 6 months
FEV<sub>1</sub>=Forced Expiratory Volume in one second, FT=Feeling Thermometer, CRQ=Chronic Respiratory Questionnaire
For example, to predict dyspnoea in 6 months, 4 predictors need to be collected, see Figure 8.1;  
1. Forced expiratory volume in one second, % of predicted (FEV₁%)  
2. The patient’s score on the FT  
3. The patient’s CRQ score for dyspnoea  
4. The patient’s CRQ score for fatigue.

**Ad 1.** To obtain FEV₁%, spirometry needs to be performed after bronchodilation. From these 4 predictors, this will be the most time-consuming; spirometry is estimated to take 30 minutes at most. Nevertheless, we don’t see this time-consuming measurement as an obstacle for using this model since spirometry is usually performed on a regular basis (e.g. annually) in COPD patients.

**Ad 2.** The score on the FT needs to be collected. The FT is a visual analogue scale presented as a thermometer with 100 marked intervals. The worst score is 0 (=dead) and the best score is 100 (=perfect health) to reflect how patients have felt in the previous seven days. This measurement won’t take more than 1 or 2 minutes.

**Ad 3.** The CRQ dyspnoea score is needed. This score is a summary of 5 questions of the CRQ, which will take 2-3 minutes to collect (question 1-5). Patients answer each question on a seven-point scale to express their degree of disability from 1 (maximum impairment) to 7 (no impairment).

**Ad 4.** Finally, the CRQ fatigue score is needed, which is a summary of 4 questions of the CRQ (question 8, 11, 15, 17). This will take another 2-3 minutes.

In summary, assuming that spirometry will be performed anyway, it takes less than 10 minutes to be able to predict the CRQ dyspnoea domain in 6 months.

**Patient 1, example:** Patient 1 feels quite stable over the last few months and she does not present any more COPD-related symptoms than usual. She visits the practice for her annual follow-up visit. Spirometry is performed, FEV₁% of predicted is 40. The other scores are as follows: FT=60, CRQ-dyspnoea score =5 and CRQ-fatigue score=3. This results in a predicted dyspnoea score of 4.65 in 6 months, a decline of 5-4.65 =0.35. This decline is less than 0.5, and interpreted as having no clinical importance.

*See figure 8.2 for explanation of deriving this score.*

Probably, this patient and her GP will discuss together to keep the treatment as it is, with regard to her dyspnoea. Additionally, it may be important to check the other domains as well using the other models in the same way.
**Figure 8.2.** Example of using the nomogram
From each predictor scale, draw a vertical line up through the points scale (upper scale) and sum all points. Next, fill in the sum value in the total points scale, draw a vertical line through the outcome variable (here dyspnoea at 6 months) and read off the predicted outcome.

*Patient 2 example, part 1:* Patient 2 feels quite stable over the last few months and he does not present any more COPD-related symptoms than usual. He visits the practice for his annual follow-up visit. Spirometry is performed, FEV₁% of predicted is 60. The other scores are as follows: FT = 50, CRQ-dyspnoea score = 6.0 and CRQ-fatigue score = 4.5. This results in a predicted dyspnoea score of 5.5 in 6 months, a decline of 6-5.5 = 0.5, which is interpreted as clinically important. Probably, this patient and his GP will discuss ways to prevent this predicted decline in dyspnoea and they will seek for appropriate treatment. Additionally, it may be important to check the other domains as well using the other models in the same way to be able to prioritize different possible treatment actions.

**Treatment options in relation to HRQL**
Pulmonary rehabilitation has a beneficial effect on all domains of HRQL, and on dyspnoea in particular, even in patients with mild disease. According to Lacasse, on average, pulmonary rehabilitation improves the CRQ dyspnoea score by more than 1 point, clearly exceeding the minimal important difference of 0.5, using the following definition of pulmonary rehabilitation: ‘any inpatient, outpatient, or home-based rehabilitation program of at least four weeks duration that
included exercise therapy with or without any form of education and/or psychological support delivered to patients with exercise limitation attributable to COPD. Our models can be used in practice to show patients their expected course on (different domains of their) HRQL. In case of an expected decline in one or more domains, they may assist the physician and patient to discuss and prioritize treatment decisions. In our example patient 2, the option of a pulmonary rehabilitation program may be discussed to prevent the expected decline in the dyspnoea domain.

Compared to other, more commonly prescribed, treatments such as inhaled bronchodilators and inhaled corticosteroids (ICS), pulmonary rehabilitation improves HRQL more\(^{36}\).

ICS, long-acting beta-agonists (LABA) and tiotropium improve HRQL more than placebo, although the potential benefits of ICS should be balanced against potential side effects, such as an increased risk of pneumonia\(^{39-44}\). Till now, no relevant improvements in HRQL are seen for combined ICS/LABA versus ICS alone\(^{45}\), combined ICS/LABA versus LABA alone\(^{46}\) or combined ICS/LABA versus tiotropium\(^{47}\). No clinically significant or relevant improvements in HRQL were found comparing ipratropium versus LABA or SABA (short-acting beta-agonists) or a combination of ipratropium/SABA\(^{48,49}\). LABA combined with ipratropium or tiotropium seems to be superior to LABA or tiotropium alone in relation to HRQL, but differences were small and inconsistent\(^{50,51}\). Although not on mortality, there was a significant benefit of lung volume reduction surgery in COPD patients with diffuse emphysema on HRQL, exercise capacity and lung function\(^ {52}\). Self-management as part of integrated disease management improved HRQL\(^ {53}\) but no other reviews could be found on self-management, nor on the effect of oxygen therapy on HRQL.

**Patient 2 example, part 2:** Since the expected decline in the dyspnoea domain, patient 2 and his GP discussed the situation and decided to follow a pulmonary rehabilitation program. After 6 weeks, directly after the rehabilitation program, the patient’s CRQ dyspnoea score was improved, 6.5 (versus 6.0 before the program). Also the score on the FT was improved, 60 (versus 50 before the program). FEV\(_1\)% of predicted and CRQ fatigue score remained stable, 60 and 4.5 respectively. Incorporating these new scores into the prediction model (See Figure 8.4), predicted CRQ dyspnoea in 6 months turned out to be 5.9. Before the rehabilitation program, dyspnoea was predicted to decline from 6.0 to 5.5 in the next 6 months but after the rehabilitation program, it was predicted to remain stable (6.0 versus 5.9).

**Other implications for clinical practice**

We found that the simple exercise capacity tests such as the STST and the hand-grip strength test are associated with mortality and HRQL. Since other tests, such as the 6MWT, are not available in primary care settings, these simple tests could be an attractive alternative.
Although the LAPAQ turned out to be unsuitable for exact measurement of PA in older adults, it still may be used to determine if the PA level of a certain person is above the recommended level. This could be enough in primary care settings since GPs are able to identify those adults that need attention with regard to stimulating being more physically active.

With regard to measuring exacerbations, we found that the use of central event adjudication or even single expert adjudication is a better method than patients’ self-report. In case exacerbations are used as outcome to study treatment effects, this could lead to misclassification. Since more patients over-reported exacerbations than under-reported exacerbations, there was more false positive than false negative misclassification, resulting in underestimation of treatment effects. Adjudicating exacerbations can substantially minimise this type of outcome event misclassification and reduce bias in estimates of treatment effects. Adjudication of exacerbations could also considerably reduce sample size requirements and costs of randomised controlled trials and observational studies.

**Implications for future research**

We have developed prediction models to predict important domains of HRQL in primary care COPD patients. For more dependable implementation of our models, external validation is needed. In the future, we intend to collaborate internationally with other COPD cohort studies and we intend to perform validation studies, in which our models will be validated in other (primary care) COPD populations. Assuming that they will perform well, we would like to incorporate treatment advices into the models by using data from randomized controlled trials and meta-analyses in order to estimate how the prediction of HRQL will change when adding treatments such as smoking cessation, pulmonary rehabilitation or specific drugs. Finally, cost-effectiveness evaluation of prediction models for HRQL in COPD should be performed to determine if it is worth the effort incorporating these models into practice. After successful completion of these steps, our models may support treatment selection based on the individual patient’s prognosis. Eventually, with regard to practicality, it would be optimal to create internet and/or smartphone applications to further enhance the use of the models.

Treatment decisions will be tailored better to the needs of the individual patient, probably resulting in a more cost-effective management of COPD with less unnecessary treatment prescriptions and a better COPD-specific HRQL.
Reference List

1 Heyworth IT, Hazell ML, Linehan MF, Frank TL. How do common chronic conditions affect health-related quality of life? Br J Gen Pract 2009 Nov;59(568):e353-e358.


32 http://www.jstor.org/stable/2346178


35 Jacome CI, Marques AS. Pulmonary rehabilitation for mild chronic obstructive pulmonary disease: a 1 systematic review. Respir Care 2013 Oct 8.


