RR 96-03

THE STUDY OF APTITUDE TREATMENT INTERACTION BY NONLINEAR METHODS: EVALUATION OF A PSYCHOSOCIAL TREATMENT FOR CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Elise M.L. Dusseldorp

Department of Data Theory

Leiden University

Leiden, The Netherlands

Foreword

This research report is a slightly different version of my "master" thesis in Psychometrics and Research Methodology at Leiden University. The thesis has been written under supervision of Dr. J. Meulman, Department of Data Theory, Prof. Dr. S. Maes, Department of Health psychology and Prof. Dr. L. van der Kamp, Department of Psychometrics and Research Methodology. All these departments are at the Faculty of Social Sciences, Leiden University, The Netherlands.

I wish to thank Stan Maes who suggested the idea of this study and Jacqueline Meulman for her expert methodological advice and support. In addition, thanks are expressed to Leo van der Kamp for referring to several relevant articles. Special thanks are extended to Jacqueline Meulman, Winnie Gebhardt and Jacqueline Dicke for their assistance with the formulation and the American-English language. Finally, I thank Jacques Commandeur for his help with the production of this research report.

Abstract

The differential effectiveness of a psychosocial treatment for chronic bronchitis and emphysema patients (n = 46) on quality of life was examined. For this purpose a special strategy of analysis was developed, regarded as an extension of the Aptitude Treatment Interaction (ATI) paradigm. The assumptions of classical ATI analysis: (1) a linear relationship between aptitude and outcome measure and (2) a numerical measurement level, were released. Furthermore, the outcome measures (i.e., 11 subscales of quality of life, reflecting 5 domains of functioning) were treated multivariately in stead of univariately. Three methods were used to test the differential effectiveness: ANOVA, MURALS and a permutation test. The permutation test was considered to be most appropriate, due to the small sample size and the skewness of the distribution of several aptitude variables.

The hypothesis that the treatment would be more successful in terms of quality of life for patients, who at the pre-test showed a more severe medical condition, a lower adherence to medical advice and a lower quality of life level could only partially be confirmed. With respect to emotional functioning, the treatment was indeed significantly more beneficial for patients who reported a higher seriousness of shortness of breath at the pre-test. On the other hand, patients who had a less severe objective medical condition (determined by duration of illness, walking distance in 12 minutes and forced expiratory volume in 1 second) profited more from the treatment; this also applies to patients who reported a higher quality of life at the pre-test. With respect to practical-social functioning the treatment was more beneficial for patients who reported a medium seriousness of shortness of breath and patients who had a less to medium severe objective medical condition.

Introduction

In most research areas of psychology, the effectiveness of a treatment is determined by inspecting the differences in average performance between subjects[1] who have received the treatment (treatment group) and those who have not (control group). A treatment is labeled as 'effective' when the difference in mean scores between treatment and control group on one or more dependent variable(s) of interest is statistically significant. This is usually determined by an analysis of variance ((M)ANOVA). With this type of analysis, individual differences within the groups are considered as 'error variance'. In this paper, analysis *between* treatment and control group will be combined with analysis of individual differences *within* the groups. This combination of analyses may not only be performed to control for individual differences (as in the widespread-used analysis of covariance: ANCOVA), but also to inspect the effectiveness of a treatment with respect to one or more individual characteristics (as in analysis of Aptitude Treatment Interaction). In Aptitude Treatment Interaction (ATI) analysis, individual differences are not considered as 'error variance', but as important sources of possible differential effectiveness[2] of treatments. An ATI approach of evaluation research makes it possible to abandon the concept of the average performer and to challenge the idea of a single best universal treatment.

Cronbach (1957) may be considered as the founder of the ATI concept. He criticized, in his historical article, the gap between "the two disciplines of scientific psychology": the experimental and the correlational tradition. The experimental approach (mostly prevailing in the area of learning or perception psychology) emphasized the differences between groups of subjects, caused by the experimental conditions: *"his goal* (i.e., the experimental psychologist) *is to control behavior, and variation within treatments is proof that he has not succeeded. (...)"* (p. 674). The correlational approach (mostly prevailing in the area of personality-, social-, and child psychology) was directed at the opposite: *" The correlational psychologist is in love with just those variables the experimenter left home to forget. He regards individual and group variations as important effects of biological and social causes. (...) His goal is to predict variation within a treatment"* (p. 674). Cronbach emphasized in particular the importance for applied psychology to combine the two disciplines: *"The greatest social benefit will come from applied psychology if we can find for each individual the treatment to which he can most easily adapt. This calls for the joint application of*

[1]The term 'subjects' refers to 'persons'.
[2]Because this effectiveness differentiates between people, it shall be referred to as *differential effectiveness* (Shoman-Salomon & Hannah, 1991). In the literature several other terms or descriptions refer to the same phenomenon: differential treatment responsiveness (Dance & Neufield, 1988), differential benefit (Stinard & Dolphin, 1981), "the degree to which treatments have different effects as a function of person characteristics" (Snow, 1991).

experimental and correlational methods" (p. 679). This "joint application" should not only imply the examination of both the variances among individuals (or animals) and among treatments, but also of the neglected *interactions* between individual characteristics and treatment variables. For this reason, research on ATI was initiated.

Before describing the basic procedure of ATI analysis, the three terms of which the concept is composed are defined. According to Snow (1991) the first term *aptitude* (sometimes called *attribute*) refers *"to any measurable person characteristic hypothesized to be propaedeutic to successful goal achievement in the treatment studied; propaedeutic means needed as preparation for response to treatment"* (p. 205). In psychological research, aptitude is usually related to psychological tests. For example, intelligence tests are considered as tests of a general cognitive aptitude including the ability to benefit from schooling (Gleitman, 1991). An ATI approach does not assume a trait model in which aptitudes are presumed both stable and continuous (Snow, 1991). The second term *treatment* has a broad meaning of any manipulable situation variable. The categories of this variable are usually fixed. The treatment variable refers to two or more different kinds of treatment (also including the comparison 'treatment' with 'no-treatment') or to one treatment over two or more trials. The third term *interaction* is defined, in the context of ATI research, as *"the degree to which results for two or more treatments differ for persons who also differ on one or more aptitude variables"* (Snow, 1991, p. 206). In other words, ATI is said to be present when for some group of subjects an aptitude variable shows a different relation to an outcome variable in one treatment than it does in another treatment.

A basic ATI design involves two experimental groups (in the present study: a treatment condition (Treatment) and a control condition (Control)), one outcome variable (O) and one aptitude variable (A). In an ATI analysis, the regressions of O on A obtained under Treatment and Control are compared. The main test concerns the question whether there is a statistically significant difference between the regression coefficient (b_t) obtained under Treatment and the regression coefficient (b_c) obtained under Control. As Pedhazur and Pedhazur-Schmelkin (1991) pointed out: *"A conclusion that there is no significant difference between the b's is tantamount to a statement that there is no interaction between the treatments and the attribute"* (p. 546). The classical technical procedure to test this difference will be described in the section 'Statistical procedure and analysis' of this paper. Graphically, an interaction means that the two regression lines are not parallel. If the regression lines cross, then the interaction is called a disordinal interaction (Figure 1, left). This interaction signifies that the rank order of the group effects (Treatment and Control) changes along the continuum of the aptitude. This rank order does not change when an ordinal interaction is present (Figure 1, right).

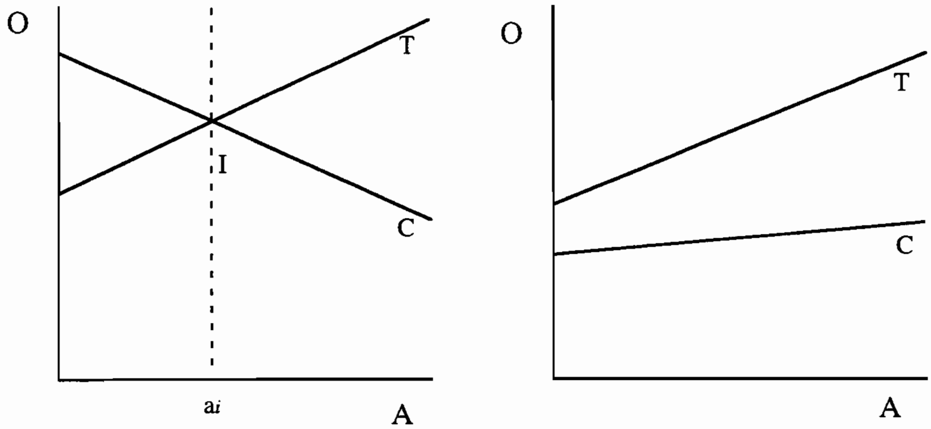

Figure 1. Left: A disordinal interaction. Right: An ordinal interaction. 'O', outcome variable; 'A', aptitude variable; 'I', intersectionpoint; 'T', treatment condition; 'C', control condition.

A disordinal interaction implies that the treatment condition is better for one group of subjects (in Figure 1, left: subjects with scores above a_i) and the control condition is better for another group (in Figure 1, left: subjects with scores below a_i). This distinction cannot be made when an ordinal interaction is present: then, one condition is better than the other, independently from the scores on A (in Figure 1, right: all subjects in the treatment condition score better on the outcome variable). If a *significant* interaction is found, several methods exist (Rogosa, 1980) to determine *the regions of significance;* the difference in performance between the experimental conditions only has statistical importance in specific regions of the continuum of the aptitude. The most commonly used method to determine simultaneous regions of significance is the Johnson-Neyman technique, extended by Pothoff (1964). The implication of a *non-significant* interaction is that the regression coefficients are homogeneous. This is one of the conditions for performing an analysis of covariance. The discovery of heterogeneity of regression slopes often caused investigators trouble *"because it casts doubt on the suitability of ordinary ANCOVA for testing the main effect. But heterogeneity is the signal that an ATI is present, and the interaction should be interpreted"* (Cronbach & Snow, 1977, p. 62).

In the literature most applications of an ATI approach to evaluation research are found in educational psychology (e.g., Cronbach & Snow, 1977; Stinard & Dolphin, 1981) and a few in clinical psychology (Dance & Neufield, 1988). Most applications in clinical psychology use a 2 x 2 analysis of variance (ANOVA) to detect possible interactions between aptitude (divided in 'high' and 'low') and treatment[3]. When

[3]Cronbach and Snow (1977) pointed out, that this procedure had low power for detecting such interactions, because a great deal of variance remained within the aptitude groups 'high' and 'low'.

classical ATI analysis (the regression approach described above) was applied, multiple aptitudes or multiple outcome measures were analyzed separately (e.g., Stinard & Dolphin, 1981). Furthermore, both a linear relationship was assumed between each aptitude and outcome variable, as well as a numerical level of measurement. This paper will focus especially on possibilities of *multivariate* analysis of ATI and explores both nonlinear relationships between aptitudes and outcome variables as well as different levels of measurement (categorical and numerical). As Cronbach and Snow (1977) recommended, plural aptitude variables should be treated simultaneously and they proposed: *"One may use multiple regression or may form a single composite, or may form several near-orthogonal composites"* (p. 35). In connection with this argument, Snow (1991) stated: *"Separate ATI findings should not be interpreted independently when outcome variables are correlated. Multiple outcome analysis can enrich the understanding of ATI substantially if it is designed to preserve the meaning of each measure"*.

No applications of ATI analysis were found in the literature of research in the field of health psychology. Before justifying an ATI approach for evaluating a psycho-social treatment for patients with Chronic Obstructive Pulmonary Disease (COPD), an introduction will be given of COPD and pulmonary rehabilitation. COPD is a generic term for patients with chronic bronchitis or emphysema. The common feature of these lung diseases is fixed expiratory airflow obstruction due to airway narrowing. The cause of the airflow obstruction is different for the two patient groups. Patients with chronic bronchitis suffer from chronic or recurrent excess mucus secretion into the bronchial tree. Emphysema patients are *"characterized by abnormal permanent enlargement of the airspace's distal to the alveoli (terminal bronchioles), accompanied by destruction of the walls of the alveoli (..)"* (American Thoracic Society, 1987 in: Van den Broek, 1995, p. 3-4). The major distinction with asthmatic patients is the ir- or reversibility of the airway obstruction. In asthmatic patients, the airway obstruction is reversible with medication. This is in contrast with COPD patients, whose obstruction is only partially reversible through the use of medication. Despite specific underlying differences, asthmatic and COPD patients share common symptoms, such as intermittent dyspnea, coughing, and wheezing. Because of these common characteristics, the term Chronic Non Specific Lung Disease (CNSLD) is often used to refer to both asthma and COPD. There are several types of treatment for COPD, such as interventions to stop smoking, drug therapy, oxygen therapy, chest physical therapy, exercise training, biofeedback and ventilatory muscle training. Since there are many various aspects of care, attempts have been made to identify which aspects of care are effective and to develop a pulmonary rehabilitation program. It is beyond the scope of this paper, however, to describe each of these treatments, their corresponding effect-studies and the development of pulmonary rehabilitation (see Hodgkin, Zorn &

Conners, 1984). An important development is the increasing role of psychology in pulmonary rehabilitation for COPD patients, both as part of the rehabilitation program (e.g., cognitive-behavioral therapy), as well as goal of rehabilitation (e.g., improvement of quality of life). The effectiveness of psychosocial treatments for COPD patients is, however, hardly investigated (Van den Broek, 1995). Therefore, a study to investigate the effectiveness of a patient education program for COPD patients was initiated by the Department of Health Psychology of Leiden University. It should be noted that this study was not designed a priori to investigate ATI, but to investigate the overall effect of the treatment. The primary analyses in this study showed a short term treatment effect for emotional and social functioning. No effects were found for practical and physical functioning, use of medical resources and adherence to medical advice. This paper involves the secondary analyses of the data from this study from an ATI perspective.

An ATI approach for evaluating health psychology research in general and the COPD-program in particular can be justified as follows. First, the idea of there being only one appropriate treatment for all patients may not be realistic. Dance and Neufield (1988) regarded ATI research as *"attempts to dispel the 'patient uniformity' myth"*. With the myth of 'patient uniformity' they meant the tacit assumption that patients with a given disorder constitute a homogeneous group. According to Maes and Van Veldhoven (1990): *"The right question is most probably not 'Do people profit from an intervention?', but 'Which people profit from which type of intervention?' Thus the conclusion of effect studies should perhaps not be whether all patients profit in the same way from a specific intervention, but which patients profit (most) in which respect from which intervention"* (p. 43-44). Second, the results of an ATI analysis could lead to guidelines regarding the development of screening measures for placing patients in the right type of care. Current trends, concerning the provision of psychosocial care for chronically ill patients, focus on those patients who are especially in need of treatment in order to increase efficiency (i.e., to lower the costs and to increase the quality of care). Therefore, it has become increasingly important that screening occurs. Third, when no overall effect or a small overall effect of a treatment is found with ANCOVA, ATI analysis can be applied to investigate possible differential effects.

The main objective of the present study is to investigate the differential effectiveness of a psychosocial treatment for COPD patients on quality of life, using the ATI-approach. The outcome measure[4] 'quality of life' (including emotional, cognitive, physical, practical and social functioning) is chosen, because of its treatment-specific

[4]The outcome variable was measured at a pre-test and five post-tests. Only data of the pre-test and the second post-test were analyzed in the present study to investigate the course of quality of life. The choice for these specific measurement points was taken on account of (1) comprehensibility (2) equality of measures and (3) a high attrition rate.

and multi-dimensional character. Aptitude measures to identify subgroups are: demographic variables (age, gender), measures of objective severity of disease (duration of illness, exercise test, lung function measure), measures of subjective severity of disease (self-reported seriousness), measures of adherence to medical advice (practice of breathing and relaxation exercises). The hypotheses, derived from a therapeutically standpoint, states that the COPD-program will be more successful in terms of quality of life for patients, who at the pre-test show 1) a more severe medical condition, 2) a lower adherence to medical advice and 3) a lower quality of life.

More specifically, the objectives of the present study are: 1) to clarify the relationships between the various aspects of the outcome measure QOL; 2) to explore relationships between the aptitude variables and QOL, taking into account possible nonlinear relationships and a mixture of measurement levels; 3) to determine exploratively which aptitude variables interact with the treatment conditions; 4) to test the ATI's and the hypothesis formulated above.

Methods

The psychosocial treatment

The psychosocial treatment for COPD patients consists of an extensive educational program (called "COPD, short of breath, but independent"). The program was offered to groups of patients and their partners at the Pulmonary-rehabilitation center Hornerheide in Horn, the Netherlands. The program encompassed eight two-hour sessions within a period of four weeks (2 sessions a week). The topics of the sessions were respectively: patho-physiological aspects, use of medicines, inhalation techniques, nutrition, physiotherapy, daily functioning, social functioning and stress. During the first hour of each session, information was given regarding one of these topics by various specialists who worked at the rehabilitation center. The second hour consisted of a form of cognitive-behavioral therapy with respect to the same topic as the first hour. Problem-solving skills were learned, in order to change inadequate coping behavior. Furthermore, beliefs and thoughts of the patients were discussed through a form of cognitive restructuring: self-talk. This second hour was led by a health psychologist. The educational program was complementary to the standard pulmonary rehabilitation program in Hornerheide, which consisted of physical training, physiotherapeutic interventions, individual psychological interventions and nutritional interventions.

Sample and design

Ninety-six patients were randomly assigned to a treatment group (Treatment) of 50 patients (who received the total educational program) and a control group (Control) of 46 patients (who received only the information part of the educational program[5]). Two measurement points were considered: a pre-test, one week before the program started and a post-test three months after discharge from the rehabilitation center. The drop-out was considerable: of the initial 50 (in Treatment) and 46 patients (in Control) from the pre-test, 23 (in Treatment) and 23 (in Control) patients continued to participate in the study at the post-test. The main reasons for drop-out were: premature discharge or transfer to a General Hospital or nursing home (20), lack of interest (7), death (7), no time to participate (7) and illness (4) (Van den Broek, 1995, p. 76). The remaining sample was characterized as follows: 72% (33)[6] was male, 28 % (13) was female; the mean age was 62, ranging from 42 through 70; 56% (26) had a low level of education, 33% (15) a medium level of education, 9% (4) had a high level of education and 2% (1) had a missing value for education.

Instruments of measurement

The General Questionnaire for CNSLD patients (GQCP) was used to assess several aptitude variables and the multi-dimensional outcome variable Quality of Life (QOL). The GQCP is a structured interview method (for the content of the GQCP, see Van den Broek, 1995, p. 141-154). The partners were asked to be present during the interview in order to enhance the reliability of the answers concerning medical adherence. The following aptitude variables were derived from this interview: age, gender, duration of illness (DURILL, derived from item 1), subjective seriousness of illness (SSER, item 2) and adherence to medical advice (ADHER, item 5 and 6). DURILL was computed by subtracting the score on item 1 (the age of onset of illness) from the age of the respondent. ADHER was derived from items 5 and 6. First, the four answer categories of item 5 (practice of breathing exercises) and 6 (practice of relaxation exercises) were pooled into two categories: 0 (no or hardly never) and 1 (sometimes or regularly). Second, ADHER was computed on the basis of three behavioral categories: 1 (no practice at all), 2 (practice of breathing or relaxation exercises[7]) and 3 (practice of both breathing and relaxation exercises).

Special attention was given to the construction of the subscales of the concept QOL, including the domains emotional, cognitive, physical, practical and social functioning. The domains emotional and social functioning were newly constructed by

[5]In other words, the control group only received the first hour of the educational program and the treatment group received both hours, including the cognitive-behavioral therapy.

[6]The frequencies are put in parentheses.

[7]The subdivision of category 2 of ADHER was neglected, because just two subjects performed only relaxation exercises.

Schlösser, Maes and Van den Broek (1990). The cognitive and physical functioning domains were derived from the Bronchitis Emphysema Symptom Checklist (Kinsman, Fernandez, Schocket, Dirks & Covino, 1983) and the practical functioning domain was adopted from Liem and Maillé (1987). Principal Component Analysis (PCA and PRINCALS)[8] and Reliability Analyses were performed for each of the domains respectively to construct subscales. These analyses were applied to the sample ($n = 96$) of the pre-test.

Emotional functioning consisted of 14 items (9aa through 9an) measured on a 5-points Likert scale, ranging from 'never' to 'always'. A three-component solution[9] was found, which explained 67.4% of the total variance. The rotated components (Varimax) are interpreted as 'depression' (consisting of feeling low, sad, bad mood[10], feeling down and depressive), 'anxiety ' (consisting of afraid, tense, worried, agitated and panicky) and 'anger' (consisting of angry, irritated, bad mood[10], aggrieved and furious). From these components three subscales are computed in two ways: by simply summing the corresponding items or by first multiplying the items with their weights[11] and then summing the weighted item scores. Although the latter procedure can be regarded as statistically more accurate (when the weights are different), the correlations between the two differently computed scales were equal to 1.0 (depression: .9991, anxiety: .9959 and anger: .9929). Therefore, the simple summation procedure was chosen to compute the scale-scores. Table 1 gives the intercorrelations of the resulting scales (together with the scales from the other domains, to be discussed below) and Cronbach's alpha for each of the separate scales.

Physical functioning (16 items of 9b) consisted of three scales[12] of Kinsman (1983): fatigue (5 items), congestion (6 items), dyspnea (5 items). The items were also measured at a 5-points Likert scale, from 'never' to 'always'. A three-component

---

[8]PRINCALS is a special generalization of PCA. PCA assumes that the items are measured at a numerical level. Because our data consisted of items measured at an ordinal level (Likert-scale), analyses with PRINCALS (see for an extensive description the section 'Statistical procedure and analysis') were performed too. Yet, in general, the results of the PRINCALS analyses corresponded highly with the unrotated component solutions of the PCA.

[9]The criterion for the number of components was an eigenvalue higher than one.

[10]This item loaded on the scale Depression as well as Anger. Because the Cronbach's alpha of both scales diminished when the item was deleted and because of the fact that the item is theoretically related to both concepts, it was decided to include 'bad mood' in the computation of both scales.

[11] The weights are equal to the component loadings divided by the square root of the eigenvalue of a principal component. They are similar to the values of the eigenvector belonging to the eigenvalue. When the rotated solution is used, post-multiplication with the rotation matrix is needed. In matrix algebra: $\mathbf{W}=\mathbf{A}\Lambda^{-1}\mathbf{R}$, where $\mathbf{W}$ is the matrix of weights, $\mathbf{A}$ is the matrix of unrotated component loadings, Λ is a diagonal matrix with the square root of the eigenvalue on the diagonal and $\mathbf{R}$ is the rotation matrix. $\mathbf{W}$ has the same order as $\mathbf{A}$: number of variables x number of components. Λ has the same order as $\mathbf{R}$: number of components x number of components.

[12] Initially a fourth scale was added to the physical domain: 'poor memory'. Then, this scale was considered conceptually as psychological functioning together with the scales of emotional functioning. Finally, because 'poor memory' strongly dominated one dimension (in a two dimensional PRINCALS solution) it was decided to treat this scale separately as 'cognitive functioning'.

solution was achieved, accounting for 63.4% of the total variance. However, the Varimax rotated components did not reflect the three scales of Kinsman (1983). Two items of congestion (chest filled up and chest tightness) loaded on dyspnea. It was decided to delete these items and perform a second PCA. The following three-component solution accounted for 68% of the total variance. The Varimax rotated components were easy to interpret: 'fatigue' (consisting of fatigued, weak, exhausted, tired, no energy), 'congestion' (consisting of coughing, mucous congestion, chest congestion and wheezing) and 'dyspnea' (consisting of feel like I need air, hard to breathe, shallow breathing, short of breath, gasping for breath). Subsequently, the two methods of adding described above were applied to compute the scales. The simple summation method was chosen once again due to very high correlations (fatigue: .9989, congestion: .9961 and dyspnea: .9786[13]).

A single component solution was found for *cognitive functioning* (5 items of 9b), including the items 'forgetful', 'forget recent things', 'poor memory' and 'difficulty remembering'. The item 'get confused' was deleted due to a low proportion of variance (.26) accounted for (VAF) by the component. The component accounted for 65.4% of the total variance. The simple summation method was used to compute the subscale, because the weights of the remaining items were very similar.

Practical functioning was indicated by 19 items (11a through 11k, 13a through 13h) measured on a 4-point scale, ranging from 'no, I cannot' to 'yes, effortless'. Liem and Maillé (1987) subdivided these items into two subscales: general daily functioning (items 11a through 11k) and special daily functioning (items 13a through 13h). This was checked for our sample. A two dimensional PRINCALS analysis showed two items did not fit[14] well: 'cut your meat' (11k) and 'drive a car' (13g). The item 'drive a car' appeared to have 40 missing cases (probably because these respondents had no license to drive a car). It was decided to delete these two items from further analyses. The second dimension distinguished between general and special functioning, with two items 'climb the stairs without rest' (11i) and 'walk 20 minutes without rest' (11j) clustering with special functioning, instead of general functioning. A forced two-component PCA[15] (without items 11k and 13g) confirmed this latter finding. It was decided to construct the scale 'special functioning' *with items 11i and 11j,* because it was also theoretically defensible. The simple summation method was used to compute the scales, because the weights were very similar.

---

[13]Although the weights were ranging from .46 (short of breath) to .09 (hard to breathe) this correlation was still high.

[14]The VAF of a variable is the reported *single fit for each variable* (see 'Statistical procedure and analysis' for description) The value of the single fit of 'cut your meat' was .087 and of 'drive your car' was .115. The values of the remaining items were all above .55 (except for '20 minutes walk without rest': .419).

[15]This solution accounted for 62.2% of the total variance.

The last domain *social functioning* was operationalized with the question 'to what extent does bronchitis/emphysema influence the following aspects of your life?' It consisted of 15 items (10a through 10o), measured at a 4 point Likert scale, ranging from 'not' to 'a lot'. Social functioning aimed to be a measure of the extent to which a patient could not fulfill his/her social role anymore. The first PRINCALS analysis showed there was little coherence between the items, probably because the items reflected a wide range of domains. On deeper reflection the items 'income', 'perspectives', 'health', 'holidays', 'neighborhood', 'social esteem' and 'different' were not considered as indicators of a particular 'social role'. Therefore a second PRINCALS analysis was performed with the remaining 8 items. This solution showed more coherence. On the second dimension the items were divided into two clusters, interpretable as 'general social functioning' (including 'family life', 'contact with family', 'contact with friends'), and as 'special social functioning' (including 'housekeeping', 'sports, 'leisure activities', 'sexuality' and 'partner relationship')[16]. The two methods of summing the items to compute the scales were applied, because the weights showed some variation. The correlations, however, between the two differently computed scales were once again close to 1.0 (general social functioning: .9980, special social functioning: .9975). Therefore, the simple summation was applied.

Table 1. Intercorrelations and Reliabilities of the scales of QOL (N = 96). Reliabilities (Cronbach's alpha) are in boldface type and shown on the diagonal. Correlations between the scales of one domain are underlined. The scales of QOL are as follows: FAT, fatigue; CON, congestion; DYS, dyspnea; DEP, depression; ANX, anxiety; ANG, anger; MEM, poor memory; GPF, general practical functioning; SPF, special practical functioning; GSF, general social functioning; SSF, special social functioning.

	FAT	CON	DYS	DEP	ANX	ANG	MEM	GPF	SPF	GSF	SSF
FAT	**0.86**										
CON	0.34	**0.86**									
DYS	0.66	0.54	**0.85**								
DEP	0.47	0.14	0.32	**0.89**							
ANX	0.56	0.13	0.43	0.59	**0.85**						
ANG	0.42	0.13	0.32	0.63	0.53	**0.82**					
MEM	0.53	0.22	0.50	0.33	0.33	0.22	**0.89**				
GPF	-0.44	-0.18	-0.33	-0.28	-0.21	-0.04	-0.36	**0.92**			
SPF	-0.58	-0.09	-0.43	-0.27	-0.27	-0.14	-0.44	0.65	**0.91**		
GSF	0.18	0.06	0.11	0.33	0.08	0.13	0.06	-0.17	-0.11	**0.76**	
SSF	0.52	0.03	0.31	0.33	0.33	0.28	0.22	-0.37	-0.41	0.45	**0.69**

Table 1 shows that the reliabilities of the final 11 scales were high (i.e., above .80), except for the scales of social functioning. But the reliability of the latter scales

[16]A PCA with the same set of items resulted also in a two component solution, accounting for 55.2 percent of the total variance. The varimax rotated components reflected the same subdivision of items.

was still moderately high: .76 and .69. In general, the correlations between scales of one domain are high (i.e., above .50), except for the correlation between fatigue and congestion, r(FAT,CON) is .34, and between general and special social functioning, r(GSF,SSF) is .45. All correlations are positive, except the correlations with the scales of practical functioning. This is due to the fact that a high score at all the scales means many complaints in the domain of functioning, except for the scales of practical functioning. At the latter scales a low score means many complaints.

The self-reported Symptom Check List (SCL90) was also used to obtain additional information about emotional functioning, due to missing data at the GQCP (see 'Statistical procedure and analysis'). Three scales of the SCL90 were used: the depression-scale, the anxiety-scale and the hostility-scale. The scale-scores were computed, according to the test manual.

In addition to the questionnaires, medical data were obtained from medical files and pulmonary function tests. Two final aptitude variables were derived: Forced Expiratory Volume in 1 second as a percentage of the predicted value (PFEV1) and 12 Minute Walking Distance (WALKD). The first variable was chosen from several lung functioning measures (including blood gas measures), because the PFEV1 is considered as the most reliable medical indicator of severity of disease (Cugell, 1988; Petty, 1988; Kaplan & Atkins, 1988). The latter variable (WALKD) resulted from a test, where the patients walked indoors as far as possible in 12 minutes.

Statistical procedure and analysis

The first phase of the analysis was directed at the clarification of the relationships between the 11 subscales of the outcome concept QOL, measured at two times. Because only 46 subjects were measured at both times, preliminary analyses involved inspection and substitution of missing data (called imputation) at the scales of QOL to keep the sample as large as possible. Two types of missing data had to be dealt with: random and systematic. Random missings originated from a missing score at one or more items of a scale. These missings were substituted by the mean score of the remaining items, belonging to a particular scale. Systematic missings were found at the scales of emotional functioning: 11 subjects had no scores at any of the items of these scales (DEP, ANX and ANG). This was probably caused by a missing page of the distributed questionnaire GQCP. However, scores of all the subjects (46) were available on comparable scales of the SCL90 (SCLDEP, SCLANX, SCLHOS).

Therefore, the missings were substituted with a certain strategy[17] by corresponding scores at the scales of the SCL90. This procedure was justified, because of high correlations between the scales of the two methods: r(DEP,SCLDEP) = .71; r(ANX,SCLANX) = .69; r(ANG,SCLHOS)[18] = .48.

The next aim was the clarification of the relationships between the 11 scales of QOL, measured at two times. Since, in a later phase, we were interested in the course of QOL in different groups of subjects (e.g., treatment and control group, severely ill and not severely ill), a special application of PRINCALS was used. Before describing this special application, adopted from Van der Ham, Meulman, Van Strien and Van Engeland (1995), a general description of PRINCALS will be given, for those who are unfamiliar with this type of analysis. PRINCALS[19] is a special generalization of classical PCA. Both methods find a small number of new variables (called: 'principal components') containing as much information as possible from the larger set of original variables. The scores of the subjects on the original variables are represented by the principal components, i.e., each component is a vector of subject scores (in PRINCALS called 'object scores' and in PCA[20] called 'component scores'). In PRINCALS, however, the original categories of the variables are replaced by category quantifications, resulting from the optimal transformation of the variable. If the variables are assumed to be 'numerical', then the PRINCALS solution is identical to PCA. However, if the variables are assumed to be nominal or ordinal, PRINCALS optimally transforms the variables; this implies that the average proportion of variance accounted for is as large as possible (given the nominal or ordinal information). The correlations between the quantified variables and the scores for the subjects (the principal components) are called component loadings. When the squared component loadings on each component are summed per variable, a measure of *the VAF per variable* is obtained (in PRINCALS called 'single fit'), indicating how much each variable contributes to the solution (see also GIFI, 1990).

The special application of PRINCALS was as follows: the two original data matrices were put below each other, resulting in one data matrix of 92 rows (the 46 subjects at two times) and 11 columns (the 11 scales of QOL). Now that the data were reshaped in this form, the PRINCALS analysis gave one set of optimal quantifications for the variables and one set of component loadings, but each subject obtained two sets

---

[17]If a respondent had, for example, a score of 40 at SCLDEP and a missing at DEP, then the scores at DEP were noted of all other respondents, having a score of 40 at SCLDEP (and no missing at DEP). The mean of these noted scores served as the substitute of the missing value at DEP.

[18]Initially, this correlation was lower: .36. Finally, because the hostility-scale of SCL90 consisted of items measuring aggressiveness as well as anger, only the sumscore of the two items measuring anger (items 11 and 24) was used to compute the correlation and to substitute the missings.

[19]PRINCALS is an acronym for PRINcipal Components analysis by means of Alternating Least Squares, and is included in the SPSS (1990) package Categories.

[20]When SPSS is used to perform a PCA the subject scores are called 'factor scores'.

of scores, one for each time, displaying the development over time (see Van der Ham et al., 1995, p. 8). The configuration of the variables[21] and the subject scores of the final PRINCALS solution served as the basis for further analysis.

The second phase was aimed to explore the relationships between the aptitude variables[22] and the outcome measure QOL. Two strategies were applied: a *centroid model approach* and *a vector model approach.* In the centroid model approach, the categories of an aptitude variable identify different groups of subjects, and the subject scores from the first phase of analysis are used to compute centroids. The centroid is the center of gravity of all subjects belonging to the same category. Graphically the centroid is a point with the means of the subject scores on each component as coordinates. The centroid model approach was performed for each time separately, in order to gain a clear understanding of the course of QOL in the different groups of subjects. An asset of this approach is that it can reveal a nonlinear relationship between an aptitude variable and the outcome measure QOL. To apply this approach to the continuous aptitude variables AGE, DURILL, PFEV1 and WALKD, these variables had to be discreticized. In order to preserve the original shape of the distribution of each variable as much as possible, an optimal discretization procedure was applied, recommended by Van Rijckevorsel, Bettonvil and De Leeuw (1985). For the discretization of a standard normal-distributed variable into 5 categories, the critical z-values are: -1.2444, -0.3823, 0.3823 and 1.2444 (see Van Rijckevorsel et al., p. 14). After standardization of a normal-distributed variable, the z-values were transformed into '1' ($z \leq -1.2444$), '2' ($-1.2444 < z \leq -0.3823$), '3' ($-0.3823 < z \leq 0.3823$), '4' ($0.3823 < z \leq 1.2444$) and '5' ($z > 1.2444$). This procedure is similar for variables with a skewed distribution, but the critical values for a positively skewed variable are: -0.1257, 0.5244, 1.0364, 1.6449 and for a negatively skewed variable: -1.6449, -1.0364, -0.5244, 0.1257.

In the vector model approach the aptitude variables are considered as 'external' variables that can be projected in the configuration of the QOL-scales (resulting from the first phase of analysis). This approach is included in the statistical technique Preference mapping (PREFMAP). The general purpose of PREFMAP is *"to relate preference information on a number of objects to a pre-existing spatial configuration of points, under the assumption of a simple class of models. Here "preference" is used as the generic name of any type of observations that indicate a conditional dominance relation among objects: thus "properties", 'attributes" (....)"* (Meulman, Heiser & Carroll, 1986, p. 2). The underlying models of PREFMAP are hierarchically ordered from simple to more complex. In the present study, the most simple model was assumed: the vector model. In this model the variable is represented in the *p*-dimensional

[21]In the output of PRINCALS called 'the plot of component loadings'.
[22]The aptitude measures were all derived from the pre-test.

configuration by a vector. The direction of the vector indicates increasing "preference". Because we were interested in the relationship between an aptitude variable and both QOL at the pre-test as well as QOL at the post-test, separate vectors were plotted for each time. The following strategy was used to compute the point of departure, the direction and the length of the vector[23]. The *point of departure* of the vector was not the origin of the configuration of QOL scales. Because only the subject scores at one time were selected ($n_1 = n_2 = 46$)[24], the mean score of these subjects on each component had to be computed. These means were the coordinates of the point of departure. The *slope* of the vector $\mathbf{b}_1$ (time 1) was computed by the projection of $\mathbf{y}_1 = \{y_i\}$(the standardized scores of the aptitude, with i as index of the subjects, with $i = 1,....,n_1$) on $\mathbf{X}_1 = \{x_{is}\}$(the data matrix of the subject scores, in deviation from column means: the rows i indicate the subjects, here at time 1, and the columns s indicate the components, with $s = 1,....,p$). The corresponding formula for time 1 is

$$\mathbf{b}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\,\mathbf{X}_1'\mathbf{y}_1 \qquad (1)$$

where $\mathbf{b}_1$ is a vector of p x 1. The *length* of the vector is determined by the variance accounted for (VAF)[25]. More precisely, the length equals the square root of the VAF. The VAF is the squared correlation between the original $\mathbf{y}_1$ and the predicted $\mathbf{y}_1$: $\hat{\mathbf{y}}_1$. The $\hat{\mathbf{y}}_1$ was computed as

$$\hat{\mathbf{y}}_1 = \mathbf{X}_1\,\mathbf{b}_1 \qquad (2)$$

where the N x p data matrix $\mathbf{X}_1$ was still assumed to be in deviation of the column means. With this procedure all the aptitude variables could be represented by two vectors (one for each time) in the configuration of the scales of QOL. The length of the vectors should be regarded as measures of 'goodness-of-fit' of the representation: the longer the vector, the higher the correlation with the components of the PRINCALS solution. In other words, the longer the vector, the stronger the relationship between the aptitude variable and the outcome variable QOL at one particular time. It should be noted that in this approach the aptitude variables are assumed to be continuous and to have a linear relationship with the outcome variable.

The third phase was directed at detecting possible interactions between the aptitude variables and the treatment variable. To explore these interactions, the above-described centroid model was applied. But this time the relationship between an aptitude and QOL was investigated *within* groups: Treatment and Control. If this relationship was different for the two groups, it signified an interaction (see the Introduction for a definition of interaction).

First, analyses were performed to determine the relationship between Treatment and Control and the outcome variable QOL. The course of QOL from pre-test to post-

[23]The algorithm was developed in APL. APL is an abbreviation of A Program Language.
[24]Sometimes the actual N was smaller because of missing values on an aptitude variable.
[25]In PRINCALS terminology the VAF of a variable equals the row sums single fit (see also note 13).

test was clarified *within* the groups, by computing the centroids of each group at each time. Then, subjects in Treatment and Control were subdivided into 'relapsers' and 'progressers'. The 'relapsers' were subjects whose QOL at the pre-test was better than at the post-test; the QOL of the 'progressers' was better at the post-test. The centroids of these groups at each time were computed and inspected.

Second, the aptitude variables were used to identify subgroups within Treatment and Control. The same discretization of the continuous variables was used as in the second phase. The variable QOL measured at the pre-test was also treated as 'aptitude'.

The last phase was aimed at testing the statistical significance of the interactions and of the hypotheses, formulated in the introduction. Three methods were applied to test this: 1) an analysis of variance; 2) a generalization of classical ATI analysis; 3) a permutation test. The first method determines the significance of effects by comparing the means of the criterion variable of the aptitude (factor A) and treatment variable (factor B) with the total group mean. In this way both the significance level of the *main* effects of the (discreticized) aptitude variable and of the treatment variable as well as the significance level of the *interaction* effect between aptitude and treatment was computed. The formula to test the interaction effect is:

$$F = MS_{AB} / MS_W \text{ with } df_{numerator} = (a\text{-}1)(b\text{-}1)\ ,\ df_{denominator} = ab(n\text{-}1) \qquad (3)$$

where n is the number of observations for each treatment; a and b indicate the number of categories of factor A and B, respectively. The following formulas are used to compute MS_{AB}:

$$MS_{AB} = SS_{AB} / ((a\text{-}1)(b\text{-}1)) \qquad (4)$$

$$SS_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y})^2 \qquad (5)$$

where $\bar{Y}_{i.}$ indicates the mean for A_i averaged over the levels of B_j (i and j refer to the categories of A and B), $\bar{Y}_{ij}$ indicates the mean of cell i, j, and $\bar{Y}$ indicates the total group mean. With a significant A×B interaction, the difference between the means of for example A_1 and A_2 is conditional on the levels of B (Edwards, 1985, p. 107).

The second method is a generalization of multiple regression analysis. The classical way to test ATI is by means of a stepwise procedure of multiple regression. The dependent variable[26] of the regression analyses is the outcome variable (in this study: a component of QOL) and the independent variables (or 'the predictors') are the aptitude, the treatment and the aptitude-treatment interaction. In classical ATI analysis the aptitude variable is continuous and the treatment variable is categorical (with effect-coding: '1', indicating Treatment, and '-1', indicating Control). The interaction term is computed by multiplying the aptitude and treatment variable. The predictors enter the

[26]In the case of two measurements (before and after treatment) the dependent variable consists preferably of gain scores or adjusted scores (with the pre-test measure as covariate) to take into account the pre-test value of the subjects.

regression equation step by step: in the first step the aptitude variable, then, the treatment variable and in the last step the interaction term. In the second step, a significant F-change value means that the two intercepts[27] are different for the two conditions, indicating a main effect of the treatment variable. In the last step, a significant F-change value means that the two regression coefficients are different for the two conditions, indicating an interaction effect (see Pedhazur & Pedhazur-Schmelkin, 1991, Chapter 21). The significance test for the interaction term can be expressed in formula as

$$F = r^2_{Y(3.12)} / ((1-R^2_{Y.123}) / (n-k-1)) \quad (6)$$

where $k = 3$ is the number of predictors; $r^2_{Y(3.12)}$ is the semipartial correlation coefficient. This coefficient is computed according to the formula:

$$r^2_{Y(3.12)} = R^2_{Y.123} - R^2_{Y.12} \quad (7)$$

where $R^2_{Y.123}$ indicates the multiple correlation coefficient of the last step and $R^2_{Y.12}$ the multiple correlation coefficient of the second step (Edwards, 1985, p.192).

In the present study the measurement level of the (transformed) aptitude variables is categorical. Moreover, the relationship between an aptitude and the dependent variable mostly turned out to be nonlinear (see Results of the second phase of analysis). Therefore a special variety of multiple regression analysis was used: MURALS (see, Van der Kooij & Meulman, 1996). In MURALS the original categories of the variables are replaced by optimal category quantifications (like in PRINCALS). The MURALS algorithm maximizes the multiple correlation coefficient between the criterion variable and a linear combination of the predictor variables, which defines optimality for the transformations of the variables. If the variables are assumed to be 'numerical', then the MURALS solution is identical to standard multiple regression. The following procedure was applied for each aptitude separately: first, a MURALS analysis was performed with a component of QOL as criterion variable and an aptitude variable as predictor. The results of this analysis were optimal category quantifications for the aptitude variable, reflecting a possible nonlinear relationship with the component of QOL. Then, a classical ATI analysis was performed (in SPSS) with as predictors: the optimal transformed aptitude variable, the treatment variable (effect coded) and the interaction term (computed by a multiplication of the transformed aptitude variable with the treatment variable).

The third method is based on the null hypotheses that all variables are independent. *"The distribution of the various statistics that we compute can be approximated by generating a permutation distribution, which mimics independent sampling of subjects"* (Heiser & Meulman, 1994). The observed category scores on the treatment variable and the (discreticized) aptitude variable were independently

[27] i.e, the intercepts of the regression equations, resulting from the regression of O (outcome variable) on A (aptitude) for the treatment and for the control group.

redistributed among the subjects, with the number of permutations set equal to 1000. The marginals of each category are preserved in each permutation. Then, for each of the permuted data sets the following scores were computed: 1) gain scores[28] on the component of QOL for each subgroup of the aptitude variable in Treatment and in Control ; 2) the difference between the treatment-gain score and the control-gain score for each aptitude subgroup. These latter scores are referred to as delta-scores and reflect a possible interaction (see the section Results of the third phase of analysis). The probability that the delta-scores for the permuted data are greater than the original delta-scores is given by the frequency of this event ('permuted' delta-scores ≥ 'original' delta-scores) divided by the number of permutations. This probability gives the significance level of the interaction effect derived from the first two methods. To compute the permutation test results special APL[23] functions were developed.

Finally, it should be noted that subjects who had a missing value on an aptitude variable (i.e., on DURILL, PFEV1 or WALKD) were deleted from analysis in all three methods.

Results

The first phase: relationships between the 11 scales of QOL

Three PRINCALS analyses were performed with the 11 scales of QOL, measured at two times: (1) with systematic[29] missings[30] and a numerical level of measurement; (2) without missings and a numerical level of measurement; (3) without missings and an ordinal level[31] of measurement. The results of (1) and (2) were compared to inspect if the imputation of the systematic missings influenced the structure of QOL substantially. If so, the applied substitution-procedure would not be justified. The results of (2) and (3) were compared to investigate if the assumption of a numerical level of measurement was too restrictive. The total proportion of variance accounted for by each PRINCALS solution was, respectively: (1) 57%, (2) 56%, and (3) 60%. These proportions can be subdivided into the average proportions accounted for by each

---

[28]A gain score is computed by subtracting the subject scores of time 2 from time 1.

[29]The random missings were substituted in all the analyses (see also 'Statistcal procedure').

[30]In PRINCALS with missing data, we do not only have weights for the variables (the normalized component loadings) but also weights for the subjects, which are the number of nonmissing observations. In other words, subjects are not listwise deleted, but only their nonmissing information is taken into account.

[31]When an ordinal level of measurement was assumed, recoding of the original categories appeared necessary. The results of an ordinal PRINCALS without recoding showed that some extreme categories with low marginals dominated the solution: these categories obtained very extreme quantifications. Therefore, some categories were merged. The results of the ordinal PRINCALS with recoded variables are described.

component, which equals the mean squared component loading obtained for the variables. The first principal component accounted for, respectively: (1) 43%, (2) 43%, and (3) 47% of the variance; the second principal component accounted for respectively: (1) 14%, (2) 13%, and (3) 13%. The correlations between the components of the three solutions are displayed in Table 2. This table shows that the correlations between corresponding components of the first two solutions were very high ($r_{c1(1),c1(2)} = 1.00$ and $r_{c2(1),c2(2)} = .88$). This result implies that the substitution of the systematic missings has not influenced the structure of QOL substantially. Therefore, the imputation strategy is considered to be successful. The ordinal solution (3) accounted for only an additional 4 % compared to the numerical solution (2). Furthermore, these solutions showed almost the same subdivision in importance of the components. This implies that the restriction of a numerical measurement level decreased the fit of the solution only slightly. Therefore, the second solution was preferred over the third.

Table 2. Correlations between the components of three PRINCALS solutions: (1) numerical with systematic missings; (2) numerical without missings (3) ordinal without missings. Correlations between corresponding components are shown in boldface type.

	(1)C1	(1)C2	(2)C1	(2)C2	(3)C1	(3)C2
(1) component 1	1.00					
(1) component 2	-.04	1.00				
(2) component 1	**1.00**	-.07	1.00			
(2) component 2	-.02	**.88**	.00	1.00		
(3) component 1	**.98**	-.06	**.99**	-.01	1.00	
(3) component 2	-.04	**.58**	-.00	**.68**	.00	1.00

The results of the second PRINCALS solution will be presented and will serve as the basis for the next phases of analysis. The component loadings of the scales on each component are given in Table 3. Figure 2 shows the graphical representation of the two components[32]. A thing to recall is that a *high* score on the scales of practical functioning means a *low* impairment of functioning. On the other hand, a *high* score on the scales of emotional, cognitive, physical and social functioning means a *high* impairment of functioning (see also Instruments of measurement).

Because all the scales correlate positively with the first component, except for the scales of practical functioning, this component can be interpreted as 'general QOL'. A high score on this component indicates a high impairment of functioning (including all the five domains of functioning) and a low score indicates a low impairment. The

[32]A three-component solution was abandoned, due to (1) the already relatively low contribution of the second component compared to the first component (the contribution of the third component is always even smaller) and (2) the clear interpretation of the two-component solution.

average proportion of variance accounted for by this component is high (43%), so this component is very important.

Table 3. Component loadings of the scales of QOL on each component. The proportion of variance accounted for per variable is given by the fit. The average proportion accounted for by each component is given by the mean square.

Scale	C1	C2	Fit
Fatigue	0.81	-0.01	0.66
Congestion	0.49	0.10	0.25
Dyspnea	0.73	-0.01	0.53
Depression	0.75	0.48	0.80
Anxiety	0.71	0.47	0.72
Anger	0.67	0.41	0.61
Memory	0.48	0.08	0.24
Special practical functioning	-0.69	0.48	0.70
General practical functioning	-0.64	0.50	0.66
General social functioning	0.46	-0.42	0.39
Special social functioning	0.71	-0.28	0.58
Mean square	0.43	0.13	

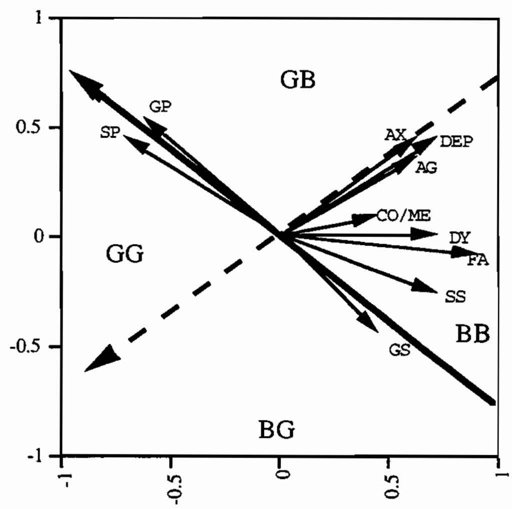

Figure 2. Relationships between the scales of QOL: graphical display of the component loadings. Second principal component (vertical axis) versus first principal component (horizontal axis). The long solid vector indicates practical-social functioning, the dotted vector indicates emotional functioning. The scales of QOL are as follows: FA, fatigue; CO, congestion; DY, dyspnea; DE, depression; AX, anxiety; AG, anger; ME, poor memory; GP, general practical functioning; SP, special practical functioning; GS, general social functioning; SS, special social functioning. The four quadrants are: GG, good practical-social and good emotional; GB, good practical-social and bad emotional; BB, bad practical-social and bad emotional; BG, bad practical-social and good emotional.

The second component differentiates between the scales: emotional and practical are separated from social functioning. Through this separation two new 'components' are formed: practical-social functioning (see Figure 2, the long solid vector) and emotional functioning (see Figure 2, the dotted vector). The two new components divide the space in four quadrants: good practical-social and good emotional functioning (GG), good practical-social and bad emotional functioning (GB), bad practical-social and bad emotional functioning (BB), bad practical-social and good emotional functioning (BG). A last thing to note is that the fit for the scales congestion and memory is low: only 25% of the variance of congestion and 24% of the variance of memory is accounted for.

Summarizing, we can conclude that (1) the main components of QOL (in this sample) are emotional functioning and practical-social functioning and (2) QOL can be considered as a 'summation' of these two components.

The second phase: relationships between aptitudes and QOL

The results of this phase are presented for each aptitude variable separately. The reader is referred to the section 'Graphical representations' for the configurations with the centroids (Figure 4a through 4g) and the configurations with the 'external vectors' (Figure 5a through 5e).

Age and QOL

As mentioned in the section 'Sample and design' the mean age of the subjects is 62, ranging from 42 to 70. There are no missing observations. The variable is negatively skewed. Optimal discretization[33] of the variable results in the following subdivision of the subjects: '1' younger than 50; '2' between 50 and 58; '3' between 59 and 62; '4' older than 62. Centroids are computed for these 4 subgroups at each time. Table 4 gives the results of the course of QOL within the subgroups: for each time the quadrant in which the centroid is located is mentioned.

Table 4 shows that the level of practical-social functioning of the different age-groups does not change from time 1 to time 2. The emotional functioning, however, improves: every subgroup has a good emotional functioning at time 2. Inspection of the configuration of the QOL-scales with the centroids of the subgroups (see 'Graphical representations', Figure 4a) shows that the order of the subgroups at the practical-social

[33]Initially the optimal discretization resulted in 5 categories: '1' younger than 50; '2' between 50 and 56; '3' 57 or 58; '4' between 59 and 62; '5' older than 62. Finally, the categories '2' and '3' were pooled because the marginal frequency of each category was only 3.

component, at the emotional component and at the general QOL component was not monotonically increasing. This indicates a nonlinear relationship between AGE and practical-social functioning and between AGE and emotional functioning.

Table 4. Course of QOL within the subgroups of AGE and GENDER. The quadrants of the centroids are given at each time: the first letter refers to practical-social functioning, the last letter to emotional functioning (B = bad; G = good).

Variable	Category	Time 1	Time 2	N
AGE	Younger than 50	BB	BG	4
	50 through 58	GB	GG	6
	59 through 62	BG	BG	10
	Older than 62	GB	GG	26
GENDER	Female	GB	BG	13
	Male	BB	GG	33

The results of the vector model approach confirm the existence of nonlinear relationships. Only 4% of the variance of AGE at time 1 is accounted for by the two components and 2% of the variance at time 2. A graphical indication of the fit per variable is the length of the vector (see 'Statistical procedure and analysis'). The vectors of AGE are very small (see 'Graphical representations', Figure 5a).

Gender and QOL

Thirteen subjects (28%) are female and 33 (72%) are male. Only the centroid model approach is applied, due to the nominal character of the variable (in fact, with only two categories, centroid and vector models are equivalent). Table 4 shows the course of QOL with respect to the four quadrants within each group. The emotional functioning of both men and women improves over time. The practical-social functioning of men improves also; on the other hand, this functioning decreases for the women. Figure 4b reveals that the improvement of emotional functioning is greater for women than for men: the women evolve from a very high impairment in emotional functioning at time 1 to a medium impairment; the men evolve from a medium impairment to a low impairment of emotional functioning.

Duration of illness and QOL

The average duration of illness at the pre-test measurement is 36 years, ranging from 1 year through 64 years. One observation is missing. The shape of the distribution is positively skewed. Optimal discretization of the variable results in the following categories[34]: '1' = 1 through 15 years; '2' = 16 through 30 years; '3' = 31 through 45

[34]Due to small marginal frequencies the initial categories '31 - 35 years' and '36 - 45 years' were pooled.

years; '4' = longer than 45 years. Table 5 shows the course of QOL within these subgroups, resulting from the centroid model approach. Once again this table shows that practical-social functioning does not change over time and emotional functioning improves over time.

Figure 4c reveals that DURILL has a linear relationship with practical-social functioning: the longer the duration of illness, the worse the practical-social functioning. Orthogonal projection of the centroids on the component of emotional functioning reveals there is little association between DURILL and emotional functioning: the centroids of the different categories at one time are close to each other.

Table 5. Course of QOL within the subgroups of DURILL and PFEV1. The quadrants of the centroids are given at each time: the first letter refers to practical-social functioning, the last letter to emotional functioning (B = bad; G = good).

Variable	Category	Time 1	Time 2	N
DURILL	Shorter than 15	GB	GG	28
	15 through 30	BB	BG	8
	31 through 45	BB	BG	3
	longer than 45	BG	BG	6
PFEV1	lower than 20	BB	BG	2
	21 through 30	GB	BG	11
	31 through 41	GB	GG	19
	42 through 50	BG	BG	5
	higher than 50	GB	GB	5

The vector model approach confirms the findings of the centroid model approach. The variance accounted for (VAF) is 18% at time 1 and 20% at time 2. These percentages indicate a moderately linear relationship (the corresponding correlations are .43 and .45). In Figure 5b, the direction of the vectors is parallel with the practical-social-functioning component and diagonally to the right, indicating a higher impairment of practical-social functioning when duration of illness gets longer.

Forced Expiratory Volume and QOL

The mean Forced Expiratory Volume in one second (expressed as a percentage of the predicted value) is 36.3%, ranging from 2% through 84%. Four observations are missing. The shape approximates a normal distribution, with many observations around the mean and some at the extremes. Optimal discretization results in the categories: '1' lower than 20; '2' 21 through 30; '3' 31 through 41; '4' 42 through 50; '5' higher than 50%. Table 5 shows the emotional functioning of all the groups is good at time 2, except for the highest group: the group with subjects whose illness is less severe. From inspection of Figure 4d, it is concluded that none of the components of QOL has a linear relationship with PFEV1.

The vector model approach confirms this: only 2% of the variance of PFEV1 at time 1 is accounted for by the components and 4% at time 2. This indicates that the relationship is nonlinear (see also Figure 5c).

Walking Distance and QOL

The mean distance walked indoors in 12 minutes is 766.8 meter, ranging from 45 meter to 1185 meter. Six observations are missing. The shape of the distribution is negatively skewed. Optimal discretization results in the categories: '1' less than 500 ; '2' 500 to 600; '3' 600 to 800; '4' 800 or more. Table 6 gives the quadrants of the centroids of the categories at each time.

Table 6. Course of QOL within the subgroups of WALKD, ADHER and SSER. The quadrants of the centroids are given at each time: the first letter refers to practical-social functioning, the last letter to emotional functioning (B = bad; G = good; M = medium, indicating that the centroid lies in the middle of a component; MM = origin of the configuration).

Variable	Category	Time 1	Time 2	N
WALKD	less than 500	BG	BG	4
	500 to 600	BB	BG	6
	600 to 800	BG	GG	10
	800 or more	GB	GG	20
ADHER	no practice	GB	GG	27
	practice of one	BB	BG	11
	practice of both	BB	BG	8
SSER	1: low seriousness	BG	GG	4
	2	BB	MM	4
	3	MM	BG	4
	4	MM	BG	9
	5: high seriousness	GB	GG	25

A first thing to note from this table is that at time 2 all the subgroups of WALKD have a good emotional functioning. The 'less than 500 meter walkers' did not change over time at both dimensions. The emotional functioning of the '500 to 600 meter walkers' and '800 meter or more walkers' improves and the practical-social functioning of the ' 600-800 meter walkers' improves (see also Figure 4e).

In the vector model approach, 15% of the total variance of WALKD at time 1 is accounted for by the components and 21% at time 2. In Figure 5d, the direction of the vectors is almost parallel with the practical-social functioning dimension, and diagonally to the left, which indicates that the higher the walking distance, the higher the practical-social functioning. This means that WALKD has a linear relationship with practical-social functioning and very little association with emotional functioning.

Adherence and QOL

Twenty seven subjects do not perform breathing or relaxation exercises, 11 subjects perform one kind of these exercises and 8 subjects perform both. Table 6 shows that the emotional functioning of all groups improves. The practical-social functioning does not change and only the 'no practicers' have a good practical functioning (see also Figure 4f). The latter finding may imply that those who have a bad practical-social functioning see the need to practice. The vector model approach is not applied, due to the categorical character of this variable.

Subjective seriousness of illness and QOL

The answer categories of Subjective severity of illness (SSER) ranged from 1, indicating a low seriousness of illness, to 10, indicating a high seriousness of illness. The mean seriousness is 6.8. No observations are missing. The shape is negatively skewed. In fact, SSER behaves more or less as a continuous variable, therefore the optimal discretization procedure is applied here as well. The categories are recoded to: '1' = below 4; '2' = 4; '3' = 5; '4' = 6 or 7; '5' = 8 through 10.

A striking result shown in Table 6 is that both the subjects with the lowest self-reported seriousness as well as the subjects with the highest ended up in the quadrant GG. The configuration of the centroids, Figure 4g, shows that SSER has a nonlinear relationship with emotional functioning and no association with practical-social functioning.

Not surprisingly, the result of the vector model approach is less spectacular: only 4% of the variance of SSER at time 1 is accounted for by the components and nothing (0%) at time 2. This result indicates that SSER has no linear relationship with any of the components of QOL (see also Figure 5e).

The third phase: Explorative analysis of ATI

The reader is referred to the section 'Graphical representations' for the corresponding configurations (Figure 6a through 6j).

The course of QOL within treatment and control group

The centroid model approach was applied to represent the treatment and the control group at each time in the configuration of the QOL scales (see Figure 6a). The centroids at time 1 are very similar, indicating that the average QOL at time 1 is equal for the two groups. Both groups show an improvement in emotional functioning, but the improvement of the treatment group is obviously greater. No improvement in practical-social functioning is found for both groups: both groups show an average practical-social functioning at each time.

Subsequently, the groups are subdivided into 'relapsers' and 'progressers' (see 'Statistical procedure and analysis'). The criterion is the first principal component, because this component refers to the 'general QOL' of a subject. If the subject score on this component is lower (or equal[35]) at time 2 than at time 1, the subject is regarded as a 'relapser'. Otherwise, the subject is regarded as a 'progresser'. A summary of the quadrants of the centroids for each subgroup is given in Table 7. The corresponding graphical representation is Figure 6b. The subgroups of the control group show a 'regression to the mean' phenomenon: the 'relapsers' and the 'progressers' both end up in the middle of the configuration. On the other hand, the subgroups of Treatment both end up with a good emotional functioning. In general, the number of 'relapsers' was higher in Control (n = 12) than in Treatment (n = 6).

Table 7. Course of QOL of the 'relapsers' and 'progressers' within treatment and control group. The quadrants of the centroids are given at each time: the first letter refers to practical-social functioning, the last letter to emotional functioning (B = bad; G = good; M = medium, indicating that the centroid lies in the middle of a component).

Condition	subgroup	Time 1	Time 2	N
Treatment	relapsers	GG	MG	6
	progressers	BB	MG	17
Control	relapsers	GG	MM	12
	progressers	BB	MM	11

Aptitude and Treatment Interaction

Table 8 gives an overview of all the identified subgroups per aptitude and the course of QOL within Treatment and Control, resulting from the centroid model approach. For several aptitudes the categories obtained by the optimal discretization procedure could not be maintained, due to the small sample size. Some categories were merged, in order to avoid categories with only one observation. The labels and marginals of the new categories are shown in Table 8 (see also Figure 6c through 6j). Inspection of the quadrants in Table 8 reveals that the level of practical-social functioning *within Treatment* does not change in 14 aptitude-subgroups, decreases in 3 subgroups and increases in 6 subgroups. The level of practical-social functioning *within Control* does not change in 13 aptitude-subgroups, decreases in 8 subgroups and increases in 2 subgroups. With respect to emotional functioning *within Treatment* 18 subgroups show an increase to a good (15) or medium (3) level of functioning, 5 subgroups remain at a good level and no subgroup decreases. *Within Control* 10

[35]Only one subject had an equal score at time 1 and at time 2, on this component. This subject was regarded as 'relapser'.

subgroups show an increase in emotional functioning, 11 subgroups do not change and 2 subgroups show a decrease in functioning. When we focus on the variable QOL (treated as an aptitude[37]), we see that the treatment has been beneficial with respect to emotional functioning for subjects with both a low as well as a high QOL at the pre-test, compared to the control-subgroups. The course of QOL in the latter subgroups confirms the 'regression to the mean' phenomenon (see also Figure 6j). It should be noted that the marginals of these subgroups are almost equal in the treatment and control group, reflecting a good result of the 'randomization' procedure (see 'Sample and design').

Table 8. Course of QOL of subgroups with different aptitudes within treatment and control group. The quadrants of the centroids are given of each time: the first letter refers to practical-social functioning, the last letter to emotional functioning (B = bad; G = good; M = medium, indicating that the centroid lies in the middle of a component).

		Total	Treatment			Control		
Aptitude	subgroup	N	T1	T2	n(T)	T1	T2	n(C)
AGE	younger than 50	4	BB	BG	2	BB	BG	2
	50 through 58	6	GB	GM	2	GG	GG	4
	59 through 62	10	BM	BG	7	BG	BM	3
	older than 62	26	GM	GG	12	GB	MM	14
GENDER	female	13	GB	BG	7	GB	MB	6
	male	33	BM	GG	16	MM	MG	17
DURILL	shorter than 15	28	GB	GG	15	GB	GB	13
	15 through 30	8	MG	GG	3	BB	BG	5
	longer than 31	9	BB	BG	4	BG	BG	5
PFEV1	lower than 31	13	BG	BG	4	GB	MM	9
	31 through 41	19	GB	BG	12	GB	GB	7
	higher than 41	10	GB	GM	4	MG	BG	6
WALKD	less than 600	10	BM	BG	6	BB	BM	4
	600 to 800	10	BG	GG	4	GM	MM	6
	800 or more	20	GB	GG	9	GM	GM	11
ADHER	no practice	27	MB	GG	15	GB	GB	12
	practice of one	11	BB	BG	6	GM	BG	5
	practice of both	8	GB	GM	2	BB	BG	6
SSER[36]	1: low seriousness	12	BG	GG	3	BB	MG	9
	2	9	BB	MG	4	GG	BG	5
	3: high seriousness	25	GB	MG	16	MB	MB	9
QOL[37]	Low	23	BB	BG	12	BB	MM	11
	High	23	GG	GG	11	GG	MM	12

A comparison between the marginals of the other aptitude-subgroups of Treatment and Control, reveals that the marginals are approximately the same, except for PFEV1 and SSER. In Treatment most subjects are in the middle category of

[36]Low seriousness refers to a reported score of lower than 6, medium seriousness refers to a score of 6 or 7, high seriousness refers to a score of 8 or higher.

[37]As mentioned before the level of QOL at the pre-test measurement was also used as 'aptitude'. The 'cut-off-point' for the subdivision of the scores in 'low' and 'high' was the median of the subject scores (selected from time 1) at the first principal component. The value of the median was 0.03.

PFEV1, while for Control the categories are more equally filled. As concerns SSER, in Treatment most subjects have reported a high seriousness of shortness of breath, while Control shows more spread over the categories of SSER.

Because we are also interested in the *magnitude* of change in QOL in the different subgroups, a measure of change has been developed. The centroids of the subgroups are projected on the directions of respectively emotional and practical-social functioning, by multiplying them with relevant direction coefficients. The direction coefficients of emotional functioning consist of the component loadings of anxiety (.71 and .47, see Table 3); the direction coefficients of practical-social functioning consist of the component loadings of general practical functioning (-.64 and .50, see also Table 3). These variables are chosen, because they are situated precisely on the components (see Figure 2). The result of the multiplication is a set of subgroup-scores on emotional functioning and a set of subgroup-scores on practical-social functioning, reflecting the position of each subgroup at each time. Subsequently, a measure of change is computed for emotional functioning by subtracting the subgroup-scores of time 2 from time 1 and for practical-social functioning by subtracting the subgroup-scores of time 1 from time 2, in order to keep the interpretation convenient. In this way a positive value of change indicates an improvement in functioning for both dimensions and a negative value a worsening. Table 9 gives an overview of the changes in each subgroup for each condition.

To gain a clear insight in the difference between the Treatment and the Control delta's (Δ EM and Δ PS) are computed, by subtracting the change of the control-subgroup from the change of the treatment-subgroup. A delta of zero reflects no treatment effect and a positive delta reflects a treatment effect. The mean values of these delta's reflect the magnitude of the overall treatment effect.

The mean values of Δ EM in Table 9 show that the improvement in emotional functioning is substantially larger for Treatment than for Control with respect to all the aptitude variables. The mean values Δ PS reveal that the improvement in practical-social functioning is also larger for Treatment than for Control with respect to all the aptitude variables, but this improvement is less obvious than that for emotional functioning (the values are less high). From this finding we can conclude that the treatment has been especially effective for emotional functioning.

The aim of this phase is to determine exploratively *interactions* between the experimental conditions and the aptitudes. An *interaction* between aptitude and treatment is found, if the delta's of the subgroups of an aptitude are not (approximately) the same. Different delta's reflect that the treatment effect differs for subjects in different aptitude-categories (see also the definition of interaction stated in the Introduction). It should be noted that an overall treatment effect is not *a condition* to find an interaction. Interaction is also possible when there is not an overall treatment effect (see, e.g., Figure 1, left: a disordinal interaction).

Table 9. Change in emotional functioning (EM) and practical-social functioning (PS) for treatment and control group. A positive value of EM or PS indicates progression, a negative value indicates relapse. The mean value refers to the average change for treatment and control group (weighted by the marginals). Delta (Δ) shows the change of the treatment group minus the change of the control group. The delta's of subgroups who profit more from the treatment are shown in *italic* type.

Apti-	sub-	Total	Treatment			Control			Δ	Δ
tude	**group**	**N**	**EM**	**PS**	**n(T)**	**EM**	**PS**	**n(C)**	**EM**	**PS**
A	<50	4	1.16	0.52	2	0.91	0.19	2	0.24	0.33
G	50-58	6	0.89	-0.22	2	-0.23	0.08	4	*1.12*	-0.30
E	59-62	10	0.82	0.44	7	-0.39	0.35	3	*1.21*	0.09
	>62	26	0.55	0.14	12	0.38	-0.05	14	0.16	0.19
	mean		**0.71**	**0.23**		**0.22**	**0.05**		**0.52**	**0.12**
G	female	13	1.11	-0.08	7	0.24	-0.10	6	*0.87*	0.02
E	male	33	0.54	0.37	16	0.22	0.10	17	0.32	0.27
N	**mean**		**0.71**	**0.23**		**0.22**	**0.05**		**0.48**	**0.20**
D	<15	28	0.70	0.18	15	0.04	-0.07	13	*0.66*	0.25
U	15-30	8	0.21	0.39	3	0.80	0.31	5	-0.59	0.08
R	>31	9	0.99	0.33	4	0.13	0.10	5	*0.86*	0.23
	mean		**0.68**	**0.24**		**0.22**	**0.05**		**0.48**	**0.22**
P	<31	13	0.60	-0.05	4	0.61	0.15	9	-0.01	-0.19
F	31-41	19	0.84	0.08	12	0.19	0.15	7	0.65	-0.07
E	>41	10	0.78	0.81	4	-0.26	-0.40	6	*1.04*	*1.21*
V	**mean**		**0.78**	**0.20**		**0.24**	**0.00**		**0.54**	**0.19**
W	<600	10	0.85	0.02	6	1.12	0.55	4	-0.27	-0.53
A	-800	10	0.67	0.87	4	0.03	-0.06	6	*0.64*	*0.93*
L	>800	20	0.76	0.09	9	0.16	-0.12	11	*0.60*	0.21
K	**mean**		**0.77**	**0.23**		**0.31**	**0.03**		**0.39**	**0.20**
A	no	27	0.73	0.30	15	0.19	-0.01	12	*0.53*	0.31
D	one	11	0.86	0.19	6	0.08	-0.41	5	*0.78*	*0.60*
H	both	8	0.18	-0.13	2	0.40	0.54	6	-0.22	-0.67
	mean		**0.71**	**0.23**		**0.22**	**0.05**		**0.46**	**0.21**
S	low	12	0.49	0.39	3	0.51	0.33	9	-0.02	0.06
S	med	9	0.62	0.55	4	-0.09	-0.55	5	*0.71*	*1.10*
E	high	25	0.78	0.13	16	0.11	0.10	9	*0.67*	0.02
R	**mean**		**0.71**	**0.23**		**0.22**	**0.05**		**0.50**	**0.24**
Q	low	23	1.03	0.41	12	1.01	0.60	11	0.02	-0.19
O	high	23	0.37	0.04	11	-0.49	-0.46	12	*0.87*	*0.50*
L	**mean**		**0.71**	**0.23**		**0.22**	**0.05**		**0.44**	**0.10**

First, an enumeration is given of the aptitude-treatment interactions with respect to emotional functioning (see the Δ EM–values in Table 9). Regarding AGE, we can conclude that the treatment has been more beneficial for subjects between 50 and 62 years old. Regarding GENDER, we can conclude that the female subjects profit more from the intervention than the male subjects. Regarding DURILL, the subjects who have a medium duration of illness (15-30 years) do not profit more from the treatment, in contrast with the other subjects who do profit. Regarding PFEV1, the subjects who have the highest forced expiratory volume profit more than the other subjects. Regarding WALKD, the subjects who walk the shortest distance in 12 minutes (i.e., shorter than 600 meter) do not profit from the intervention, in contrast with the other subjects. Regarding ADHER, the subjects who perform no exercises or only one type of exercise profit more from the intervention. Regarding SSER, the subjects who report the lowest seriousness of shortness of breath do not profit from the treatment, in contrast with the other subjects. Regarding QOL, the subjects who report a high quality of life at the pre-test profit more from the treatment.

Second, an enumeration is given of the aptitude-treatment interactions with respect to practical-social functioning (see the Δ PS–values in Table 9). Regarding AGE, GENDER and DURILL we can conclude there are no obvious interactions (the delta values do not differ very much). Regarding PFEV1, the subjects who have the highest forced expiratory volume profit more from the treatment than the other subjects. Regarding WALKD, the subjects who walk the shortest distance in 12 minutes (i.e., shorter than 600 meter) do not profit from the intervention, in contrast with the other subjects. Regarding ADHER, the subjects who perform no exercises or only one type of exercise profit more from the treatment. Regarding SSER, the subjects who report a medium seriousness of shortness of breath profit from the treatment, in contrast with the other subjects. Regarding QOL, the subjects who report a high quality of life at the pre-test profit most from the treatment.

The hypotheses stated in the introduction can only partially be confirmed by these exploratively found interactions. In terms of quality of life the treatment has been more successful for patients who showed the following characteristics at the pre-test 1) a more *subjective* severe medical condition (measured by SSER) and 2) a lower adherence to medical advice (measured by ADHER). However, the treatment has not been more successful for patients who showed at the pre-test 1) a more *objective* severe medical condition (measured by DURILL, PFEV1 and WALKD) and 2) a lower quality of life.

The fourth phase: confirmative ATI analysis

As mentioned in the section Statistical procedure and analyses three methods are used to test the significance of the exploratively found ATI's. The subject scores on the criterion variables were identical in the three methods: these are the change scores[38] on the emotional functioning component and the change scores on the practical-social component. These scores were computed by first projecting the *subject scores* of the

[38] In the literature these change scores are usually called 'gain scores'.

PRINCALS solution (of the first phase) on the directions of respectively emotional and practical-social functioning. Then, the procedure to compute the *change* scores was the same as formulated above for the centroids.

The results of the analyses of variance (ANOVA), the generalized multiple regression analyses (MURALS) and the Permutation tests are given in Table 10 for emotional functioning and in Table 11 for practical functioning. The number of subjects in each analysis are indicated by N. This number is equal in all three methods.

Regarding the ANOVA solutions, only the cell means of the aptitude categories are shown, because the cell means of the treatment categories and interaction categories are already displayed in Table 9. For example: in the ANOVA of AGE and Treatment with the change scores on emotional functioning as dependent variable, the cell means of the treatment categories are 0.71 and 0.22 (see the mean change scores on EM of Treatment and Control in Table 9) and the cell means of the interaction categories are, respectively 1.16, 0.89, 0.82, 0.55, 0.91, -0.23, -0.39 and 0.38 (see the change scores of the AGE-subgroups on EM of Treatment and Control in Table 9). Table 10 also contains the significance level of the main effect of each aptitude variable, the main effect of the treatment variable and the interaction effect of the ANOVA's.

Regarding the MURALS solutions, only the optimal category quantifications of the aptitude variable are shown and the significance levels of the main effects and the interaction effects. The treatment variable is coded -1 (Control) and 1 (Treatment) and the interaction variable is a multiplication of aptitude and treatment variable (see Statistical analysis and procedure). Comparison of the cell means of the aptitude categories in ANOVA and the optimal category quantifications of MURALS reveals that the quantifications reflect the order of the cell means. A higher quantification means a higher change score, i.e., a higher improvement of emotional functioning or practical-social functioning (except for cell means which are very similar).

Due to the fact that MURALS optimizes the relationship between the aptitude and the criterion, the significance levels of the main effect of the aptitude are lower in the MURALS solutions than in the ANOVA solutions. However, the ANOVA solutions of GENDER and QOL are (almost) identical to the MURALS solutions, because these aptitude variables are dichotomous. With respect to *emotional functioning* only the interaction effect of QOL×Treatment is significant in ANOVA and MURALS. The main effect of Treatment is significant (for PFEV1 and QOL) or approximates significance (p = .06 or .07) in the ANOVA and MURALS solutions for all the aptitude variables, except for WALKD[39].

[39]The cell means of the change scores in Treatment and Control have the smallest difference (Δ EM=0.39, see Table 9), probably because of the deletion of subjects who had a missing value on WALKD, leaving only 40 subjects for analysis. Therefore, the main effect of the treatment variable is not significant.

Table 10
Results of significance tests for emotional functioning. Significance values of main effects and interaction effects for each aptitude, obtained with ANOVA and MURALS. Significance values of the Δ Em- values with Permutation tests; *: *p*-value ≤ .05.

		ANOVA		MURALS		Permutation Test	
Effect Category	N	Means	Sign.	Quant	Sign.	Δ Em	Sign
AGE	46		.48		.11		
<50 years		1.03		2.82		0.24	.45
50-58 years		0.15		-1.53		1.12	.08
59-62 years		0.45		-0.05		1.21	.02*
>62 years		0.46		-0.06		0.16	.32
Treatment			.06		.06		
AGE×Treatment			.35		.51		
GENDER	46		.26		.24		
Female		0.71		1.59		0.87	.03*
Male		0.37		-0.63		0.32	.14
Treatment			.06		.06		
GENDER×Treatment			.33		.33		
DURILL	45		.69		.56		
<15 years		0.39		-0.76		0.66	.01*
15-30 years		0.58		1.67		-0.59	.81
>31 years		0.51		0.86		0.86	.06
Treatment			.06		.06		
DURILL×Treatment			.16		.20		
PFEV1	42		.40		.17		
<31%		0.60		0.59		-0.01	.51
31-41%		0.60		0.54		0.65	.06
>41%		0.16		-1.79		1.04	.05*
Treatment			.05*		.06		
PFEV1×Treatment			.38		.28		
WALKD	40		.24		.06		
<600 meter		0.96		1.69		-0.27	.68
600-800 meter		0.29		-0.93		0.64	.12
>800 meter		0.43		-0.38		0.60	.06
Treatment			.14		.14		
WALKD×Treatment			.36		.15		
ADHER	46		1.00		.66		
no practice		0.49		0.34		0.53	.07
practice of one		0.50		0.73		0.78	.07
practice of both		0.34		-2.15		-0.22	.64
Treatment			.07		.06		
ADHER×Treatment			.52		.25		
SSER	46		.60		.35		
low seriousness		0.51		0.26		-0.02	.50
medium		0.22		-2.01		0.71	.11
high seriousness		0.54		0.60		0.67	.02*
Treatment			.06		.06		
SSER×Treatment			.56		.75		
QOL	46		.00*		.00*		
low		1.02		1.00		0.02	.48
high		-0.08		-1.00		0.87	.01*
Treatment			.02*		.02*		
QOL×Treatment			.02*		.02*		

With respect to the ANOVA and MURALS solutions of *practical-social functioning* (see Table 11) the main effects and interaction effects are not significant, except for the main effect of QOL ($p = .00$). An interesting thing to note is that, although PFEV1 and practical-social functioning are completely unrelated (the multiple correlation is zero and therefore, the p-value of the main effect is one), the interaction effect PFEV1×Treatment has a p-value of .06 (MURALS).

Classical ATI analysis assumes a linear relationship between the aptitude and the treatment variable. The optimal quantifications show that almost all aptitude variables have a nonlinear relationship with the outcome measure. Therefore, the strategy to first optimally quantify the aptitude variables and then perform a stepwise regression analysis is considered to be justified. It should be noted, however, that the quantifications do not reflect the results of the second phase of analysis. For example, the conclusion of the second phase was that DURILL and WALKD had a linear relationship with practical-social functioning. The quantifications indicate the opposite (see Table 11). But these quantifications reflect the relationship with the *change* of emotional functioning, not with emotional functioning at time 1 or at time 2.

Both procedures, ANOVA and MURALS followed by linear multiple regression, assume a multivariate normal distribution of the variables. This assumption is strongly violated, because the distribution of AGE, DURILL, WALKD, ADHER and SSER is skewed and the discretization procedure of the aptitudes has preserved the original shape of the variables. The third method, the permutation test, does not assume a multivariate normal distribution. Therefore, the third method was preferred over the other methods. Another important advantage of the permutation test is that it tests the significance of each category of the interaction term. In this way it shows precisely for which category of the aptitude variable the interaction effect is significant. This result is comparable with the result of the Johnson-Neyman technique, which is a technique especially developed for this purpose (see Introduction).

The results of the permutation tests in Table 10 and 11 partially confirm the findings of the explorative phase of ATI. With respect to **emotional functioning** subjects between 59 and 62 years old profit significantly more from the treatment ($p = .02$). Also the female subjects profit significantly more than the male from the treatment ($p = .03$). Regarding the *objective* severity of the medical condition, the treatment is significantly more beneficial for subjects with a *less* severe medical condition, that is, a shorter duration of illness (DURILL < 15 years, $p = .01$) and a higher expiratory volume (PFEV1 > 41%, $p = .05$). Regarding the *subjective* severity of the medical condition, subjects who report a higher seriousness of shortness of breath (SSER > 8) profit significantly more from the treatment ($p = .02$). Regarding the general QOL reported at the pre-test, subjects with a high QOL profit significantly more ($p = .01$). Because the overall treatment-effect for emotional functioning is (almost) significant, the interactions can be considered ordinal (see Introduction).

Table 11. Results for practical-social functioning. Significance values of main effects and interaction effects for each aptitude, obtained with ANOVA and MURALS. Significance values of the Δ PS- values with Permutation tests; *: p-value ≤ .05.

		ANOVA		MURALS		Permutation Test	
Effect Category	N	Means	Sign.	Quant	Sign.	Δ PS	Sign.
AGE	46		.69		.16		
<50 years		0.35		1.34		0.33	.36
50-58 years		-0.02		-0.86		-0.30	.64
59-62 years		0.41		1.57		0.09	.46
>62 years		0.04		-0.61		0.19	.28
Treatment			.64		.62		
AGE×Treatment			.94		.89		
GENDER	46		.23		.23		
Female		-0.09		-1.59		0.02	.51
Male		0.23		0.63		0.27	.17
Treatment			.41		.41		
GENDER×Treatment			.64		.64		
DURILL	45		.65		.39		
<15 years		0.06		-0.73		0.25	.22
15-30 years		0.34		1.79		0.08	.46
>31 years		0.20		0.70		0.23	.35
Treatment			.40		.38		
DURILL×Treatment			.97		.82		
PFEV1	42		.99		1.00		
<31%		0.09		1.26		-0.19	.66
31-41%		0.11		-0.10		-0.07	.55
>41%		0.08		-1.45		1.21	.01*
Treatment			.42		.43		
PFEV1×Treatment			.10		.06		
WALKD	40		.51		.25		
<600 meter		0.23		0.77		-0.53	.82
600-800 meter		0.32		1.21		0.93	.05*
>800 meter		-0.02		-0.99		0.21	.29
Treatment			.44		.45		
WALKD×Treatment			.16		.80		
ADHER	46		.40		.23		
no practice		0.16		0.10		0.31	.17
practice of one		-0.08		-1.47		0.60	.12
practice of both		0.38		1.66		-0.67	.84
Treatment			.33		.32		
ADHER×Treatment			.31		.17		
SSER	46		.41		.26		
low seriousness		0.34		1.47		0.06	.44
medium		-0.06		-1.46		1.10	.03*
high seriousness		0.12		-0.18		0.02	.45
Treatment			.30		.32		
SSER×Treatment			.24		.21		
QOL	46		.00*		.00*		
low		0.50		1.00		-0.19	.70
high		-0.22		-1.00		0.50	.08
Treatment			.48		.48		
QOL×Treatment			.12		.12		

With respect to **practical-social functioning** three significant interactions are found, by means of the permutation method. The treatment is significantly more beneficial for subjects with a less to medium *objective* severe medical condition (PFEV1

> 41%, $p = .01$; WALKD = 600-800 meter, $p = .05$) and a medium *subjective* severe medical condition (SSER = 6 or 7). Because the overall treatment-effect for practical-social functioning is not significant, these interactions can be considered as disordinal.

Conclusions and discussion

It has been shown that the Aptitude Treatment Interaction (ATI) approach is an appropriate way to evaluate the differential effectiveness of a treatment. In ATI-analysis differences within subjects are not treated as 'error variance' (as in analysis of variance), but as important sources of possible differences in the effects of treatments. Thus, ATI is said to be present when for some group of patients an aptitude variable (i.e., a person characteristic) shows a different relation to the outcome measure in the treatment group than it does in the control (or other treatment) group. This implies that the statistical (and practical) importance of the treatment effect differs between the various categories of the aptitude variable. The term 'differential effectiveness' is chosen to refer to this phenomenon.

The ATI paradigm has been frequently applied to research in the area of educational psychology. However, several limitations of classical ATI-analysis have been described: (a) the use of mainly *univariate* analysis: one outcome variable, one treatment variable and one aptitude variable; (b) the use of only continuous variables (except for the treatment variable, which is categorical); (c) the assumption of a numerical level of measurement and (d) the assumption of a linear relationship between aptitude and outcome variable. The present study extends the ATI-paradigm by:

- the use of multivariate analysis
- the incorporation of a mixture of measurement levels (categorical and numerical)
- allowing nonlinear relationships.

The newly developed analysis strategy has been applied to the evaluation of a psychosocial treatment for COPD on quality of life. The research hypothesis stated that the treatment would be more successful in terms of quality of life for patients, who at the pre-test showed (1) a more severe medical condition; (2) a lower adherence to medical advice and (3) a lower quality of life. The corresponding study objectives were, respectively: 1) to clarify the relationships between the subscales of QOL; 2) to explore relationships between the aptitude variables and QOL; 3) to determine exploratively which aptitude variables interact with the experimental conditions; 4) to test the ATI's and the hypothesis.

A randomized pre-posttest design was used with two conditions. The treatment group (n = 50) received information and cognitive-behavioral therapy, and the control group (n = 46) received information only. The preliminary analyses of this study, with the subjects from the pre-test (n = 96), involved the clarification and reliability of the measures. First, the subscales of the outcome measure QOL were derived from items of the General Questionnaire for CNSLD-patients (GQCP). Eleven subscales were constructed on the basis of the results of Principal Component Analyses. These scales reflected five domains of functioning: physical functioning (fatigue, congestion, dyspnea), emotional functioning (depression, anxiety, anger), cognitive functioning (memory), practical functioning (special and general) , and social functioning (special and general). The reliability of the scales was high (varying from 0.82 to 0.92), except for the scales which measure social functioning (0.76 and 0.69). Adjustment of the items in the GQCP of the latter scales is needed to increase the reliability of the instrument. Second, aptitude measures were derived from the GQCP and medical data. In addition to age and gender, five other aptitude measures were chosen on basis of the hypothesis. The severity of the medical condition was measured by 'subjective seriousness' (i.e., a 10-points-scale) and 'objective seriousness' (i.e., 'duration of illness', 'forced expiratory volume in 1 second' and 'walking distance in 12 minutes'). Adherence to medical advice was measured by 'practice of breathing and/or relaxation exercises'.

Subsequently, four phases of analyses were performed to address the study objectives. Only the 46 patients who completed both measurements remained in the analysis. In the first phase of analysis, Principal Components Analysis (PRINCALS) of the eleven subscales measured at two times revealed that two main components could be distinguished: emotional functioning and practical-social functioning. QOL was a 'summation' of these two components. The physical functioning component did not account for an important additional amount of the total variance of the subjects. A plausible explanation for this finding is that all of the subjects received physical training as part of the standard care program at the Hornerheide center. The new components divided the configuration of the eleven subscales in four quadrants: GG, good practical-social and good emotional functioning; GB, good practical-social and bad emotional functioning; BB, bad practical-social and bad emotional functioning; BG, bad practical-social and good emotional functioning.

In the second phase, the relationships between the aptitudes and the main components of QOL were studied by means of a centroid model approach and a vector model approach. In the *centroid model approach,* the categories of an aptitude variable identify different groups of subjects. The centroid is the center of gravity of all subjects belonging to the same category. Graphically it is the point which has the means of the subject scores on each component (from the first phase) as coordinates. An asset of this approach is that it can reveal a nonlinear relationship between an aptitude variable and the outcome measure QOL. The centroids were computed for each time separately to inspect the course of QOL (expressed in the position of the centroids at time 1 and time 2 in the quadrants) within each aptitude-category. Continuous aptitude variables had to be discreticized before computing the centroids. This was done by an optimal discretization procedure, which follows the original shape of the distributions as closely as possible. In the *vector model approach* the aptitude variables are considered to be 'external' variables which can be projected in the configuration of the QOL-scales (resulting from the first phase of analysis). The length of the vector indicates the strength of the relationship (i.e., the amount of variance of the aptitude variable accounted for by the components). Some important results were found. The medical indicator 'forced expiratory volume in 1 second as a percentage of the predicted value' (PFEV1) has a nonlinear relationship with both emotional and practical-social functioning. The vector model approach showed that PFEV1 has almost zero correlation[40] with QOL. In the literature concerning QOL and COPD, however, most studies yielded a strong relationship between FEV1 and QOL (Curtis, Deyo & Hudson, 1994, p. 166). The present study showed in addition, that the aptitudes 'duration of illness' and 'walking distance' have a linear relationship with practical-social functioning and that they are not associated with emotional functioning. 'Subjective seriousness of illness' has a nonlinear relationship with emotional functioning, but is not associated with practical social functioning.

The third phase was directed at exploring possible interactions between the aptitude variables and the treatment variable (ATI's), by means of the centroid model approach. This time centroids were computed for each category of an aptitude at each time *within* groups: the treatment group (Treatment) and the control group (Control). If the course of QOL would be different between Treatment and Control for the same aptitude category, this would imply an interaction. The results reveal an overall treatment effect for emotional functioning and no treatment effect for practical-social functioning. Secondly, and most important in view of the research objectives, ATI's were found with respect to both emotional functioning and practical-social functioning. Thus, the treatment effect was found to be more pronounced for particular subgroups of patients. For which subgroups the interactions had statistical importance was confirmed in the fourth phase of analysis.

By means of three methods the significance of the ATI's was tested, respectively: ANOVA, MURALS and a permutation test. Classical ATI analysis (i.e., stepwise multiple regression) was considered to be inappropriate, due to the nonlinear relationships between aptitudes and both components of QOL. Therefore, a special generalization of multiple regression was applied: MURALS. The MURALS algorithm maximizes the multiple correlation coefficient between the criterion variable and a linear

---

[40] r^2(PFEV1,QOLt1) = .02 ; r^2(PFEV1,QOLt2) = .04.

combination of the predictor variables, which defines optimality for the transformations of the variables. The permutation test is a form of resampling, in which observations are independently redistributed among subjects. In this study, the observed category scores on the treatment variable and the (discreticized) aptitude variable were permuted to test the significance of the exploratively found ATI's. An ATI was expressed as the difference between the treatment-gain score and the control-gain score on each component of QOL for each aptitude subgroup. These latter scores were referred to as delta-scores. The probability that the delta-scores for the permutated data are greater than the originally observed delta-scores, was estimated by the use of 1000 permutations.

The ANOVA and MURALS solutions confirmed the existence of an overall treatment effect for emotional functioning and not for practical-social functioning. This result corresponds with the results of primary analyses of the data (Van den Broek, 1995). However, Van den Broek has found a significant treatment effect for social functioning as well, a result which is not replicated in this study. The difference in results between the two studies might be explained by the use of a multivariate approach of QOL instead of a univariate approach, as applied by Van den Broek.

Due to the small sample size the ANOVA approach and the MURALS approach appeared to be not powerful enough to detect significant aptitude-treatment-interaction effects. Therefore, the permutation test is preferred to the other methods, because it is less sensitive to small sample sizes. Moreover, the test does not assume a multivariate normal distribution, which is an additional advantage since the distribution of several aptitudes is skewed. The permutation tests showed that with respect to *emotional functioning* subjects between 59 and 62 years old profited significantly more from the treatment than patients between 42-58 and between 63-70 years old. Also the female subjects profited significantly more than the male subjects. The hypothesis could only be confirmed regarding the *subjective* severity of the medical condition. The treatment was significantly more beneficial for subjects who reported a high seriousness of 'shortness of breath' (8-10) than for those with a lower estimated seriousness (1-7). On the other hand, the permutation tests showed that the treatment was significantly more beneficial for subjects with a *less severe objective* medical condition (i.e., a shorter duration of illness (< 15 years) and a higher forced expiratory volume (> 41%). In addition, the treatment was significantly more effective for subjects who report a high QOL at the pre-test than for subjects with a low QOL. No significant ATI was found concerning the adherence to medical advice. With respect to *practical-social functioning* the hypotheses could not be confirmed. The treatment was significantly more beneficial for subjects who reported a medium severe subjective medical condition (6 or 7) and a less to medium severe objective medical condition (forced expiratory volume > 41%; walking distance in 12 minutes = 600-800 meter).

The main purpose behind this study was to underline the importance of Aptitude-Treatment-Interaction analysis for the evaluation of treatments with regard to their effectiveness within the field of health psychology. Whenever an overall treatment effect is found, ATI analysis may reveal for which subgroups of patients this effect is of particular statistical and therefore practical significance. But also in cases when no significant treatment effects are found, ATI analysis may give insight in differential effects. This is very important, particularly because some previous studies on the effectiveness of psychosocial treatments for COPD showed little or no overall treatment effect (in: Van den Broek, 1995).

More research on the differential effectiveness of psychosocial treatments for COPD is needed before guidelines can be formulated regarding the development of screening measures for the referral of patients to an appropriate type of care. Although the sample of this study is a fair representation of the total COPD population with respect to gender and age (e.g., Cugell, 1988), the number of patients who completed the pre- and posttest is rather small (23 in Treatment and 23 in Control) and does not allow drawing far-reaching conclusions. Cronbach and Snow (1977) have argued that *"an ATI study with S's assigned at random to one of two treatments ought to employ something like 100 S's per treatment"* (p. 46). Their argument is based on the finding that the sampling error of a difference in correlation (i.e., in regression slopes) is 1.4 times the sampling error of each correlation. Furthermore, due to the fact that the number of subjects included in the analyses was small, categories of aptitude variables had to be merged, which has led to a loss of information. The random assignment of the subjects to Treatment or Control can be considered successful with respect to QOL, but not with respect to the various aptitudes. In ATI research stratification or matching of patients on important person characteristics is recommended, to avoid that a category of an aptitude is underrepresented.

This study initiates extension of the ATI paradigm by the use of multivariate analysis, a mixture of measurement levels (categorical and numerical) and nonlinear relationships. Figure 3 gives a framework of possible models for ATI research.

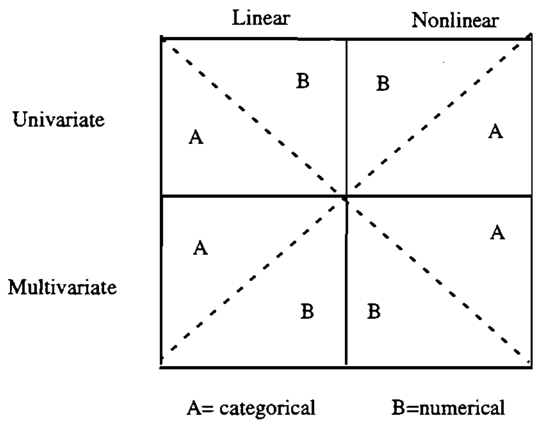

Figure 3. Framework of possible models for ATI research.

The present study used a multivariate approach to study outcome measures. However, the different aptitudes were still treated separately, and the relationships between the aptitudes were not examined. Tate (1981) proposes a multivariate approach of both aptitudes and outcome measures, whenever the classical assumptions for multivariate analyses are met. The aim of this study was to identify subgroups of patients for whom the treatment was more beneficial. When the research is directed at the individual subject, Rogosa (1991) proposes three interesting ATI models for individual growth. Further study is needed to elaborate different kinds of ATI models and the relationships between the models to overcome the short-comings of classical ATI-analysis.

Graphical representations

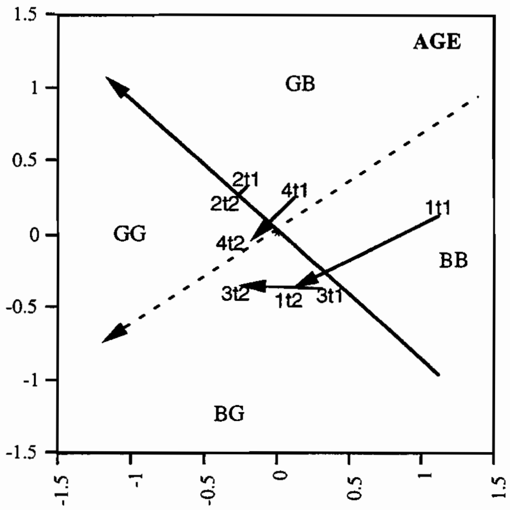

Figure 4a. Configuration of the centroids at time 1 (t1) and time 2 (t2) for AGE. First number indicates the category: '1', younger than 50 years; '2', 50-58 years; '3', 59-62 years; '4', older than 62. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'. GG, good practical-social and good emotional; GB, good practical-social and bad emotional; BB, bad practical-social and bad emotional; BG, bad practical-social and good emotional.

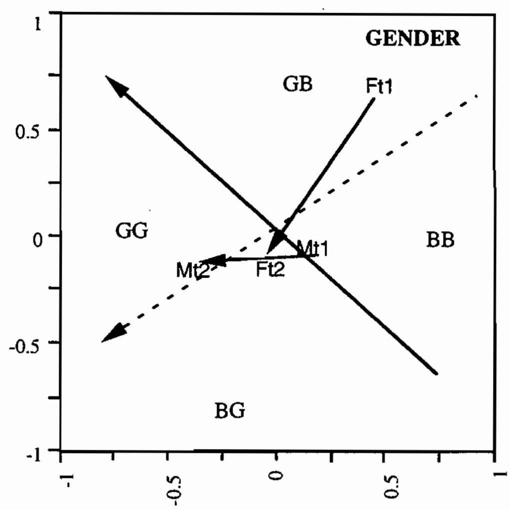

Figure 4b. Configuration of the centroids at time 1 (t1) and time 2 (t2) for GENDER. First letter indicates the category: 'F', female; 'M', male. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

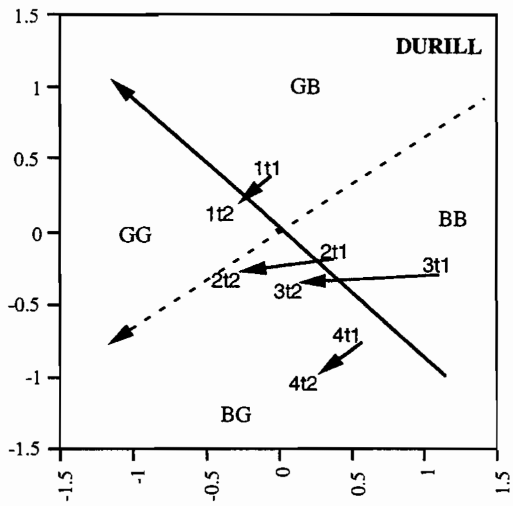

Figure 4c. Configuration of the centroids at time 1 (t1) and time 2 (t2) for DURILL. First number indicates the category: '1', shorter than 15 years; '2', 15-30 years; '3', 31-45 years; '4', older than 45. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

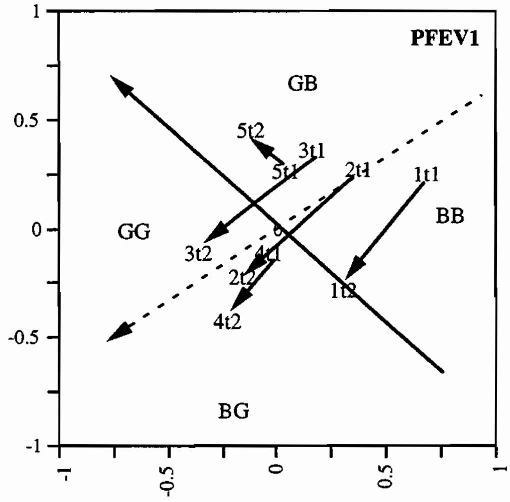

Figure 4d. Configuration of the centroids at time 1 (t1) and time 2 (t2) for PFEV1. First number indicates the category: '1', lower than 20%; '2', 21-30%; '3', 31-41%; '4', 42-50%; '5', higher than 50%. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

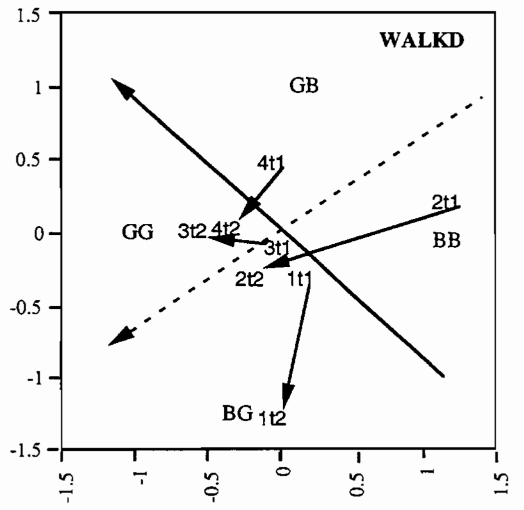

Figure 4e. Configuration of the centroids at time 1 (t1) and time 2 (t2) for WALKD. First number indicates the category: '1', less than 500 meter; '2', 500-600 meter; '3', 600-800 meter; '4' 800 meter or more. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

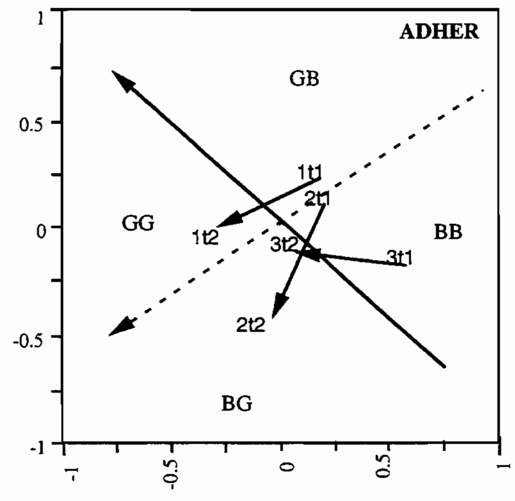

Figure 4f. Configuration of the centroids at time 1 (t1) and time 2 (t2)for ADHER. First number indicates the category: '1', no practice; '2', practice of one; '3', practice of both exercises. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

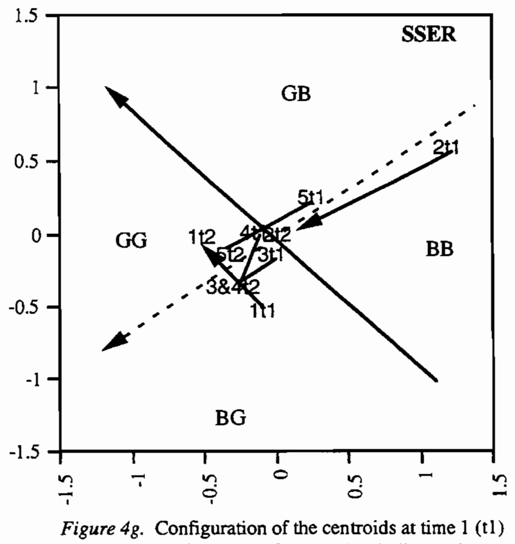

Figure 4g. Configuration of the centroids at time 1 (t1) and time 2 (t2) for SSER. First number indicates the category: '1', not severe; '3', medium severe; '5', very severe shortness of breath. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

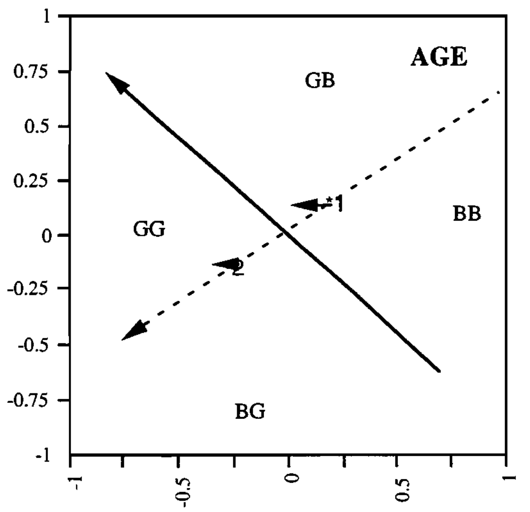

Figure 5a. Configuration of the vectors at time 1 (1) and time 2 (2) for AGE, projected in the configuration of the QOL- components; '*' is point of departure. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'. GG: good practical-social, good emotional; GB: good practical-social, bad emotional; BB: bad practical-social, bad emotional; BG: bad practical-social, good emotional.

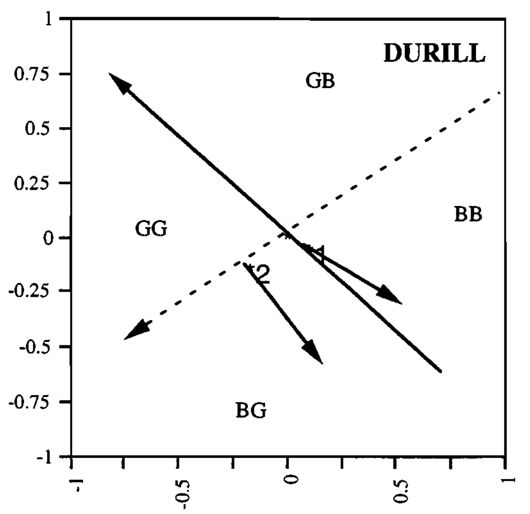

Figure 5b. Configuration of the vectors at time 1 (1) and time 2 (2) for DURILL, projected in the configuration of the QOL-components; '*' is point of departure. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

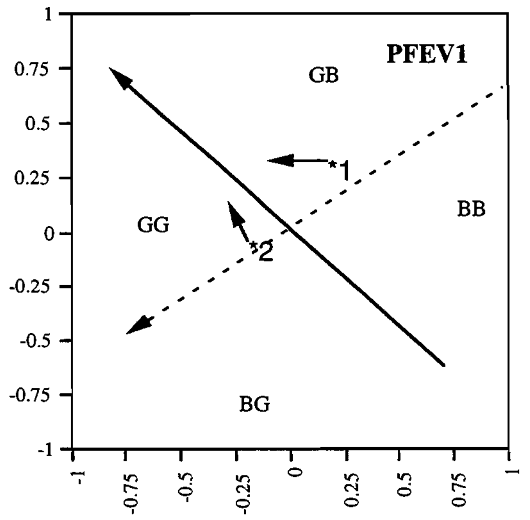

Figure 5c. Configuration of the vectors at time 1 (1) and time 2 (2) for PFEV1, projected in the configuration of the QOL-components; '*' is point of departure. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

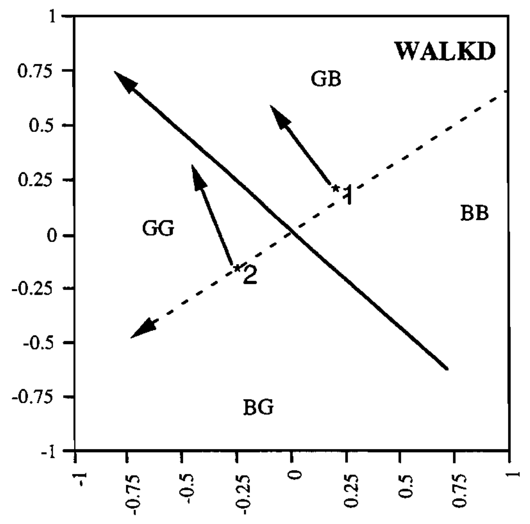

Figure 5d. Configuration of the vectors at time 1 (1) and time 2 (2) for WALKD, projected in the configuration of the QOL-components; '*' is point of departure. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

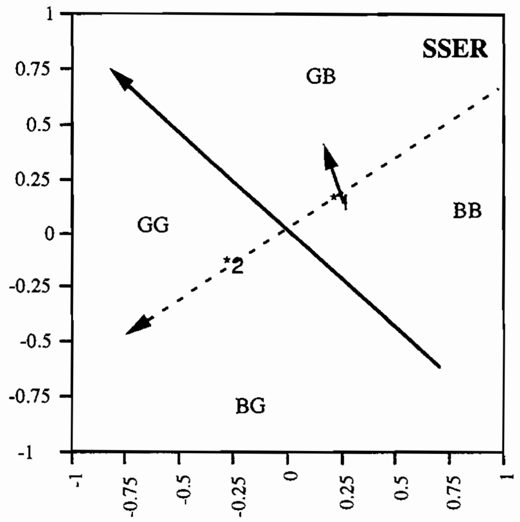

Figure 5e. Configuration of the vectors at time 1 (1) and time 2 (2) for SSER, projected in the configuration of the QOL-components; '*' is point of departure. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

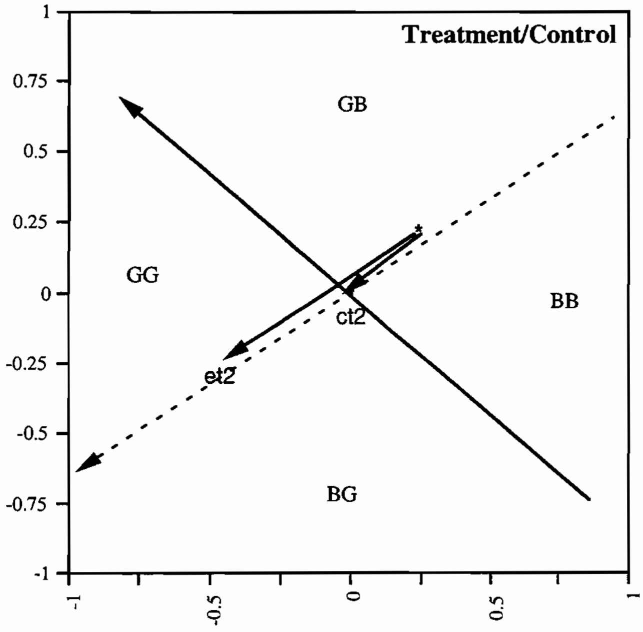

Figure 6a. Configuration of the centroids at time 1 (*) and time 2 (t2) for Treatment and Control. First letter indicates the group: 'e', experimental group (i.e., treatment); 'c', control group. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'. GG, good practical-social and good emotional; GB, good practical-social and bad emotional; BB, bad practical-social and bad emotional; BG, bad practical-social and good emotional .

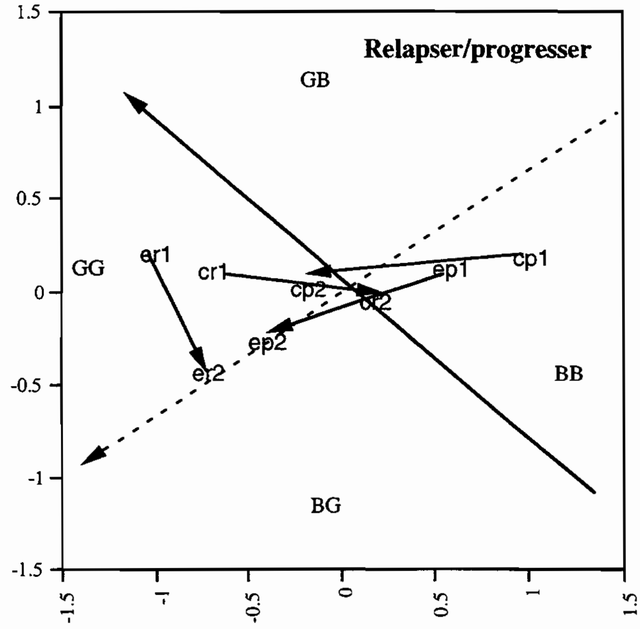

Figure 6b. Configuration of the centroids at time 1 and time 2 for relapsers and progressers in Treatment and Control. First letter indicates the group: 'e', experimental group (i.e., treatment); 'c', control group. Second letter indicates the subgroup: 'r', relapsers; 'p', progressers. Last number indicates the time: '1', time 1; '2', time 2. Lond solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

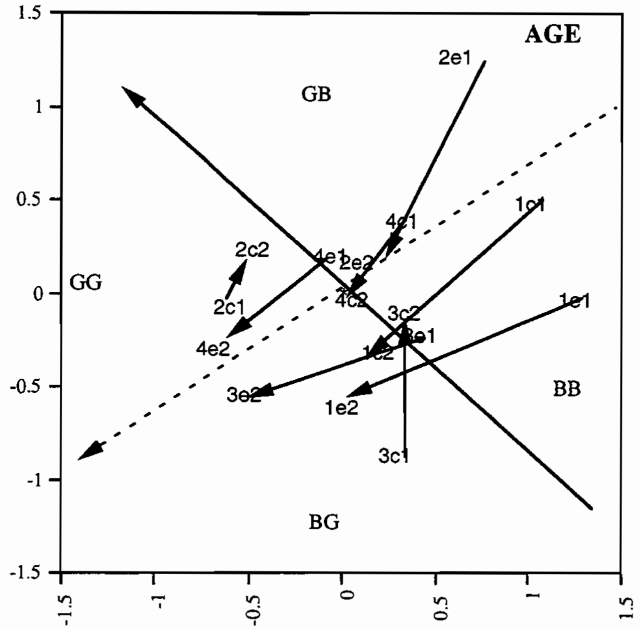

Figure 6c. Configuration of the centroids for subgroups of AGE in Treatment and Control. First number indicates the subgroup: '1', younger than 50 years; '2', 50-58 years; '3' 59-62 years; '4' older than 62. The letter indicates the group: 'e', experimental group; 'c', control group. The last number indicates the time: '1', time 1; '2', time 2. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

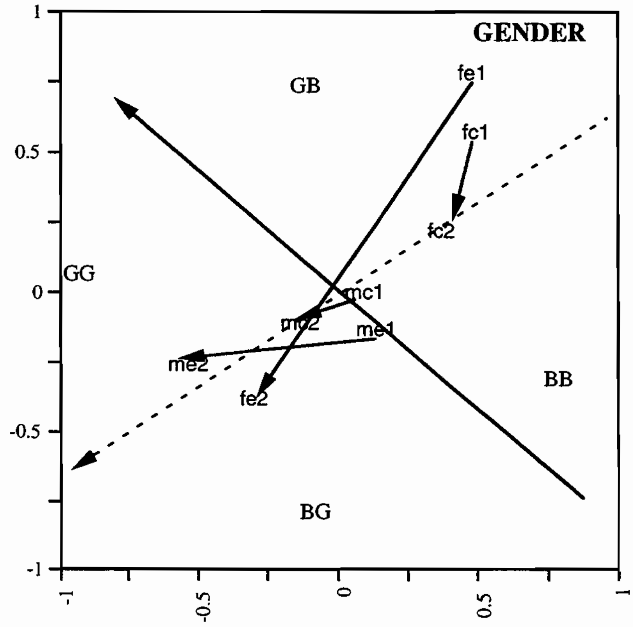

Figure 6d. Configuration of the centroids for subgroups of GENDER in Treatment and Control. First letter indicates the subgroup: 'f', female; 'm', male. The second letter indicates the group: 'e', experimental group; 'c', control group. The number indicates the time: '1', time 1; '2', time 2.

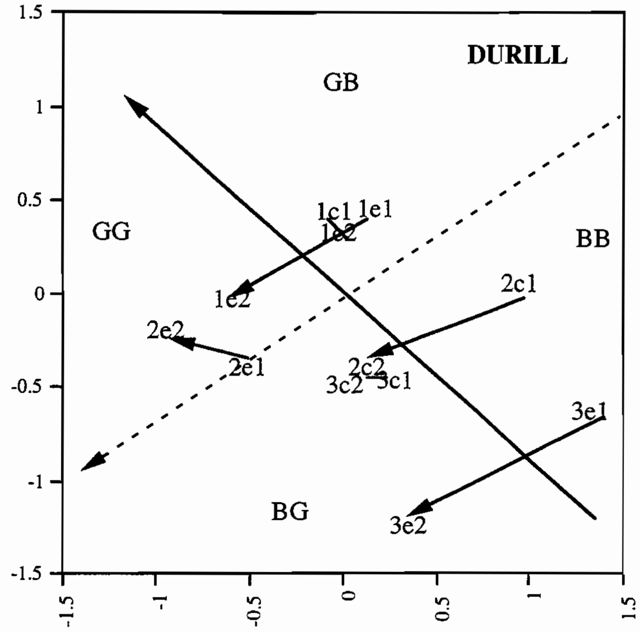

Figure 6e. Configuration of the centroids for subgroups of DURILL in Treatment and Control. First number indicates the subgroup: '1', shorter than 15 years; '2', 15-30 years; '3' longer than 62. The letter indicates the group: 'e', experimental group; 'c', control group. The last number indicates the time: '1', time 1; '2', time 2. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

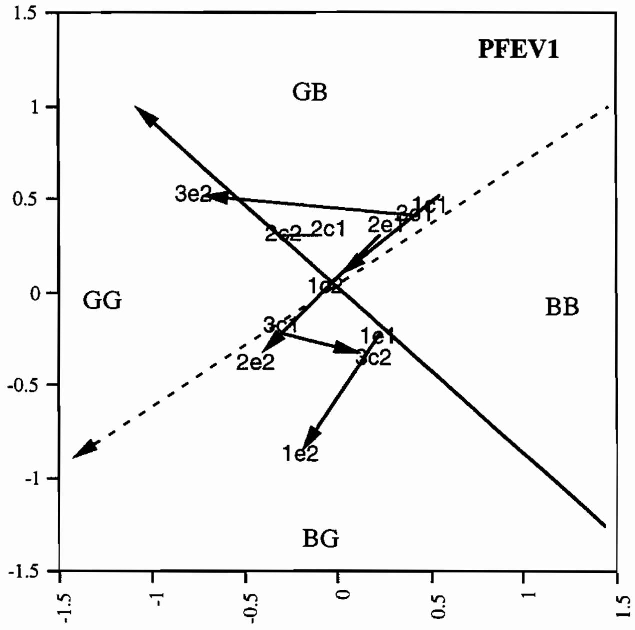

Figure 6f. Configuration of the centroids for subgroups of PFEV1 in Treatment and Control. First number indicates the subgroup: '1', lower than 31%; '2', 31-41%; '3' higher than 41%. The letter indicates the group: 'e', experimental group; 'c', control group. The last number indicates the time: '1', time 1; '2', time 2. Long solid vector is 'practical-social functioning'; dotted vector is 'emotional functioning'.

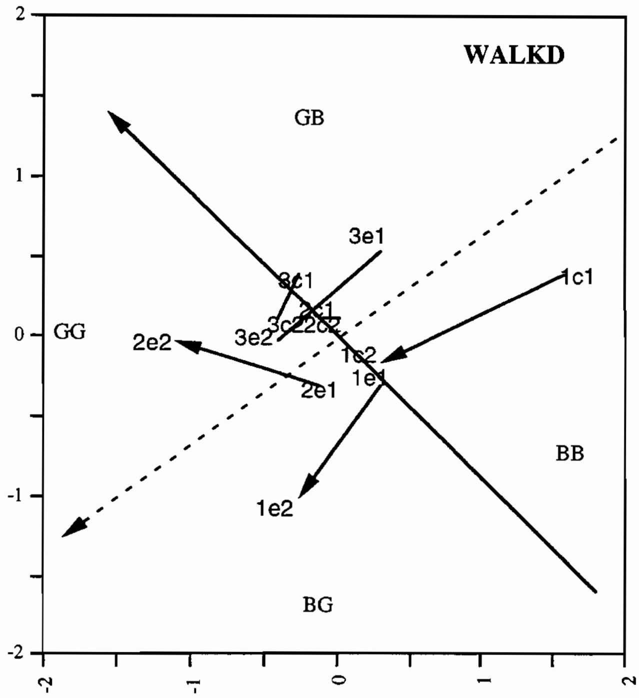

Figure 6g. Configuration of the centroids for subgroups of WALKD in Treatment and Control. First number indicates the subgroup: '1', less than 600 meter; '2', 600-800 meter; '3' 800 meter or more. The letter indicates the group: 'e', experimental group; 'c', control group. The last number indicates the time: '1', time 1; '2', time 2.

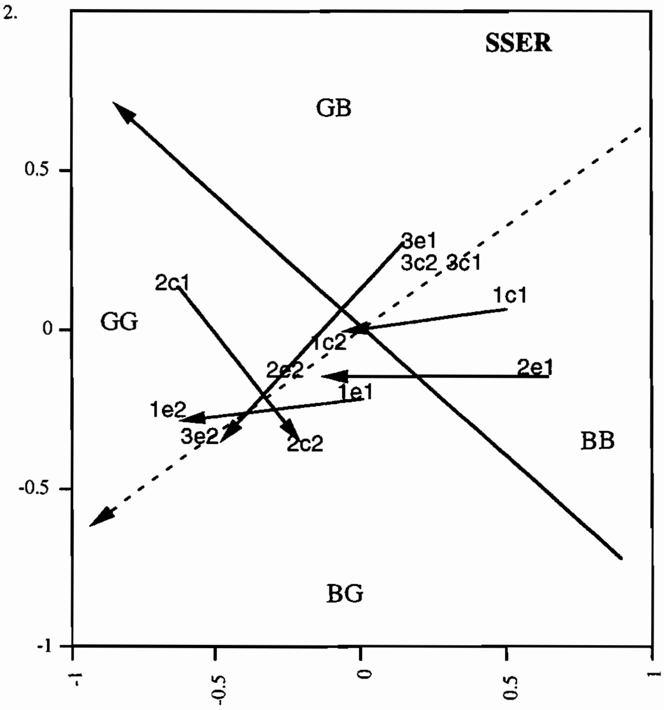

Figure 6h. Configuration of the centroids for subgroups of SSER in Treatment and Control. First number indicates the subgroup: '1', low seriousness; '2', medium seriousness; '3' high seriousness. The letter indicates the group: 'e', experimental group; 'c', control group. The last number indicates the time: '1', time 1; '2', time 2.

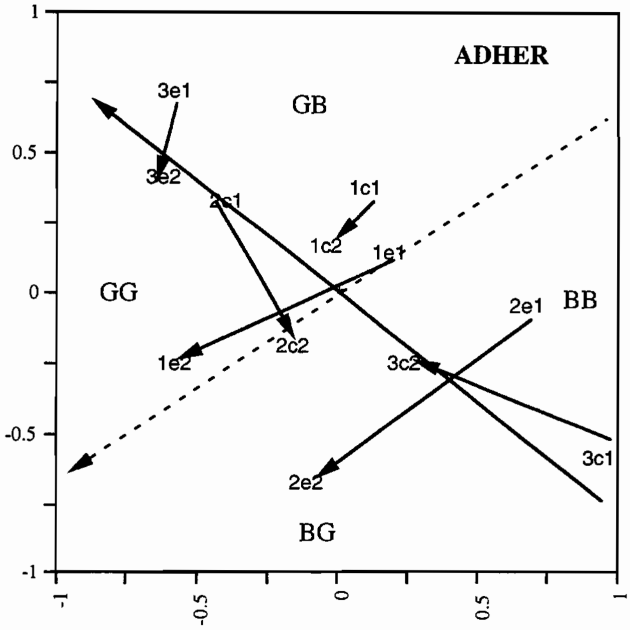

Figure 6i. Configuration of the centroids for subgroups of ADHER in Treatment and Control. First number indicates the subgroup: '1', no practice; '2', practice of one; '3' practice of both exercises. The letter indicates the group: 'e', experimental group; 'c', control group. The last number indicates the time: '1', time 1; '2', time 2.

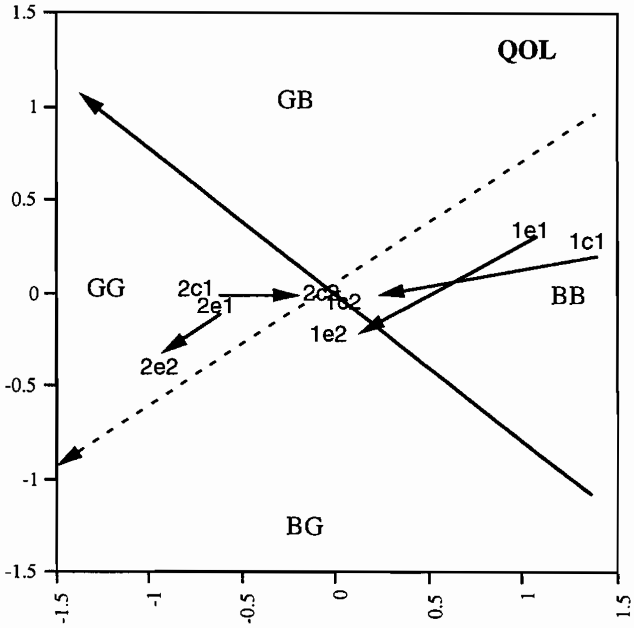

Figure 6j. Configuration of the centroids for subgroups of QOL in Treatment and Control. First number indicates the subgroup: '1', low quality of life at pre-test; '2', high quality of life at pre-test. The letter indicates the group: 'e', experimental group; 'c', control group. The last number indicates the time: '1', time 1; '2', time 2.

References

Cronbach, L. J. (1957). The Two Disciplines of Scientific Psychology. *The American Psychologist, 12,* 671-684.

Cronbach, L. J. and Snow, R. E. (1977). *Aptitudes and instructional methods.* New York: Irvington Publishers, inc.

Cugell, D.W. (1988). COPD: A Brief Introduction for Behavioral Scientists. In A.J. Mc Sweeny & I. Grant (Eds.), *Chronic Obstructive Pulmonary Disease.* New York: Marcel Dekker, Inc.

Curtis, J. R., Deyo, R.A. & Hudson, L.D. (1994). Health-related quality of life among patients with chronic obstructive pulmonary disease. *Thorax, 49,* 162-170.

Dance, K.A. & Neufield, R.W.J. (1988). Aptitude-Treatment Interaction in the Clinical Setting: A Review of Attempts to Dispel The "Patient Uniformity" Myth. *Psychological Bulletin, 104(2),* 192-213.

Edwards, A. E. (1985). *Multiple Regression and the Analysis of Variance and Covariance.* (2nd Edition). New York: W.H. Freeman and Company.

Gifi, A. (1990). *Nonlinear multivariate analysis,* Chichester: Wiley.

Gleitman, H. (1991). *Psychology* (3rd Edition). New York: W.W. Norton & Company.

Heiser, W. J. & Meulman, J. J. (1994). Homogeneity Analysis: exploring the distribution of variables and their Nonlinear Relationships. In M. Greenacre & J. Blasius (Eds.), *Correspondence Analysis in the Social Sciences* (pp. 179-226). San Diego: Academic Press.

Hodgkin, J.E., Zorn, E.G. & Conners, G.L. (Eds.) (1984). *Pulmonary Rehabilitation. Guidelines to success.* Boson: Butterworth publishers.

Kaplan, R.M. & Atkins, C.J. (1988). Behavioral Interventions for Patients with COPD. In A.J. Mc Sweeny & I. Grant (Eds.), *Chronic Obstructive Pulmonary Disease.* New York: Marcel Dekker, Inc.

Kinsman, R.A., Fernandez, E., Schocket, M., Dirks, J.F. & Covino, N.A. (1983). Multidimensional Analysis of the Symptoms of Chronic Bronchitis and Emphysema. *Journal of Behavioral Medicine, 6(4),* 339-357.

Liem, S.D. & Maillé, A. R. (1987). Preventieve gezondheidsvoorlichting ouderen: effecten van GVO-cursussen in Tilburg. [Preventive health education for the elderly: effects of health education courses in Tilburg]. Tilburg: Tilburgse Kruisvereniging.

Maes, S. & Van Veldhoven, M. (1990). From Health Behavior to Health Behavior Change. In: A.A. Kaptein, H.M. van der Ploeg, B. Garssen, P.J.G. Schreurs & R. Beunderman (Eds.), *Behavioral Medicine* (pp. 33-47). Chichester: Wiley.

Meulman, J.J., Heiser, W.J. & Carroll, J.D. (1986). *Prefmap-3 User's Guide.* Murray Hill, NJ: AT&T/Bell Laboratories.

Pedhazur, E. J. & Pedhazur-Schmelkin, L. (1991). *Measurement, Design, and Analysis: an integrated approach* (Student ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Petty, T.L. (1988). Medical Management of COPD. In A.J. Mc Sweeny & I. Grant (Eds.), *Chronic Obstructive Pulmonary Disease.* New York: Marcel Dekker, Inc.

Pothoff, R.F. (1964). On the Johnson-Neyman Technique and some extensions thereof. *Psychometrika, 29,* 241-256.

Rogosa, D. (1980). Comparing Nonparallel Regression Lines. *Psychological Bulletin, 88(2),* 307-321.

Rogosa, D. (1991). A Longitudinal Approach to ATI research: Models for Individual Growth and Models for Individual Differences in Response to Intervention. In R.E. Snow & D. Wiley (Eds.), *Improving inquiry in social science. A volume in Honor of Lee J. Cronbach* (pp. 221-248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schlösser, Maes & Van den Broek (1990). General Questionnaire for CNSLD-Patients. Translated from original Dutch version: Algemene vragenlijst voor CARA-patiënten.

Shoham-Salomon, V. & Hannah, M.T. (1991). Client-treatment interaction in the study of differential change processes. *Journal of Consulting and Clinical Psychology, 59,* 217-225.

Snow, R. E. (1991). Aptitude-treatment-interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology, 59,* 205-216.

SPSS *Categories* (1990). Chicago: SPSS Inc.

Stinard, T. A. & Dolphin, W. D. (1981). Which students benefit from Self-Paced Mastery Instruction and Why. *Journal of Educational Psychology, 73,* 754-763.

Tate, R. L. (1981). Multivariate ATI analysis. *Multivariate Behavioral Research, 16(2),* 243-259.

Van den Broek, A.H.S. (1995). *Patient education and Chronic Obstructive Pulmonary Disease.* Health psychology series, **1**.

Van der Ham, T., Meulman, J.J., Van Strien, D.C. & Van England, H. (1995). Empirically based subgrouping of eating disorders in adolescents by principal components analysis: a longitudinal perspective. *Research report RR-95-01,* Leiden: Department of Data Theory. Accepted for publication in British Journal of Psychiatry.

Van der Kooij, A. & Meulman, J.J. (1996). *Murals Manual.* Unpublished paper.

Van Rijckevorsel, J., Bettonvil, B. & De Leeuw, J. (1985). Recovery and Stability in nonlinear Principal Components Analysis. *Research report RR-85-21,* Leiden: Department of Data Theory.