Outcome assessment in inpatient pulmonary rehabilitation: clinical results and methodological aspects
van Stel, H.F.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Summary and discussion
This gave me no concern, as I have ever confined myself to facts.

Rudolph Erich Raspe
The Surprising Adventures of Baron Munchausen, 1785
8.1 Introduction

This dissertation describes the clinical outcome of comprehensive, multidisciplinary inpatient pulmonary rehabilitation for patients with asthma or COPD and addresses several problems with measuring that outcome. Two meta-analyses have demonstrated that pulmonary rehabilitation is effective in patients with COPD: it reduces symptoms, especially breathlessness; and it improves health status and functional exercise capacity [1;2]. However, most studies on pulmonary rehabilitation deal with relatively stable patients with COPD treated in an outpatient setting. Much less is known about patients with other chronic lung diseases, such as asthma, about patients with unstable disease and about inpatient pulmonary rehabilitation (IPR).

The general introduction in chapter 1 describes the basic topics of asthma and COPD: definitions and clinical features, epidemiology, comorbidity, exercise limitation, psychosocial functioning, health status and pharmacotherapy. Pulmonary rehabilitation and the measurement of outcome in pulmonary rehabilitation are described more extensively.

8.2 Outcome study of IPR

Chapters 2 and 3 of this dissertation describe the characteristics and the treatment outcome of 140 patients participating in the IPR programme of Asthmacenter Heideheuvel in Hilversum, The Netherlands. This study includes 56 patients with asthma and 84 patients with COPD who were referred to the asthmacenter during 1996 and 1997. The main reason for referring patients to pulmonary rehabilitation is when patients still have disabilities and handicaps despite optimal medical treatment. The major disabilities and handicaps are dyspnoea, reduced exercise capacity, and restrictions in daily and social activities. There are several reasons to refer patients to inpatient instead of outpatient pulmonary rehabilitation [3]. When patients are severely handicapped, they may not have enough functional ability to attend outpatient sessions. This happens often to patients with recent and/or recurrent hospitalizations. When an outpatient or home-based programme is not available in the vicinity of a patient, inpatient rehabilitation may be the only alternative. For some patients, intensive monitoring is required for an integrated description of the illness behaviour of a patient. Furthermore, patients may be referred for specific interventions such as nutritional therapy, behavioural interventions to correct psychosocial problems, and teaching of coping skills.
8.2.1 Outcome of IPR: illness severity

The patients referred to Asthamacenter Heideheuel appear to be more severely ill than other patient groups reported in the scientific literature on outpatient and inpatient pulmonary rehabilitation. First of all, 80% of the patients were hospitalized in the year before IPR, an average of 53 days with a total of almost 6000 days. This is much higher than the 9 to 20 days in hospital in the year before IPR reported in other studies on IPR, including several studies in the Netherlands. The illness severity of this group of patients is further illustrated by the large amount of comorbidity: 80% had at least one comorbid condition, mainly cardiovascular diseases, mobility disorders, overweight, psychosocial disorders and allergy. On top of that, half of the patients with asthma and three-quarters of the patients with COPD had osteoporosis. The number of exacerbations, which is an important parameter of illness severity and often a primary outcome in drug trials, was very high: 7 per year in asthma and 10 per year in COPD, causing 80 admissions to the intensive treatment facility of the asthamacenter. This exacerbation rate is much higher than reported in several studies on outpatients with COPD [4-6].

8.2.2 Outcome of IPR: main results

The major results of the IPR programme of Asthamacenter Heideheuel were a decrease in the use of oral corticosteroids, clinically relevant improvements in health status and emotional well-being and a large decrease in hospitalization. The median daily dose of oral corticosteroids in the asthma group dropped from 5 to 0 mg. The decrease in oral corticosteroids in the COPD group was smaller, but still highly significant. There were three-to tenfold decreases in the number of hospital admissions, the number of patients with admissions and the number of days in hospital in the year post-IPR as compared to pre-IPR. Furthermore, patients reported large subjective improvements on knowledge about lung disease, medication and correct use of medication; exercise capacity; disease symptoms; and performance of daily, social, and leisure activities. All domains of the Quality of Life for Respiratory Illness Questionnaire (QoLRIQ) showed improvements above the threshold for clinically relevant change. In several domains even the lower limit of the confidence interval was above this threshold. Emotional well-being, experienced invalidity and self-assessed health status showed moderate to large effect sizes. An important finding of this study is that patients with asthma not only improved, but also maintained their improvements in health status and well-being, despite small, non-significant deteriorations in the 12-month follow-up period. In contrast, the patients with COPD deteriorated in the follow-up period almost to the pre-treatment level, as was found by others.
8.2.3 Outcome of IPR: deterioration during follow-up for COPD

There is a contradiction between the diminishing of the initial treatment effects on health status and psychosocial functioning on the one hand and the large decrease in hospitalization post-IPR in the group of patients with COPD. It should be studied why patients report deterioration in health status and well-being while they clearly improved on clinical parameters. Furthermore, the quick deterioration of health status in patients with COPD is worrying, both from the perspective of the patients involved and from the perspective of third-payer parties. Therefore, the effectiveness of methods to maintain the large initial treatment effects should be studied. Maintenance exercise programs may be useful for preserving the effects of exercise training [7,8], but the benefits of booster sessions [9], after-care programs [10], repeating pulmonary rehabilitation [11], and voluntary sports groups have not been studied sufficiently yet. The development and analysis of a maintenance programme for Asthmacenter Heideheuel is however hampered by the supra-regional function of this asthmacenter: regular maintenance sessions on an outpatient basis are only possible for patients living within close travelling distance, which makes it very difficult to recruit sufficient patients within an acceptable research period.

8.2.4 Outcome of IPR: no improvement in exercise capacity?

The IPR programme did not result in an improvement in exercise capacity - on average. The overall lack of improvement in walking distance in this study is not unexpected, although almost all studies on outcome of pulmonary rehabilitation show improvement of exercise tolerance. The focus of the treatment programme described in this study on both self pacing to prevent exhaustion and dyspnea, and exercise training to enlarge endurance and muscle strength, will cause a number of patients to walk less after the treatment and others to improve their walking distance. This is reflected in the wide range of change, from -193m to +298m. Composite analysis (explained in chapter 5) showed that 62% of the patients with asthma and 36% of the patients with COPD improved in 2 or more factors of functional exercise tolerance, while 22% and 28% respectively showed deterioration.

Although most outcome studies of pulmonary rehabilitation report both improvement in health status and in exercise capacity, there seems to be no relation between change in health status and change in walking distance. Reardon and coworkers suggested that in a multifaceted pulmonary rehabilitation program that addresses several functional areas, it is possible that health status might be improved without substantial improvement in exercise endurance [12]. Another explanation may be that it is difficult to improve exercise capacity in severely dyspnoeic patients, as was recently found by Wedzicha et al., who randomised patients with MRC grade 5 to outpatient or home-based pulmonary rehabilitation and found no improvement in both groups [13]. The majority of the patients referred to
Asthmacenter Heideheuvel also have MRC grade 5: 77% of the patients with asthma and 83% of the patients with COPD.

**8.2.5 Outcome of IPR: dropout and imputation**

There was a substantial dropout, both in the IPR-phase and in the follow-up phase of the study. About twenty percent of the patients dropped out at each subsequent assessment, totalling up to 64% dropout. This is comparable to the dropout rates of 50% and 60% in some studies on IPR, but higher than in other studies on IPR (range 15 to 40%). The dropout was equally divided between illness- or treatment-related dropout and study-related dropout. There were no significant differences at baseline or in change between completers, illness/treatment related dropouts and study dropouts, implying that there was no selective dropout of the most severely ill patients, as was found by Ketelaars et al. [14]. However, the high drop out rate threatens the validity of the results described in chapter 2 and 3. Therefore, imputation of missing data was done to test the robustness of the findings (see chapter 4). Two major outcome measures were selected for this analysis: health status (the total score of the QoLRIQ) and hospitalization (number of hospitalizations and days in hospital). Imputation with specialised imputation software gave counterintuitive results for hospitalization. Therefore, a decision model for imputation was developed, using 4 different scenarios: *optimistic*, *realistic*, *sombre* and *pessimistic*. In the optimistic scenario, both treatment dropouts and study dropouts are assumed to improve as much as the completers. This obviously leads to a confirmation of the positive outcome of IPR described above. In contrast, the pessimistic scenario assumes that study dropouts do not improve at all and that treatment/illness-related dropouts deteriorate twice the minimal important difference in health status during IPR, and will be twice as much in hospital in the year post-IPR then in the year before IPR. This clinically implausible scenario does not change the improvements in health status and hospitalization described above. The improvement in health status during IPR remained significant and above the threshold for clinically relevant change in both the asthma group and the COPD group. The decrease in hospitalizations in the COPD group remained highly significant in all scenarios. The decrease in hospitalizations in patients with asthma was significant up to the sombre scenario. The results of this imputation analysis show that the positive outcomes of IPR are very robust and valid, despite the large number of missings.
8.3 Factor analysis of the 6MWT

Chapter 5 describes an analysis of exercise performance in the six minute walking test. An important treatment goal in pulmonary rehabilitation is improvement of functional performance, which can be achieved by exercise training and by training of self-pacing skills. However, exercise training and training of self-pacing skills may have contradictory effects on walking distance, which is the usual outcome parameter for exercise capacity. Changes in other aspects of exercise capacity, such as dyspnoea, oxygen saturation and walking technique may be of equal importance for improvement of functional performance of an individual patient, but are seldom reported.

The first aim of the study in chapter 5 was to describe baseline performance in the six minute walking test with more factors than only walking distance. An exploratory factor analysis was performed on physiological measurements, dyspnoea ratings, and walking distance recorded during pre-treatment six minute walking testing in 83 patients. The second aim was to assess if the use of multiple factors adds to walking distance in describing change in exercise performance. Factor analysis is a data-reduction technique that consists of two steps: clustering of variables with shared variance, which yield factors, and then simplifying the factor structure by varimax rotation, which improves interpretability. The factor analysis resulted in a clinically interpretable 4-factor structure which explained 78% of the total variance. Performance in the six minute walking test can be described by four factors: heart rate pattern, endurance capacity, impairment of oxygen transport, and perceived symptoms.

The added value of using more factors to describe change in exercise capacity is shown in two ways. First, by using composite analysis, which is here defined as a simultaneous qualitative outcome analysis of several related factors. Composite analysis showed that 29 of 53 patients improved in three or four factors, and 7 patients deteriorated in three factors, while there was no mean change in walking distance. Second, multiple regression analysis showed that self-perceived change in exercise tolerance is mainly explained by change in walking distance, but also by less desaturation and less dyspnoea. This may explain why the majority of the patients (42 out of 53) reported improvement in exercise tolerance despite the overall lack of change in walking distance.

The main advantage of using multiple factors to describe performance is the possibility of assessing change in several aspects of exercise capacity simultaneously. There is however a clinical problem in judging the importance of the observed changes in all factors together. It will depend on the specific treatment goal for patients and if the size of changes in the improvement in one factor outweighs the deterioration in other factors. This judgment may be aided by assessing the self-perceived change in exercise tolerance and by comparing the
size of the changes to the minimal important difference for each factor (as far as these are known).
For Asthmacenter Heideheuvel, it will be interesting to assess the outcome of different treatment goals related to functional capacity by describing and testing of the expected outcome in terms of the four factors of the six minute walking test described above.

8.4 Longitudinal properties of the QoLRIQ

The question of the size of a clinically relevant change in an outcome measure, the minimal important difference, is addressed in chapter 6. Improvement of health status is one of the major goals of pulmonary rehabilitation, so therefore a validated and responsive health status questionnaire should be used to assess the outcome of the treatment. The outcome study described in chapter 2 and 3 used the Quality of Life for Respiratory Illness Questionnaire (QoLRIQ), which is a health status questionnaire developed for and validated in both patients with asthma and patients with COPD. Because the longitudinal measurement properties had not been evaluated yet, these were studied in 108 patients who completed the QoLRIQ and related outcome measures pre- and post-treatment; and a global rating of change in disease symptoms post-treatment. The QoLRIQ uses a 7-point response scale.

First, the sensitivity to change was assessed by computing the statistical significance and relative magnitude of changes detected by the QoLRIQ. Second, the longitudinal validity of the QoLRIQ was assessed by computing correlation coefficients between change scores in QoLRIQ-domains and change scores from related outcome measures. Third, the reliability of the change score was assessed by computing the intraclass coefficient of change. Fourth, the size of a minimal important difference (MID) was computed with a retrospective global rating of change question. Because several authors have questioned the validity of retrospective assessment of change [15;16], the validity of that method was studied, and the MID was also determined with alternative methods such as computing the standard error of measurement and using the benchmarks for effect sizes.

8.4.1 QoLRIQ: sensitivity to change

All domains and the total score of the QoLRIQ showed highly significant changes (all p < 0.0002). The relative magnitude of change was assessed by computing standardised response means. A standardised response mean is interpreted as an effect size: 0.2 represents a small change; 0.5 a moderate change and changes of 0.8 or higher are interpreted as a large change. Standardised response means ranged from 0.46 to 0.90. The largest absolute and relative changes were seen in the total score and domains representing
daily functioning (general activities, daily/domestic activities, social: activities).
QoLRIQ-change scores were highly correlated with self-rated change in disease symptoms,
with change in self-assessed health status and with change in several domains of the Rand-
36 and the Medical Psychological Questionnaire for Lung Patients. The intraclass correlation
coefficient of change in the QoLRIQ-total score was 0.90, indicating a high reliability of this
change score.

8.4.2 QoLRIQ: size of the MID
The MID was assessed with both anchor-based and distribution-based methods: from
retrospective assessment of change in disease symptoms, from change in self-assessed
health status, by computation of the standard error of the measurement and from the effect
size benchmarks. The 'retrospective' method gave a MID for the QoLRIQ-total score of 0.5
point in both positive and negative direction. Categorizing by one-unit changes in self-
assessed health status gave positive and negative MIDs for the QoLRIQ-total score of 0.4
and 0.6 respectively. The standard error of the measurement for the domains, ranging from
0.4 to 0.65, had a mean of 0.49 points. Using a moderate effect size gave MIDs quite
similar to the MIDs based on the standard error of the measurement. So, 0.5 seems the best
point estimate for the MID, within a range of 0.4 to 0.6. The mean value for the standard
error of the measurement is similar to the values found by Wyrwich and coworkers [17-19],
confirming a value of 0.5 as the threshold for meaningful change in domains of
questionnaires with a 7-point response scale.

8.4.3 QoLRIQ: validity retrospective computation
The retrospective computation of MIDs proved not to be valid in this group of patients.
First, the retrospective assessment of change was significantly correlated to post-treatment
health status but not to pre-treatment health status, as was found by Norman et al [15]. This
indicates that the global assessment of change is determined by post-treatment health
status. Second, retrospective assessment of change was significantly higher than serial
assessment of change, as was found by Fischer et al [16]. Third, the mean change in the
'unchanged' group had a large 95% confidence interval and was significantly different from
zero.

8.4.4 QoLRIQ: conclusion
This study shows that the QoLRIQ is sensitive to change, has a good longitudinal validity
and reliability, and has a MID of 0.5 points for patients with moderate to severe asthma or
COPD. The MID can be used for power calculations in future studies and to assess whether
observed changes are clinically relevant. These results enable the use of the QoLRIQ as an
outcome measure in clinical trials with patients with moderate to severe asthma or COPD.
The remaining questions pertain to the longitudinal measurement properties in less severely ill patients and the use of the QoLRIQ and its MID in clinical practice.

8.5 Assessing the patient's view of outcome

Because of the large variation in individual problems and the essential role of motivation in pulmonary rehabilitation, a major feature of the inpatient pulmonary rehabilitation program of Asthmacenter Heideheuvel is the use of individualised treatment goals. After a 1-week multidisciplinary diagnostic phase, an extensive integrated description of the patient specific problems is constructed. Individualised treatment goals based on this problem description are formulated by the treatment team in consultation with the patient. There are however some problems with the usual, standardised outcome assessment when assessing the outcome of a multidisciplinary treatment programme consisting of several simultaneous interventions. Most outcome measures address common problems, but they neither represent the specific problems experienced by an individual patient nor the specific intervention for addressing that problem [20]. Outcome assessment in pulmonary rehabilitation may be improved by looking at the individual treatment goals of a patient, and by including the patient's view of outcome. Chapter 7 describes a new and complementary method for assessing the patient's view of outcome by asking patients to assess subjectively their attainment of the individualised treatment goals they had aimed to achieve during the IPR program. The therapists were asked to give a similar assessment. Attainment of the treatment goal was scored on a 6-point response scale: was the goal attained not at all, barely, a bit, partly, largely or completely (1 — 6). The 79 patients who participated in this study had a total of 540 treatment goals. 60% of the 488 goals scored by the patients were labelled as successful (score 5 or 6). The patients had a significantly higher median attainment score than the therapists (5 versus 4). The sensitivity to change of attainment scoring was assessed by computing the standardised response mean which, with a value of more than 3, is much higher than the already large standardised response mean of 1 of the QoLRIQ (in 42 patients who also participated in the outcome study described in chapter 2 and 3). The relative efficiency of measuring change with attainment scoring versus the QoLRIQ total score was high. Despite the lack of correlation between attainment scores from patients and change in QoLRIQ total score, there was sufficient evidence for the construct validity. Several treatment goals with at least 10 occurrences showed moderate to high correlations with change in a closely related external outcome measure. The correlations were higher with increasing specificity of the external criterion. The need for assessing the outcome of individual treatment goals is confirmed by the large number of treatment goals that did not have an equivalent in the standardised outcome.
measures. The reliability of attainment scoring, which was assessed by computing the interrater agreement, was very low. The low reliability and the difference in the attainment scores between patients and therapists was expected and may be explained partly by the different points of view, but also because there was no common criterion for assessing the amount of change. A necessary improvement of attainment scoring is including a clear description of the different expected outcome levels in verifiable terms, such as in Goal Attainment Scaling. Two different standardised versions of Goal Attainment Scaling (which is reliable, valid, and responsive but time-consuming) seem very promising for measuring outcome in multiple individualised interventions [21;22].

8.6 "Complex problems"

An important characteristic of the patients referred for inpatient pulmonary rehabilitation in AsthmaCenter Heideheuvel is the large variation in individual problems associated with or interacting with the chronic lung disease. In Heideheuvel, this is called "complex problems". "Complex problems" may result in inadequate disease behaviour and insufficient self-management. However, the term "complex problems" is not defined or operationalised, which hinders both treatment and research. Without operationalising "complex problems", it remains very difficult to demonstrate that the patients referred to Heideheuvel are indeed a special group that requires interdisciplinary individualised care. Furthermore, if one cannot distinguish clearly between different levels and types of "complex problems", it is not possible to develop and validate new treatment modalities, which requires well-defined patient groups and well-described, replicable interventions [23].

A possible definition of "complex problems" is the combination of and interaction between somatic and psychosocial morbidity. A framework for operationalising could be based on the International Classification of Functioning, Disability and Health, developed by the World Health Organisation [24]. The degree of disablement in body functions, body structures, activity limitations and participation restrictions can be measured using standardised and validated function tests and questionnaires. This allows a detailed description of the illness severity of a patient or a group of patients. The characterisation and quantification of the interaction between disease, comorbidities and disease behaviour is much more difficult. This is an essential component of "complex problems" and should be observed and rated by the treatment team. This requires an unambiguous observation and rating method which needs to be developed and validated. Maybe a standardised "illness complexity index" can be devised, consisting of severity stagings for lung function impairment, comorbidity, exercise capacity, psychological functioning, social functioning, and health status; types of disease behaviour; and types of interaction.
Another way of gaining insight into the complexity of problems may be to use the functional status concept from Leidy [25]. The dimensions of functional status are functional capacity, functional performance, functional reserve and functional capacity utilization. Clarifying the relation between actual performance, its precursors (such as pain, sleep, attitudes and emotions) and influential variables may help to untangle the complexity.

8.7 The control group problem

Another point that requires more discussion is the lack of a control group. Study designs with randomisation to experimental and control groups eliminate bias and show clearly if the experimental treatment is better than the control treatment. There were however two problems with using a randomised design. The first problem is that inpatient pulmonary rehabilitation is not an "experimental" treatment in the Netherlands. Therefore, randomising patients to conventional care would imply withholding them an accepted treatment, which is an ethical dilemma. The control group problem [26] plays a role in almost all studies on outcome of inpatient pulmonary rehabilitation: until now only one randomised controlled trial of IPR versus conventional care has been published [27]. That study was feasible because of a 3 to 6 months waiting list and by excluding unstable patients [28]. This trial established the effectiveness of IPR by showing positive long-term effects on health status and exercise capacity for IPR versus conventional care in patients with COPD.

The second problem that prohibited a randomised design was a practical one: the short waiting list on the one hand, and the long treatment duration and long follow-up period on the other hand, made it impossible to form a control group. The asthmacenter would have been half empty, which is economically unacceptable; the patients and control group would have had to wait for about a year and a half before receiving IPR, which is ethically and clinically unacceptable. An alternative may be to form a control group of patients from hospitals that only seldom refer to pulmonary rehabilitation programs. However, this will introduce selection bias because the criteria for referral to IPR are not strict. Furthermore, the question is not if IPR is better than "standard" care - that has already been shown by the randomized trial by Goldstein and colleagues [27]. However, there remain a number of questions in IPR which can and should be studied with randomized controlled trials. First, the effectiveness of IPR in patients with asthma has not yet been established, so a randomised trial is badly needed. Second, the optimal treatment setting and treatment content for both patients with asthma and patients with COPD is still unknown, so randomised controlled trials of inpatient versus outpatient pulmonary rehabilitation and trials comparing treatment options within IPR should be set up. These designs are ethically, clinically and economically acceptable. Evaluation of IPR will however remain difficult,
because the variety in additional problems experienced by the patients hampers the development and replicable description of the intervention, which is required for performing a sound evaluation of a complex intervention [23].

One of the major reasons for randomising patients is to prevent bias. All patients who started treatment in two years were included, so it is not likely that the positive results found in this study are influenced by recruitment bias. Only patients with insufficient knowledge of the Dutch language (n=4) and patients with a primary disease other than asthma or COPD (n=4) were excluded. Furthermore, outcome assessment was independent from the treatment given and not performed by the therapists. These arguments, together with the missing data analysis and imputation, suggest that the internal validity of this outcome study is satisfactory. However, the results of this study may be not generalizable to other inpatient pulmonary rehabilitation programs because of allocation bias: a substantial number of patients are specifically referred to Heideheuvel, and not to other asthmacentres.

Because the current study had no control group, alternative methods to assess the statistical probability and clinical relevance of the observed changes were applied. The first method is to control for multiple testing by decreasing the level for accepting significance to a very stringent level. A p-value of 0.001 provides reasonable evidence against the null-hypothesis [29]. Hospitalization, use of oral corticosteroids, self-assessed health status, most QoLRIQ-domains, emotional reaction and in patients with COPD also emotional well-being and experienced invalidity, improved with p-values of 0.001 or lower. A second method is to check if the lower limit of the confidence interval of change is above the minimal important difference. This was true for several QoLRIQ-domains. This shows that there is a clinically relevant improvement in almost all patients.

The improvements found in this observational outcome study of IPR are moderate to large, which is in the same range as other studies on inpatient or outpatient pulmonary rehabilitation. Only hospitalizations showed a much larger decrease than in other studies, which is partly caused by the large number of pre-IPR hospitalization, giving much room for improvement. Some authors have suggested that observational studies consistently give more positive results than randomised controlled trials [30;31], but this view is recently contradicted by others [32-34].
8.8 Further research

There remain a number of questions on both pulmonary rehabilitation and the assessment of outcome in pulmonary rehabilitation, such as the cost-effectiveness and the contribution of different components of comprehensive programs to the total outcome [35]. Recommendations for research on psychosocial interventions are given by Fishman [35]: what are the optimal interventions and to which patients should they be given? Information is needed about training of coping skills, social support, motivation, vocational rehabilitation, proper timing of interventions, and the influence of depression on rehabilitation outcomes. Gosselink and coworkers formulated several questions on exercise training in patients with COPD including optimal intensities and modalities, which setting and how to maintain training effects [36].

It is still unclear which test is the best for assessing change in exercise capacity. Solway [37] suggested the six minute walking test because this test is the most extensively described submaximal exercise test for patients with COPD. However, the endurance shuttle walk test is very responsive to change in endurance capacity [38] and eliminates some of the standardisation and motivation problems of the six minute walking test. Assessment of exercise capacity may also be improved by using predicted values for walking distance [39;40].

Selection of outcome measures should reflect the content of the treatment programme: for highly standardised home-based or outpatient pulmonary rehabilitation programmes, it will be sufficient to use a standardized, minimal set of outcome measures including a disease-specific quality of life questionnaire and a walking test; comprehensive inpatient and outpatient pulmonary rehabilitation programmes require an extended set of outcome measures for use in diagnosis and reporting back to the referring pulmonologist. Scientific studies in comprehensive inpatient and outpatient pulmonary rehabilitation programmes require an extensive set of outcome measures covering all areas in which change due to treatment is expected. However, evaluating comprehensive pulmonary rehabilitation programmes is hindered by the lack of validated and responsive outcome measures to assess specific interventions in non-medical areas: physiotherapy, nursing, vocational and recreational therapy, and psychosocial counselling.
8.9 Reference List

20. Wright JG. Evaluating the outcome of treatment. Shouldn't we be asking patients if they are better? J Clin Epidemiol 2000; 53:549-553.