Quality of life assessment in adult congenital heart disease: A validation of formal instruments

by

Michael D. Cooper, MD

A Thesis

Presented to the
Department of Public Health and Preventative Medicine
and the Oregon Health Sciences University
School of Medicine
in partial fulfillment of
the requirements for the degree of
Master of Public Health

June 1999

School of Medicine

Oregon Health Sciences University

Certificate of Approval

This is to certify that the M.P.H. Thesis of

Michael D. Cooper, MD

Has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

Associate Dean for Graduate Studies

Table of Contents

Acknowledgment:

The author would like to thank Cynthia Morris, Ph.D., MPH, for her support and guidance through this entire project. The study would not have been possible without the aid and access to the congenital heart disease registry she provided.

Additionally, the services of Mr. Robert Shruff of the General Clinical Research Center at Oregon Health Sciences University were essential in establishing and maintaining the study's database.

This study was supported in whole by the General Clinical Research Center at Oregon Health Sciences University, PHS Grant # 5MO1RR00334.

Abstract:

Introduction: Quality of life has become an important outcome measure for clinical research. Past evaluations of quality of life in adult congenital heart disease (ACHD) have been scant and have not utilized available standard formal assessment instruments.

Objectives: Our objective for this study was to determine the validity of several health related quality of life assessment tools with regard to their physical health domains and provide criterion validation for these instruments in our target population.

Methods: A cohort of subjects with ACHD was randomly selected from an existing database for invitation to participate in this protocol. Additional inclusion of subjects from an active ACHD medical clinic enhanced the study population in both illness severity and defect variety. After informed consent was given, subjects were asked to complete the Medical Outcomes Study Short Form 36, Short Form 12, Sickness Impact Profile, and Duke Activity Status Index questionnaires. They then underwent exercise testing with respiratory gas analysis. The quality of life questionnaire physical domain scores were then correlated with objective measures of physical ability. Specifically the physical domain assessments were compared to peak oxygen consumption, work performed, metabolic equivalents achieved, and anaerobic threshold.

Results: A moderate correlation was found between the physical domain score of instruments tested and peak oxygen consumption ($r=0.5$, $p<0.01$). The same moderate degree of correlation was found between these scores and work and metabolic equivalents achieved ($r=0.6$, $p<0.01$). There was no correlation between anaerobic threshold and the physical domain of quality of life.

Conclusion: Despite finding a significant correlation between the physical domain of perceived quality of life and objective exercise endpoints, the strength of the correlation was modest. The variance in perceived physical aspects of quality of life was only partially explained by maximal exercise abilities.

Introduction:

Adults with congenital heart disease are a new and expanding patient population. It has been estimated that over the last 25 years, more than 500,000 children with congenital cardiac defects have reached adulthood. [1] By the year 2000, adults with congenital heart disease will exceed 900,000 in number. [5] . Using a birth prevalence for congenital heart disease of 8 per 1,000 live births, some 20,000 additional patients will reach adulthood each year. [43] This new population is a direct result of the progression of surgical technique. As technique has improved, so has survival. With this growing population, clinical questions of optimal screening, medical therapy, surveillance strategies and reproductive safety arise. We currently do not have sufficient research data to make confident, cogent therapeutic recommendations. Adult congenital heart disease is truly a developing arena.

To better understand the evolving natural history of adult congenital heart disease, initial attempts at data acquisition have focused on hard endpoints. Excellent data using this approach is available for large cohorts of individuals with congenital heart disease. [2,3,4] For example, rates of late cardiac related deaths depend on the era of repair and the defect repaired. For the earliest era of open-heart surgery (1958-1964), the 25 year mortality rates range from 1% (patent ductus arteriosus) to 17% (aortic stenosis). [4] Independent of operative era, potential late sequelae of surgical repair exists for all but the most minor defects. Of primary concern are the incidences of cardiac arrhythmias, infective endocarditis, congestive heart failure and re-operation. Such morbidity is common across all types of congenital defects. Rates for such morbid events are available. Any one of these events has the potential to compromise cardiac function and thus health related quality of life. The events themselves, however, fail to shed light on quality of life. Health related quality of life has not been adequately investigated in adult congenital heart disease.

Investigation of "quality of life" can be difficult, for *"the concept of quality of life is subject to numerous interpretations."* [9] Although somewhat artificial, for research purposes quality of life can be practically and operationally defined as an individual subjective perception of life that incorporates all circumstances of living (physical function, mental function, social interaction, work, recreation, etc.). [12] The medical research utilizing quality of life has tended to focus on functional status[9], which is a component of overall quality of life. Arguably, functional status could be equated to the physical aspects of quality of life.

Functional status has been shown to be a predictor of overall mortality and is also important in prognosis, especially for surgical and maternal risk. [5] Functional status can generally be defined as the ability to physically carry out tasks and activities associated with life. Measurement of functional status, however, proves to be much more difficult than dichotomous endpoints; such as, mortality, re-operation, endocarditis or heart failure. To complicate matters, the measurement of functional status is somewhat philosophical. Should one measure performance of daily activities, maximal exercise capacity, or patient perception of functional ability? It has been shown that these various estimates of "functional status" measure different abilities. [6] Laboratory measurements typically evaluate maximal exercise capacity, while questionnaires typically record the ability to perform routine tasks or behaviors. Guyatt et al. looked at individuals with chronic heart and pulmonary conditions, only to find that four different functional status questionnaires provided poor correlation with laboratory measures of exercise capacity. [6] They concluded that laboratory measures of "functional capacity" should be differentiated from the ability to undertake physical activities of daily living. [6] Similar findings have been established by other investigators in non-cardiac conditions. [3]

This discrepancy between laboratory derived measures and patient perception is at the heart of the trend in measurement of functional status and health related quality of life. Traditional quantifiable endpoints (e.g. years of life, cardiac output,

tumor size, etc.) have been replaced with patient centered measures. The goals of health care in general are shifting. Making patients live better, not just longer has become a new paradigm. [7,8,9,10] This is particularly true of areas such as oncology and incurable chronic disease, where improved survival did not necessarily translate into better overall patient care. [9] The altruistic use of quality of life measures allows for comparison of patient outcomes between treatment strategies, health care systems, specialty care, intensity of resource use, etc. [8,10,11] Such comparisons using an "objective measure" of a very subjective entity would hopefully allow tracking of patient perception of their response to care. This data can then be used for program and policy changes, which would improve patient outcomes.

Health related quality of life measurement has also been adopted for program and policy formulation from a cost utility perspective. This area of application is somewhat contentious as it sheds the altruistic goals and focuses upon a monetary bottom line. In attempts at cost containment of skyrocketing medical expenses, third party payers needed a way to assess if the product was worth the cost. [8,9,12] Given that over 80% of ambulatory patients are not grossly dysfunctional, the previous use of simple functional classification schemes (e.g. New York Heart Association Functional Class) became inadequate as criterion for resource allocation. [12] A more discriminative tool was necessary. The typical tool used has been the "quality adjusted life year" (QALY) or "well-year". [8,12] This type of utility measure can be developed through means such as the "standard gamble" or "time trade off" techniques. The standard gamble asks patients to choose between their own health and a gamble of immediate death or achieving full health for the rest of their lives. The degree of gamble they are willing to accept is considered reflective of their current quality of life. [8] The greater the gamble, the poorer the quality of life. Alternatively, one can ask patients how many years of life at their current health status are they willing to trade for a shortened life span at full health. [8] The gamble accepted or time traded can be converted into a "well year" or "QALY". This measure is then combined with the

monetary cost of providing medical care and services designed to restore health. A dollar amount can then be attached to each "well year" or "QALY" gained. Many ethical concerns arise from formulating broad medical care policy based upon cost-utility analysis. [9] Many poorly defined components (particularly the estimated costs used in such analysis) subtract from the credibility of this approach.[9] Nonetheless, these types of studies are currently used for decision making, without regard to their soft scientific underpinnings. [9,12,13]

Regardless of final intent (clinical or economic maximization of treatment strategies), long term follow-up in clinical studies has increasingly identified health related quality of life and functional status as an important aspect of outcomes. The ability to accurately and reproducibly measure such a subjective entity as quality of life has greatly improved in recent years. [8,11] Health related quality of life measurement was previously confined to simple questionnaires and patient self-assessment. Currently, several formal and standardized instruments have evolved. [7,8,12,13] Various authors describe the health related quality of life questionnaires as instruments, questionnaires or tools. These terms will be used interchangeably in this text.

In the arena of congenital heart disease, past studies aimed specifically at measuring health related quality of life relied on broad and inconsistent determinants of "quality of life". Measures such as, education level, employment, marital status, insurability and New York Heart Association classification were the typical markers used.[14,15,16,17] The largest review of quality of life in adult congenital heart disease was undertaken by the Mayo Clinic, but even this was limited to only three specific defect types. [14] Conclusions from this cohort were based upon self-perception of general health, employment, education level, marriage status, and insurability. These investigators found no differences in these characteristics between their cohort and the general U.S. population. [14] Even if a difference had been found, it would be presumptuous and incorrect to

ascribe a lower "quality of life" by assessing education level or job classification. The pitfalls in using such nonspecific markers for quality of life are evident.

As noted above, the population with congenital heart disease is reaching adulthood and rapidly growing. Adequate data on which to base care plans for this population is lacking. [4] It has been suggested that ongoing observation for post-operative residua and sequela, as well as functional status will be essential in forming this much needed data base. [4] This brings one back to the primary question of how to best measure functional status in the congenital heart disease population.

It is evident from other studies that quality of life and laboratory measures of functional status do not necessarily correlate. [3,6] In looking at a population with pulmonary disease, Hajiro et al., found that pulmonary disease specific quality of life questionnaires (the St. George's Respiratory Questionnaire, the Breathing Problems Questionnaire, and the Chronic Respiratory Disease Questionnaire) had only weak correlations with maximal oxygen consumption and pulmonary function testing ($r=-0.24$ to -0.36).[3] Similar findings were reported by Guyatt et al. in patients with chronic lung and heart conditions. [6] These authors reported no statistically significant correlations ($r=0.1$ to 0.3) between maximal exercise time and four different functional assessment instruments (the Rand instrument, the Baseline Dyspnea Index, the Oxygen Cost Diagram, and the Specific Activity Scale). They did however find moderate correlations between these same questionnaires and a six-minute walk test ($r=0.47$ to 0.59).[6] This is notable, because a six minute walk test is examining something different from a maximal exercise test. Maximal exercise testing represents an assessment of complete cardiopulmonary function, including cardiopulmonary reserve. A walk test may never encroach upon the limits of cardiopulmonary reserve. From these two examples alone, it can be gleaned that in these chronically ill populations, patient reported functional status is different from objective measurements. Additionally,

all objective measures are not equal in their abilities to reflect patient self-perception of function.

Add to this the evidence that most individuals with congenital heart disease do not have normal exercise capacity, [15,18,19,20,21,22] severely limits the ability to extrapolate any prior finding to the unique adult congenital heart disease population. It is necessary, therefore, to assess standardized quality of life instruments specifically in the adult congenital heart disease population. Only after we have evaluated the validity of these tools in this population can quality of life assessment be incorporated into the data desperately needed for development of future treatment strategies.

The psychometric scale concept of measuring quality of life is based upon the principle of segmenting "life" into several different domains. The domains are typically such things as physical health, mental health, social function and role function. [13] There is variation among these domains from instrument to instrument. Some assessment tools also divide these domains into separate categories. Regardless of how the instrument is partitioned, it evaluates each domain by a series of questions, which are then scored. The domain scores can be viewed separately, in various combinations, or compiled for a global score depending on the specific instrument. General tools, as well as disease specific instruments have been created and tested. [8,25] For example, specific instruments exist for illnesses such as cancer, asthma, AIDS, and congestive heart failure. To date no formal instrument has yet been applied to or developed for congenital heart disease.

The primary evolution of functional status assessment by questionnaire has occurred in the last 50 years. [23] Only since the 1970's has the concept of utilizing domains or scales to reflect quality of life been developed. The psychometric techniques of scale construction spawned the entire field of health related quality of life measurement. [23] The largest, most well recognized,

milestone in development of such instruments was the Rand Corporation Health Insurance Experiment. [44] This project demonstrated the ability of scale constructed instruments to accurately and reliably measure changes in health status across a large population. [23] The Medical Outcomes Study took this approach and applied it to medical conditions. The hope here was to test if such scaled instruments would be able to tract health status and changes in status in an older, sicker population. [24] The success of the Medical Outcomes Study provided the opportunity to finally describe and quantify differences in functioning and perceived well-being for a wide array of illness and health states. Comparison across disease states and after interventions was now possible.

Several quality of life assessment tools exist, but a few specific instruments deserve highlighting. The Medical Outcomes Study (MOS) Short Form 36 (SF-36) and Short Form 12 (SF-12) have been extensively studied in a broad range of health and disease. These tools are widely available and frequently used in the literature. Both the SF-36 and SF-12 are gaining recognition as the standard for quality of life assessment. [23]

The MOS SF-36 and SF-12 consist of eight different domains: physical function, role-physical, bodily pain, vitality, social functioning, general health, role-emotional, and mental health. The physical function domain samples a range of mild to severe physical limitations and allows responses, which reflect severity of any specific limitations noted by the respondent. Beyond the actual limitation, the role physical domain evaluates the impact any health related limitation has on perceived quality of life. The contribution pain or discomfort has towards interference in daily life is assessed by the bodily pain domain. The vitality domain captures a subjective sensation of well being in terms of fatigue and energy level. This is in contrast to the general health domain, which assesses the subjective sensation of well being from an illness resistance and health outlook standpoint. Mental Health and role emotional domains are parallel to the physical function and role physical domains. The mental health domain defines

psychologic barriers to function, while role-emotional domain elucidates the extent to which those mental problems actually limit function. Finally, the social functioning domain extends beyond the individual to evaluate how the respondent is able to interact with others as a social being.

The majority of these domains reflect health status as the absence of limitation in function. Thus, an individual without any disability would achieve a maximal score in that domain. Maximum scores have been normalized to 100. Three of the domains (general health, vitality and mental health) are "neutral". These three domains allow for the subject to be either positive or negative in their response. A score of 50 reflects average, where as a score of 100 reflects a perception of better than average status (e.g. "full of energy all the time").

Testing of an instrument's reliability and validity is an ongoing process, which advances with every application of that test. All standardized quality of life instruments are evaluated for reliability and validity in essentially the same manner. The assessment endured by the SF-36 and SF-12 will be described in detail as an example of the general process.

Reliability and validity of the MOS SF-36 and 12 have been extensively evaluated. Reliability as judged by test-retest evaluation and "split half" testing is excellent. The "split half" method simply assumes that if all questions reflecting a domain are split in half, each half when scored separately should give the same final score. By taking an average of all possible "split half" reliability scores, and adjusting for the total number of questions, an alpha-coefficient (Cronbach alpha) can be determined. The mean reliability coefficients (averaged across fifteen separate studies) for all domains of the SF-36 are greater than 0.80 (indicating 80% reliability), with the exception of the social functioning domain, which has a mean value of 0.76. The physical function domain consistently has a reliability coefficient greater than 0.90. [23]

Validation of the MOS SF-36 and 12 has been carried out in several chronic illnesses as well as in health. [23,24,26,27] Validity of an instrument is a difficult concept to prove, as there is no "gold standard" of quality of life measurement against which any instrument can be compared. The field of psychologic testing has provided several approaches to establish validity while lacking a gold standard. These approaches include different conceptual constructs of validity. The major types of validity are content validity, criterion validity, and construct validity. [28] Overall validity of the SF-36 has been successfully supported by these approaches. [29]

Content validity assesses if the questions used to determine the domain truly reflect that domain. For example, the percent of respondents who can walk more than one block (question 3i of the SF-36) steadily decreases as the physical functioning score decreases. Other items such as dressing and bathing (question 3j of the SF-36) and vigorous activity (question 3a of the SF-36) also distribute as expected. This means that as the physical function score decrease, the likelihood that a respondent will report a need for assistance with dressing and bathing, or undertake no vigorous activity will increase. Content validity has been demonstrated for every domain of the SF-36. [23,29]

Criterion validity tests the domain score against a criterion, which is not part of the domain items. For instance, the ability to hold a paying job should segregate well with scores in the domain of physical function. This is indeed the case for the SF-36. There is an almost perfect rank ordering of the physical function domain score with reported disability. [23] Similar tests of criterion validity have been done for all domains of the SF-36.

Finally, a process of factor analysis has been used to assess *construct validity* in the SF-36. [29] Briefly, this involved taking all eight domains and using them as variables in a factor analysis algorithm to determine "principle components" (factors). When this process was carried out in the later development of the SF-

36, the domains loaded in the appropriate factors, as hypothesized. That is, those domains hypothesized to best describe physical characteristics of quality of life (physical function, role physical, and bodily pain) were highly correlated with a single factor ("loaded high"). The same held true for those domains hypothesized to best describe mental health (designated as a separate factor in the analysis). The three domains that were hypothesized to be shared by mental and physical aspects of quality of life (vitality, role-social and general health) displayed reasonably equal loading between these two factors (physical and mental). This type of analysis provides support that the general constructs used in development of the SF-36 (and subsequently SF-12) were indeed valid. [23] Factor analysis also lead the way for simplification of eight domains into two component summary scores.

For the SF-36 and SF-12, the Physical Component Summary (PCS) and Mental Component Summary (MCS) scores are the distilled assessment of quality of life into two variables. The PCS is composed of those domains, which had the highest loading values into the "physical factor" of the factor analysis process mentioned above. However, the three domains that demonstrate significant correlation with both factors ("physical" and "mental") were further separated by "orthogonal rotation" of the factors. Using this mathematical technique, the "general health" domain now correlated more strongly with the "physical" factor, while "vitality" and "social functioning" loaded in favor of the "mental" factor. Although there still exists a statistically significant correlation (greater than 0.35 factor coefficient) with the weaker loading factor, it was felt acceptable to assign these three domains in a binary fashion. In the final analysis, PCS is a summary score for physical functioning, role physical, bodily pain, and general health domains, while MCS reflects the mental health, role emotional, social functioning, and vitality domains. [29] The PCS and MCS both are normalized to a score of 50. The mean, standard deviation, and factor coefficient scores used to normalize the PCS and MCS came from the general US population. [29]

In an attempt to simplify the SF-36 without loss of information, the SF-12 was developed. [30] The SF-12 is a 12-question subset of the SF-36. The original 36 questions were analyzed individually for their predictive powers (weighted essentially) and then pared down to only those that were the most robust. This left 12 questions; two each for four of the domains, and only one each to represent the remaining four domains. Extensive correlational studies demonstrating the ability of the SF-12 to reproduce the results of the SF-36 have been done. [30] For the general US population, correlations of greater than r=0.95 have been found between the SF-36 and SF-12. [30]

Another instrument developed in near parallel with the MOS SF-36 was the Sickness Impact Profile (SIP). It is a more extensive (contains 136 questions) questionnaire, which too has been widely applied and validated in several different populations. [10,31] The development of the SIP was undertaken in 1972. The goal was to provide a measure of health care outcomes that would be sensitive to minimal levels of dysfunction and reliable in detecting any slight changes. [10] Adopting a parallel approach to the MOS, Bergner et al. hypothesized that a broad base of questions regarding daily activities would provide the intended measure. Over 300 questions were collected and grouped into categories. After a series of field tests on a variety of types and severity of illness, the final 136 questions were selected as the most robust in representing their category. These categories were then tested for reliability and validity in three specifically chosen populations. [10,31] These first tests were carried out in 45 subjects, equally distributed between patients with arthritis, hyperthyroidism, and recent hip replacements. These patient groups were selected because the clinicians involved felt reliable clinical measures, which reflected patient functional status, existed.

For the categories, test-retest reliability was excellent (r=0.92), as was internal consistency (r=0.94).[31] Further reliability testing obtained alpha-coefficients (Cronbach's) of 0.63 (Eating category) to 0.90 (Body care and movement

category) have been obtained.[31] Criterion validity (using the predefined clinical measure specific for each group) ranged from moderate to good (correlations of 0.4 to 0.8) across the three groups. Content and construct validity (as assessed by factor analysis) have also been accomplished for the SIP.[31] Much like the MOS SF-36, the finding of high intercategory correlations lead to the development of summary domains.

The SIP consists of three summary domains: physical, psychosocial and independent categories. The physical domain is composed of three categories: "ambulation", "mobility and body care" and "movement". The psychosocial domain utilizes the "communication", "alertness behavior", "emotional behavior" and "social interaction" categories. The independent categories are "sleep and rest", "eating", "work", "home management", and "recreation and pastimes". Each domain is scored on a zero to 100 scale, with zero being without any disability and 100 being completely disabled. The summary domain scores come from weighting and summing the positively endorsed responses with in the contributing categories.[31] The SIP is also one of the few instruments that can be collapsed into a single overall score attempting to reflect health related quality of life.

Quality of life assessment instruments, such as those listed above, are not the only means to assess functional status. In an attempt to focus on assessment of the physical aspects of functional status alone, the Duke Activity Status Index (DASI) was created. The DASI is the only instrument currently available that provides an estimate of maximal oxygen consumption (a surrogate of exercise ability). The DASI is limited to questions of activity behavior and allows for only positive or negative responses. Development consisted of asking 50 consecutive individuals undergoing clinically indicated maximal oxygen consumption testing to fill out an extensive questionnaire involving specific functional abilities. The results of the maximal oxygen consumption test were then used as the dependent variable for a regression equation involving the questions as

independent variables. Twelve questions fell out as predictive of maximal oxygen consumption. The questionnaire and its weighted regression equation was then tested prospectively on the next 50 consecutive patients. A correlation of 0.58 (Spearman's) between the predicted and actual maximal oxygen consumption was found in the validation group. [32] Unfortunately, the DASI has not been widely tested by other investigators.

As noted earlier, formal assessment of health related quality of life with a standardized instrument has not been demonstrated in the congenital heart disease population. If fact, there has not been an attempt to determine if the current instruments, including those listed above, are even valid in this very unique population.

In looking specifically at the physical domain of quality of life for individuals with congenital heart disease, rigorous validation of instruments is necessary. Prior studies relying upon patient self-report have demonstrated both over estimation [18] and under estimation [33] of performance. "*Exercise testing in adults with congenital heart disease, in general, has demonstrated reduced exercise capacity*" (Kaplan, et al. [19]), yet the vast majority of these same patients rate themselves as unimpaired. [19] This leaves great latitude in assessing the impact mild to moderate physical limitation has on health related quality of life. The ramifications of this decrement in physical ability upon health related quality of life has not yet been quantified.

As stated previously, functional status can be assessed by various means. Laboratory measures do offer some distinct advantages. Assessment of exercise capacity as a marker of physical ability is reproducible and well accepted. Maximal exercise testing with concomitant respiratory gas analysis provides a complete assessment of the subjects' physiologic state. [37] Cardiac and pulmonary status are paramount in this type of testing. The lungs' ability to exchange oxygen and carbon dioxide in conjunction with the heart's ability to

deliver that oxygen to working muscle are all evaluated. Additionally, the active muscles' efficiency in utilizing the delivered oxygen is incorporated into this type of testing. Not only are the maximal levels of cardiopulmonary and muscular function (and thus these systems' reserve) assessed, markers of less than maximal performance, such as anaerobic threshold, can also be defined by this type of exercise testing. These qualities render maximal exercise testing with oxygen consumption analysis an ideal measure of global physical capacity. As noted above, however, prior investigators have failed to find a strong correlation between maximal exercise capacity and reported functional status in chronic acquired illnesses. [3,6] For individuals with congenital heart disease, the relationship between laboratory measures of functional status and perceived physical aspects of quality of life remain untested. The study described here assesses this very relationship.

The population of individuals with surgically corrected or palliated congenital heart disease is new and growing. These patients are now reaching adulthood and long term outcomes and care strategies are being defined in the absence of extensive research. Future studies of congenital heart disease will require the ability to assess health-related quality of life. The goals of the study at hand is to provide criterion validation for the physical domains of the MOS SF-36, SF-12, and SIP using several objective measures of exercise capacity. Additionally, this study will assess the predictive ability of the DASI in the adult congenital heart disease population.

Should these instruments validate well in this population, they will be useful in future research assessing health related quality of life as an endpoint for developing care strategies in adult congenital heart disease. These instruments may ultimately prove to be useful in determining which reparative surgical procedures performed on infants provide the best outcomes once adulthood is reached.

Methods:

Study objectives:

The objectives of this study are two fold. The primary goal is to provide criterion validation for the physical domains of the MOS SF-36, SF-12, and SIP using objective measures of exercise capacity, namely peak oxygen consumption. The secondary goal of this study is to assess the predictive ability of the DASI in the adult congenital heart disease population.

Study design:

A cross-sectional descriptive assessment of both quality of life (by formal instruments) and objective exercise capacity (by maximal exercise testing) was carried out. The study population was obtained from a population based cohort and an active medical subspecialty clinic. These individuals underwent evaluation in a consecutive, prospective fashion.

Ethical considerations:

The letters of invitation, consent forms, questionnaires and study protocol all were reviewed and approved by the Oregon Health Sciences University human subjects Institutional Review Board.

Study population:

A total of 29 subjects were included in this study. Subject recruitment was from two sources over an eleven month period (May 1998 to April 1999). The primary source was an established population based registry of over 4200 individuals with surgically altered (corrected or palliated) congenital heart defects. This registry was established by Dr. C. Morris in 1982, and includes all Oregon residents who have had reparative surgery for one of 14 major heart defects.

[Table 1] The registry begins with the first year of congenital heart disease surgery performed in Oregon (1958). Enrollees underwent surgical intervention at less than 19 years of age. To form the registry, medical records departments of all Oregon hospitals that performed cardiac or thoracic surgery were asked to identify cases by using both procedure and diagnostic codes. Pertinent information was then extracted retrospectively from the medical charts. Long term follow up of multiple endpoints for subjects in the registry has been obtained every two years via mailed questionnaire. Subjects not responding to the mailed questionnaires are traced and contacted with a formatted phone interview. Current information is available for approximately 90% percent of the registry subjects and new surgical cases are ascertained and added yearly.

Table 1. Congenital Defects Included in Registry Population

Tetralogy of Fallot
Ventricular septal defect (VSD)
Atrial septal defect (ASD)
Coarctation of the aorta
Aortic valve stenosis
Pulmonary stenosis
Transposition of the great vessels
Patent ductus arteriosus
Partial atrioventricular canal
Complete atrioventricular canal
Pulmonary atresia
Pulmonary atresia with VSD
Truncus arteriosus
Total anomalous pulmonary venous return

From this registry, potential subjects were limited to the greater Portland Metropolitan area, Southwest Washington and Northwest Oregon. This limitation of subjects was necessary given reasonable expectations for voluntary subject travel to the research center. Confining the registry to this geographic area was accomplished by using telephone area codes representative of the regions of interest.

Additional exclusion criteria included age greater than 18 years of age (relieving the need for parental consent, and ensuring that subjects had had their surgical

correction or palliation several years prior.) and absence of mental disability. Mental disability was a necessary exclusion because the quality of life assessment tools used in this study have not been validated in age groups less than 14 years old, and require the ability to read, as they are self-administered.

Since congenital cardiac anomalies occur with different incidence rates, their representation with in the database is weighted toward the more common defects. To avoid over representation of any single defect in the study group, random selection stratified by the fourteen defect categories was necessary. Among the individuals in the database fitting all inclusion criteria, ten potential subjects from each defect category were randomly selected. This selection method was used to ensure a balanced subject group. Even utilizing this strategy, five of the fourteen defect categories had less than ten potential subjects over the age of 18 years living with in the defined geographic area. These represented the most rare congenital defects [Table 2].

Table 2. Potential Subjects by Congenital Defects	**Number Eligible for Random Selection**
Atrial septal defect (ASD)	139
Patent ductus arteriosus	121
Coarctation of the aorta	112
Tetralogy of Fallot	102
Ventricular septal defect (VSD)	72
Pulmonary stenosis	68
Aortic valve stenosis	45
Transposition of the great vessels	17
Partial atrioventricular canal	17
Complete atrioventricular canal	4
Pulmonary atresia with VSD	4
Total anomalous pulmonary venous return	3
Pulmonary atresia	1
Truncus arteriosus	1
	706 Total

A total of 103 letters of invitation were sent out. An initial response rate of 14% was achieved. A second round of invitational letters were mailed out to the same individuals approximately two months after the first letter. The second mailing obtained an overall response rate of 23%.

To evaluate the generalizability of our population to the greater population with congenital heart defects, the baseline quality of life surveys (MOS SF-12) of all registry subjects were utilized. The concern of a response bias in our study group was evaluated by comparing the quality of life scores for the registry to the quality of life scores for the participants in this project.

The preeminent goal of this recruitment strategy was to obtain a diverse study population, which spanned individuals with corrected minor cardiac defects to those with severe cardiac compromise. The population based cohort has the potential for survival bias in those over age 18 and consists predominantly of common, repairable defects. This raises the possibility that the registry population may be biased towards a higher functioning status. Therefore, in an attempt to achieve the stated goal, a second population was drawn upon to enhance the defect variety and illness severity of the study population. By inviting patients from the Adult Congenital Heart Disease (ACHD) Clinic at Oregon Health Sciences University, subjects with active medical concerns related to their cardiac condition were included. The ACHD Clinic also provided individuals with uncorrected or complex congenital malformations not present in the 14 categories of the registry. Invitation to this protocol from the ACHD Clinic was accomplished by giving scheduled patients an invitational letter upon arrival for their appointment. Additionally, patients were asked to fill out the same baseline demographic and quality of life survey used in the registry population. All baseline surveys were collected regardless of willingness to participate in the complete protocol. The participants from the ACHD Clinic had no overlap with those from the registry who had received letters of invitation.

A total of 29 subjects ultimately completed the entire protocol, five of whom came from the ACHD clinic.

Quality of life assessment:

Subjects were aware only that they would be asked to fill out a series of questionnaires and then be asked to exercise. They did not know that any attempt to relate the questionnaires to their exercise was to be done. All questionnaires were completed prior to exercise as to not influence any answers about perceived physical abilities.

After informed consent was given, subjects self-administered the battery of quality of life instruments. These were the MOS SF-36, SF-12, SIP and DASI. Verbal clarification of instructions for completion of the questionnaires was minimized. All subjects were referred to the written instruction provided. Study personnel did this to minimize any potential influence on answers and to ensure standardization from subject to subject.

Physical Exam:

Once the questionnaires were complete, subjects underwent a brief physical exam followed by exercise testing.

A brief physical exam was conducted on every subject by a single investigator (M.C.) to assess for potential contraindications to vigorous exercise. Findings of unstable vital signs (heart rate or blood pressure), overt heart failure, severe valvular stenosis, or musculoskeletal abnormalities prohibiting running would have excluded participation. No subject was declined due to physical exam findings.

Exercise Testing:

Prior to conducting exercise and respiratory gas analysis, every subject underwent spirometry and maximum minute ventilation testing. This was done to

evaluate for potential coexisting pulmonary processes that could impact exercise abilities. Forced expiratory volume in one second (FEV1), functional vital capacity (FVC) and maximum voluntary ventilation (MVV) were measured. This was accomplished using the pneumotach and flow analysis components of the Medgraphics metabolic cart and "BreezEx" 3.0 software package. The pneumotach was calibrated with a three liter (3 L) syringe before every test. Only variability of less than 3% was accepted for calibration.

Exercise testing consisted of standardized exercise treadmill protocols and respiratory gas exchange analysis. [35,36] Available protocols were the Bruce protocol, modified Bruce protocol, and Naughton protocol. Protocols of varying intensity were used to ensure that at least six to eight minutes of exercise was obtained. Completing this duration of exercise was important, as it allows subjects to give adequate effort in obtaining their peak oxygen consumption.[35,36,37] A Bruce protocol would have been too aggressive for many of the subjects and resulted in early fatigue and failure to reach anaerobic threshold. Selection of an appropriate exercise protocol allowed achievement of anaerobic threshold in 90% of the subjects.

During treadmill exercise, all subjects were monitored with continuous electrocardiography to determine if abnormal cardiac rhythms or evidence of ischemia developed. Noninvasive blood pressure monitoring using a cuff at the brachial artery was done every two minutes during exercise and during recovery to ensure hemodynamic stability and capture peak blood pressure response to exercise.

Respiratory gas analysis was carried out to determine oxygen consumption and carbon dioxide production at baseline as well as throughout exercise duration. These measurements were accomplished using the Medgraphics metabolic cart and "BreezeEx 3.0" software package. The apparatus was calibrated prior to every test session. This process included calibration of O_2 and CO_2 analysis

modules against known reference gases, zeroing and calibrating the pneumotach (as described above), as well as changing the drying cartridge for the gas sampling line. Additionally, values for current temperature, barometric pressure and relative humidity were entered prior to every test. These values were obtained from a local meteorologic source. Subjects were fitted with either a mask that included the nose or mouth piece and nose clips (for those with beards) to ensured complete gas collection. All masks were tested for fit and air leak prior to gas analysis and exercise.

Multiple variables were collected during the exercise portion of the protocol assessing peak exercise performance, exercise work, and anaerobic threshold. [Table 3] It should be noted that the ability of ventilatory anaerobic threshold (VAT, as determined by the V-slope method [38]) to accurately represent true anaerobic threshold (measured by blood lactate levels) has been confirmed elsewhere. [39]

Table 3 Exercise Related Variables Obtained During Study

Variable	Symbol	Explanation/units
Peak oxygen consumption	$VO2_{peak}$	Milliliter/min/kg
$\%VO2_{max}$ of predicted normals	%VO2	The % of O2 consumption achieved compared to normal predicted value
Ventilatory Anaerobic threshold, absolute	VAT	Determined by V-slope method, Milliliter/min/kg
Anaerobic threshold, as percent of $VO2_{peak}$	VAT%	Percent of $VO2_{peak}$ (no units) at which VAT occurred
Respiratory exchange ratio	RER	VCO2/VO2 (no units)
Metabolic equivalents	METS	METS is the unit of measure
Work performed	Work	Measured in Watts
Maximum heart rate	HR_{max}	Beats per min (BPM)
Maximum systolic blood pressure	SBP_{max}	Millimeters of mercury
Forced expiratory volume in one second	FEV1	Liters
Functional vital capacity	FVC	Liters
Maximum voluntary ventilation	MVV	Liters/min

Computer algorithms incorporated in the software package supplied predicted normal values for the spirometry and exercise variables. Many variables could then be expressed in terms of percentage of predicted normal values. The references for the algorithms from which predicted normal values were calculated (based on age, weight, height and gender) are provided in Appendix A.

Data handling:

To avoid potential human recording error, all demographic data, SF-36, SF-12, and DASI questionnaires were scanned directly into the computer database. Once in the computer, the SF-36 and SF-12 raw data could be checked and scored. Any questions left blank were left as missing data, which was then accounted for by the scoring algorithm. This adjustment involved determining an interim score for the domain using the other questions pertaining to that domain. The missing item was then coded using the mean of the other responses given for that domain. A domain was not scored if more than 50% of the responses were missing. [23] This scoring strategy was developed and tested by the creators of the SF-36 and SF-12. In our study, only one question was left unanswered.

If multiple responses were endorsed for a single question, the lowest score (reflecting the greatest impairment) endorsed was chosen during the verification process. This approach was established prior to the protocol initiation to ensure a conservative estimate of quality of life. The alternative strategy (accepting the higher score) would have potentially over estimated health related quality of life scores. This strategy was applicable to all questionnaires, except the SIP. It was applied in two instances only: one response in the DASI, and one response in the SF-36. These instances occurred in separate patients.

Once DASI responses were scanned into the database, estimation of maximal oxygen consumption from those scores could be accomplished without human

computation. The activity weights and the regression equation are shown in Appendix C. Note that the weighting is not reproduced on the actual administered questionnaire. Subjects would potentially be biased by those weights if they were to appear on the form.

The SIP did not lend itself to automated scoring, but required hand computation with computer aggregation.

Computerized databases were password protected, and hard copies of exercise data and original questionnaires were stored in locking file cabinets.

Sample Size Estimation:

Prior investigators have used correlation coefficients of 0.6 to indicate a significant positive correlation between health related quality of life instrument scores and objective testing. [23,25,29,30,32] For this protocol, we hoped to achieve a similar degree of correlation. At an alpha (2-tailed) of 0.05 and power of 0.90. (Beta = 0.10), a minimum sample size of 25 subjects was calculated as necessary to demonstrate this level, or greater, of correlation between the continuous scale variables. This was derived from the following equation:

$$(N-3)^{1/2}[1/2 \ln ((1+r)/(1-r))] = Z_{alpha} + Z_{beta}$$ [34]

Analysis:

The primary outcome variable was the objectively derived peak oxygen consumption ($VO2_{peak}$). Peak oxygen consumption was chosen as the primary objective measurement as it is a classically accepted value which represents total cardiopulmonary capacity. [37] VAT, Work and metabolic equivalents (METS) achieved were considered secondary outcome variables. These variables were chosen as they too are classic measures of cardiopulmonary fitness. [37] $VO2_{peak}$ was evaluated as percent of predicted normal value and absolute value, while Work, METS and VAT were assessed as absolute values only. Other endpoints

are the numeric scores derived from the health related quality of life questionnaires.

All endpoints, objective measures and quality of life scores are of continuous scale. Exercise measures all can theoretically range from a value of zero to infinity. For the quality of life measures, the PCS score for the SF-36 and SF-12 have been normalized to a value of 50, but scoring does allow for both higher and lower values. The SIP score is a defined range from zero to one hundred, allowing for values between integers as well. Additionally, exercise measures and quality of life scores are known to be normally distributed in the general population. These characteristics suggest that the endpoints measured lend themselves to analysis by correlational evaluation.

Accordingly, each objective variable was correlated separately against the physical ability score obtained from each of the health related quality of life assessment questionnaires completed. Simple Pearson's correlation coefficients were used to evaluate these relationships. Since four different questionnaires were tested for significant correlations, the individual p-values considered significant was set more stringently at a level of 0.01. Using this level of significance and the following equation, an overall p-value of 0.05 can be calculated.

$$p^* = 1-(1-p)^n, \text{ where } p^* \text{ is overall p-value}$$ [45]

Finally, in an attempt to better assess the various contributors to prediction of peak VO2, a multiple linear regression analysis was undertaken. This regression model included not only PCS (of either the SF-12 or SF-36), but also variables such as age, sex, NYHA classification, pulmonary function, and MCS. A regression model to assess the contribution $VO2_{peak}$ makes to PCS36 was also constructed.

Results:

Study group characteristics:

The study population was evenly split by gender (M:F = 15:14), and had an age range of 27 years (from 18 to 45 years old). [Table 4] All but one subject had had corrective surgery. That subject had undergone a palliative procedure in infancy and was later deemed inoperable for complete correction of the congenital defect. This was the only cyanotic individual. All participants were free of clinical heart failure and severe valvular stenosis by physical exam. One patient was under treatment for clinical depression, otherwise no mental health concern were present. Finally, only one subject developed cardiac arrhythmias during exercise testing. Fortunately, this occurred at peak exercise and did not require early termination of the test. NYHA classification and medication usage are listed in Table 4.

Table 4 Baseline Characteristics of Study Group (N=29)

Characteristic	n (%)
Age	
18 - 25	9 (31.0)
26 - 30	3 (10.3)
31 - 35	6 (20.7)
36 - 45	11 (37.9)
Drugs	
Antiarrhythmic	0 (0)
Beta blocker	1 (3.4)
Digoxin	3 (10.3)
Diuretic	3 (10.3)
Warfarin	4 (13.8)
ACE Inhibitor	5 (17.2)
Any Drug	9 (31.0)
NYHA Classification [40]	
I	21 (72.4)
II	7 (24.1)
III	0 (0.0)
IV	1 (3.4)
Routine exercise training	4 (13.8%)
Cyanotic (O_2 saturation < 85%)	1 (3.4%)

Additionally, subjects were well distributed across the congenital defects represented in the registry cohort. [Table 5]

Table 5 Subjects by Congenital Defects	n	% of total group
Tetralogy of Fallot	4	13.8
Coarctation of the aorta	4	13.8
Complete atrioventricular canal	4	13.8
Transposition of the great vessels	3	10.3
Partial atrioventricular canal	3	10.3
Atrial septal defect (ASD)	2	6.9
Patent ductus arteriosus	2	6.9
Ventricular septal defect (VSD)	1	3.4
Aortic valve stenosis	1	3.4
Pulmonary stenosis	1	3.4
Congenitally corrected transposition	1	3.4
Pulmonary atresia	1	3.4
Tricuspid atresia	1	3.4
Complex congenital defect	1	3.4
Total anomalous pulmonary venous return	0	0
Truncus arteriosus	0	0
Total	29	100%

Study variables:

Quality of life measures:

Quality of life scores are listed in Table 6. A few key characteristics should be highlighted. The PCS12 scores of the group that participated in this study did not differ from the entire registry population. This point is salient, because it helps to eliminate concern that the volunteers had a lower or higher perceived physical quality of life than those that declined or were not invited. Having equality of PSC12 scores supports the supposition that the study group was representative of the entire ACHD registry population in regards to their perceived physical quality of life. Additionally, the mean SF-12 and SF-36 physical domain summary scores (PCS) were no different than norms for the general US population. [30] [Table 6]

Similarly, using a prior study estimate, it can be seen that the SIP Physical Domain scores for the study group also did not deviate from the general US population either. [31] [Table 6]

Table 6 Subject Physical Quality of Life Questionnaire Results

	Score (mean ± SD)				
Instrument	Study group	Registry Cohort	p [8]	General US[30,31]	p [#]
SF36 - PCS36	51.3 ± 7.8			50.1 ± 9.9	0.41
SF12 - PCS12	51.4 ± 6.5	52.8 ± 7.5	0.95	50.1 ± 9.5	0.27
SIP Physical Domain	1.2 ± 1.9			1.3 ± 4.8	0.81
DASI Predicted $VO2_{max}$	31.5 ± 4.7 ml/kg/min				

* p = 2-tailed significance for Study group versus Registry
\# p = 2-tailed significance for Study group versus General US population

For completeness, it should be noted that the mental component scores of the SF-36, SF-12, and SIP for the subject group were also no different from the general US population.[30,31] [Table 7] The SIP total summary score was also no different between the study group and the general US population. [31] [Table 7]

Table 7 Subject Mental Quality of Life Questionnaire Results

	Score (mean ± SD)		
Instrument	Study group	General US[30,31]	p [#]
SF36 - MCS36	49.5 ± 11.1	50.0 ± 10.8	0.81
SF12 - MCS12	50.2 ± 9.5	50.0 ± 9.6	0.91
SIP Mental Domain	4.8 ± 10.7	1.3 ± 4.8	0.67
SIP Total Score	3.6 ± 5.9	2.5 ± 6.5	0.31

\# p = 2-tailed significance for Study group versus General US population

The fact that the SF-12 is a subset of the SF-36 allows for testing of intra-subject reliability. The correlation between the PSC12 and PCS36 for the study group was excellent ($r=0.86$, $p<0.001$), as was that between the MSC12 and MSC36 ($r=0.96$, $p<0.001$).

Exercise measures:

Exercise endpoints are presented in Table 8. The primary measure of exercise capacity for this study was $VO2_{peak}$. The mean $VO2_{peak}$ for the group was only 80% of the predicted normal value. This demonstrates a measurable decrement in the exercise capacity of adults with congenital heart disease, even after complete correction. $VO2_{peak}$ is, however, effort dependent and accurate assessment relies on a vigorous attempt at exercise by the study subject. Maximal oxygen consumption is the point at which metabolic oxygen consumption actually plateaus and can not increase further. $VO2_{peak}$ represents the most oxygen these subjects were able to utilize. Although $VO2_{peak}$ is not technically maximal, it is likely near the true maximal level. This assumption can be readily made, because on average the study cohort gave a good exercise effort. In support of this assertion, all but three subjects achieved anaerobic threshold and the group achieved a high double product (HR_{max} x SBP_{max}). [Table 8]

Table 8 Subject Exercise Results

Exercise Variable	Mean ± SD	Median	Range
HR_{max}	163.2 ± 22.1 bpm	168	112-195
SBP_{max}	153.1 ± 21.5 Torr	150	110-210
RER_{max}	1.2 ± 0.1	1.2	0.95-1.4
Double product	25,057 ± 5117	25,600	13440-33540
Percent of predicted normal FEV1	85.7 ± 17.2 % (N=28)	83.5	49-123
Percent of predicted normal FVC	85.1 ± 15.6 % (N=28)	85	49-123
Percent of predicted normal MVV	93.0 ± 20.3 % (N=28)	90.5	56-142
$VO2_{peak}$	27.1 ± 10.1 ml/kg/min	25.7	9.6-47.4
Percent of predicted normal $VO2_{max}$	80.0 ± 24.9%	78.8	27-137.2
VAT	14.8 ± 5.0 ml/kg/min (N=26)	15.1	4.4-27.4
VAT%	52.0 ± 10.6%	53.4	31.9-75.2
Work	204.8 ± 85.2 Watts	202	23-418
METS	8.0 ± 2.8	8.0	2.7-13.5

Additionally, no subject had limiting pulmonary abnormalities on spirometry testing prior to exercise* and only four subjects reported participation in a routine exercise program of more than 30 minutes, three times per week. The results of exercise endpoints should, therefore, not be skewed by the impact of poor effort, exercise training or pulmonary disease.

Additionally, there was a trend for decreased exercise capacity in those individuals requiring a modified Bruce or Naughton protocol, as compared to a full Bruce protocol. This did not reach statistical significance. This finding should be expected. Differing protocols were used based on pre-test estimated subject abilities. The modified Bruce and Naughton are less physically demanding and were used specifically for this reason. Despite their attenuated physical demands, these gentler protocols still allow for maximal exertion like a full Bruce protocol. The use of differing protocol should, therefore, also not bias the above results.

Correlations:

A Pearson's correlation was calculated comparing $VO2_{peak}$ against PCS36 ($r = 0.53$, $p=0.004$). [Figure 1] Testing of $VO2_{peak}$ against PCS12 yielded similar results ($r=0.49$, $p=0.006$). [Figure 2] When $VO2_{peak}$ and the SIP Physical score are correlated, a significant relationship is again found ($r= -0.50$, $p=0.006$). [Figure 3] (Note the correlation is negative because the SIP scores complete health as zero (0), and complete infirmity as 100.) [Table 9]

Despite the slightly larger r value, the PCS36 can not be considered a better predictor of $VO2_{peak}$ than the PCS12 or SIP Physical Score. A Fisher's Z transformation can be used to assess for significant difference between correlation coefficients.[41] Using this statistic (at alpha = 0.01, $Z_{0.01/2}=2.576$), it is

* One subject was unable to complete spirometry.

evident that no significant difference between these determined Pearson's correlations exists ($Z=0.17 < 2.576$).

Additionally, actual $VO2_{peak}$ correlated significantly with the DASI predicted $VO2_{max}$ ($r=0.60$, $p=0.001$). [Figure 4] [Table 9] It is interesting to note that the correlation between the DASI predicted $VO2_{max}$ and $VO2_{peak}$ found in this ACHD group was as good a correlation as the DASI developers found in their own validation sample (which was $r=0.58$). [32]

Despite the theoretic advantage of anaerobic threshold as an effort independent marker of physical ability, there were no significant correlations found between VAT and any physical quality of life measures (PCS12, PCS36, SIP Physcial). [Figures 5,6 and 7] [Table 9]

As an alternative measure to $VO2_{peak}$, maximal exercise capacity can also be judged by work performed and metabolic equivalents achieved. These were assessed for correlation with the physical domains of the quality of life measures. Again, a significant correlation was found. [Figures 9-13] [Table 9] This result was predictable as there is significant colinearity between $VO2_{peak}$ and both Work and METS ($r=0.69$ for $VO2_{peak}$ and Work, and $r=0.87$ for $VO2_{peak}$ and METS, $p<0.001$ for both). All objective measures of maximal exercise demonstrate approximately the same degree of correlation with the physical domains assessed. Based on the degree of correlation and R squared values (see Figures), these exercise endpoints account for 25 to 36% of the variability seen in the physical domain scores of the instruments.

Table 9 Correlations of Exercise Measures with Quality of Life Scores

	PCS36	PCS12	SIP Physical	DASI Predicted VO2
$VO2_{peak}$	0.53 (0.004)	0.49 (0.006)	-0.50 (0.006)	0.60 (0.001)
VAT	0.39 (0.05)	0.28 (0.16)	-0.34 (0.09)	
Work (watts)	0.60 (0.001)	0.57 (0.001)	-0.55 (0.002)	
METS	0.56 (0.001)	0.57 (0.002)	-0.57 (0.001)	

2-tailed significance provided beneath correlation (p=)

Secondary Analysis:

To assess if individuals with marked reduction in exercise capacity might be more likely to perceive and report limitations than those with minimal deficits, a subgroup was analyzed. Subjects with $VO2_{peak}$ measures less than 85% of normal predicted values were selected (n=18). This subgroup had a mean $VO2_{peak}$, which was 64.9% of predicted normal (SD=6.1%). For this subset, the PCS36 correlation with $VO2_{peak}$ had a larger r value ($r=0.78$, $p<0.001$) than that of the total study population. The correlation with PCS12 was similarly increased ($r=0.63$, $p=0.005$). Despite these larger r values, when tested with the Fisher Z transformation [41], they are not significantly different (at a 2 tailed alpha=0.01 level, $Z_{0.01/2}=2.576$) from those correlations based on the entire study group ($Z=1.36 < 2.576$).

Stratification by NYHA Classification [Appendix B] segregates the study group by *perceived* functional ability, rather than measured functional ability. When the correlations of $VO2_{peak}$ with quality of life scores are obtained for NYHA Class I and Class II separately, it can be seen that Class II individuals demonstrate a much greater correlation. [Table 10]

Table 10 Correlations of $VO2_{peak}$ with Quality of Life Measures By NYHA Class

	PCS36	PCS12	SIP Physical	DASI predicted VO2
Class I (n=21)	0.16 (0.49)	0.12 (0.62)	-0.02 (0.97)	0.25 (0.29)
Class II (n=7)	0.81 (0.03)	0.83 (0.02)	-0.69 (0.09)	0.93 (0.002)

2-tailed significance provided beneath correlation (p=)

Dramatic differences between r values and significance are shown in Table10. Despite these differences, the correlations between $VO2_{peak}$ and the physical domain scores are actually no different between NYHA Classes (at a 2 tailed alpha=0.01 level, $Z_{0.01/2}$=2.576) when subjected to the Fisher Z transformation [41] test (Z_{PCS36}=1.73, Z_{PCS12}=1.93, $Z_{SIP\ Physical}$=1.48). However, correlations between $VO2_{peak}$ and the DASI predicted VO2 are indeed different between the two functional classes. (Z_{DASI} =2.58 > $Z_{0.01/2}$=2.576). This indicates that the DASI is better at predicting oxygen consumption in physically compromised individuals.

When only female subjects are considered, there are no significant correlations ($p<0.01$) between $VO2_{peak}$ and quality of life measures. However, if only male subjects are evaluated, the $VO2_{peak}$ correlates with PCS36, and the SIP at a significance of $p<0.01$. [Table 11]

Table 11 Correlations of $VO2_{peak}$ with Quality of Life Measures By Gender

	PCS36	PCS12	SIP Physical	DASI predicted VO2
Females (n=14)	0.42 (0.15)	0.59 (0.03)	-0.47 (0.09)	0.48 (0.08)
Males (n=15)	0.66 (0.007)	0.55 (0.04)	-0.66 (0.008)	0.72 (0.002)

2-tailed significance provided beneath correlation (p=)

Despite the apparent difference in significant correlations between males and females, there fails to be a difference between the absolute values of the correlation when subjected to the Fisher Z transformation.[41] Neither the

significant correlation for males with PCS36 or DASI predicted VO2 is different from the nonsignificant correlation found for the female subjects (Z=0.83 for PCS36 correlations and Z=0.92 for DASI correlations).

Age may also affect the correlation between these instruments and objective endpoints. The median age of the study population, age 33, was selected as the cut off for this analysis. For those 33 years of age or older, there is a trend for all physical quality of life measures to be decreased compared to those younger than 33 years. [Table 12] These differences are, however, not statistically significant (p>0.50 by Student t test).

Table 12 Quality of Life Score Questionnaire Results by Age Group

Instrument	Score (mean ± SD)	
	Age =>33 years	Age<33 years
SF36 - PCS36	48.5 ± 8.9	54.5 ± 4.9
SF12 - PCS12	49.7 ± 7.3	53.2 ± 5.1
SIP Physical Domain	1.4 ± 2.3	0.9 ± 1.2

In looking at correlations between $VO2_{peak}$ and the physical aspects of the instruments, the younger group showed no significant correlation. The older subset, however, demonstrated correlations that reached significance at the 0.01 level, except that with PCS12, which was still significant at the 0.05 level. [Table 13]

Table 13 Correlations of $VO2_{peak}$ with Quality of Life Measures By Age

	PCS36	PCS12	SIP Physical	DASI predicted VO2
Age >= 33 years (n=15)	0.69 (0.005)	0.60 (0.02)	-0.69 (0.004)	0.75 (0.001)
Age < 33 (n=14)	0.17 (0.58)	0.34 (0.23)	0.10 (0.72)	0.31 (0.28)

2-tailed significance provided beneath correlation (p=)

Again, despite this trend identified when the study population is split by age, these differences in correlation between subgroups fail to be significant when

tested. The Fisher Z transformation [41] scores for the PCS36, SIP Physical and DASI predicted VO2 correlations all are nonsignificant (Z_{PCS36}=1.6, $Z_{SIP\ Physical}$=2.3, $Z_{DASI\ Predicted\ VO2}$=1.6).

As shown above (Table 9, Figure 1), the PCS36 only accounts for 28% of the variance seen in peak VO2. In an attempt to further define the contributors to $VO2_{peak}$, a linear multiple regression analysis was done. PCS36, MCS36, age, NYHA class, gender, and percent of predicted normal values for pulmonary function measures (FEV1, FVC and MVV) were used as independent variables. These variables were chosen for inclusion because these demographic descriptors displayed trends in segregating the study group in the secondary analyses above. Additionally, one's mental state (as measured by MSC) could be conceivably effect exercise performance and pulmonary function can certainly impact exercise capacity. Thus these variables were included as independent factors in the model. $VO2_{peak}$ was the dependent variable. This yielded a model where only 61% of the variance in $VO2_{peak}$ could be accounted for by all these factors. [Appendix H] When the values obtained from the SF-12 for PCS and MCS are used in place of those from the SF-36, the variance accounted for in $VO2_{peak}$ drops to 52%. [Appendix H] Additionally, when PCS36 is used as the dependent variable and $VO2_{peak}$ is included as one of the independent variables, it can be seen that this constellation of data accounts for only 46% of the variance in PCS36. [Appendix H]

Discussion:

This study confirms the findings of others. [12,19,21] For those with congenital heart disease, patient self report of physical ability does not accurately reflect maximal exercise capacity. This unique population, on average, over estimates their functional abilities. [12,19,21]

Despite the fact that this study provides evidence of a significant moderate correlation between objective endpoints and quality of life assessment tools, this

correlation only accounts for approximately one half of the variance seen in physical quality of life measurement. For adults, who have lived with corrected or palliated congenital heart disease, the standard quality of life assessment tools used in this study do not completely capture maximal exercise abilities in their measurement of physical functional status.

In attempts to account for the above findings, one can consider a number of explanations. The most obvious is that physical "quality of life" does not depend on maximal exercise ability. [6] There exist two potential pathways to arrive at this conclusion. Either subjects do not perceive maximal exercise capacity as important in their physical quality of life, or the instruments used here do not measure maximal abilities. It should be stressed that only the SF-36 and DASI have questions that ask specifically about *vigorous* activity.

In the modern era, few if any, routine activities depend on maximal physical ability. Some vigorous recreational activities potentially dip into this physical reserve, but few individuals challenge themselves to this level. An alternative measure of physical capacity that normally occurs at 40 - 60% of peak oxygen consumption is anaerobic threshold.[38] For this study group, anaerobic threshold (measured as VAT) occurred at 52 % of $VO2_{peak}$. Anaerobic threshold may, therefore, better reflect the levels of exertion reached in routine daily life.

Additionally, VAT was assessed in this protocol as an effort *independent* measure of exercise ability, where as $VO2_{peak}$ is an effort *dependent* measure. The more physical fit an individual, the longer they will be able to maintain aerobic metabolism, and thus extend their anaerobic threshold.[38] Despite these two theoretical advantages of VAT (effort independence, and potentially better representation of a threshold routinely reached in daily life), no correlation of VAT with any of the physical quality of life scores was demonstrated.

Potential insight into the results of the study at hand may come from the analysis stratified by NYHA Classification. The instruments tested here may not be sensitive enough to pick up limitations in individuals who perceive themselves as completely functional (Class I). Rather, once a subject is mildly hindered (Class II), the instruments' performance improves. This would reflect the discriminative ability of the instrument. More gross deficits are revealed, while the subtler dysfunction is missed. The trend seen for dramatic improvement in correlation between quality of life measures and $VO2_{peak}$ as functional class declines is difficult to ignore. [Table 10] Despite the lack of statistical significance between the different correlations stratified by NYHA Class, one suspects this trend would be borne out if a greater number of subjects were analyzed. This suspicion is strengthened by the statistically significant difference NYHA Class has on correlation between $VO2_{peak}$ and DASI predicted VO2 consumption. [Table 10]

If physical quality of life depends on ones ability to simply care for oneself, do typical chores, and only intermittent moderate physical activity, the rigorous endpoints used in this study were perhaps, in retrospect, wrongly chosen. A simple six minute walk test (distance covered by walking for six minutes) or strength testing may be more reflective of the characteristics of physical health related to perceived quality of life. [3,6]

Alternatively, perhaps the concept of physical aspects of quality of life is different for a population that has been limited since birth. Those individuals who have never known normal maximal abilities may perceive physical capacity very differently than those who have acquired such limitations later in life. If a child is physically limited, it is easy to speculate that other interests are likely to develop. Non-exercise oriented activities may become a prominent focus. The perceived quality of life is still quite good (even that specific to physical abilities), as evidenced by the scores for this group that match the general US population. All is perceived as "normal", despite the measurable deficit in maximal physical ability.

Regardless of how one attempts to explain the results found here, it is apparent that for clinical purposes the current standard quality of life questionnaires have limited utility in estimating the maximal exercise capacity (measured as either $VO2_{peak}$, Work, or METS), or anaerobic threshold of an individual ACHD patient. An incongruity exists: either maximal exercise capacity is not important for physical quality of life or the instruments tested here do not fully assess maximal exercise ability. Even the DASI, which was specifically designed to estimate maximal oxygen consumption, has only a modest correlation with measured oxygen consumption. Such quality of life surveys would be inadequate if exercise endpoints were necessary to determine employability or safety for potential pregnancy or surgery. If it is important to know cardiovascular reserve for clinical indications, these endpoints must be measured directly. An informal or even formal assessment using a questionnaire would not be adequate.

In hopes that questionnaire data plus additional clinical variables would better predict $VO2_{peak}$, several linear regression models were tested. Variables reflecting typical demographic data (age, sex, NHYA class) and pulmonary function substantially added to the quality of life measures' ability to predict $VO2_{peak}$. If, however, the PCS and MCS quality of life variables are removed from the model completely, the clinical variables alone account for 48% of the variance seen in $VO2_{peak}$. These findings are consistent with the correlations found in the primary analysis. Only a modest portion of peak oxygen consumption can be predicted by quality of life questionnaires, even when physical and mental assessments are included.

Furthermore, when a linear multiple regression model is constructed switching PCS36 to the dependent variable and including $VO2_{peak}$ with the independent variables, only 46% of the variance in PCS36 can be predicted. This would imply that variables other than maximal exercise capacity, lung function, and simple demographic data contribute to the physical aspects of quality of life. Again, it is

apparent that for clinical purposes the current standard quality of life questionnaires have limited utility in estimating the maximal exercise capacity, and maximal exercise capacity does not adequately describe the physical domains of quality of life.

However, for larger cohort evaluation and tracking, the physical domains of quality of life instruments may provide some insight into physical abilities. The moderate correlations found in this study provide validation that the quality of life surveys do reflect a measurable physical endpoint. It remains unknown, however, if these instruments can assess *change* in physical abilities for this unique population. This would require a test - retest model to assess for sensitivity to change in health status. To establish an instrument's sensitivity to a change in health status, baseline and follow-up scores would need by obtained over a time period where objective measures of health stability or change could be obtained for comparison. The current study provides only a cross-sectional, single point in time evaluation.

In an attempt to improve questionnaire sensitivity, the potential to develop a quality of life questionnaire specific for congenital heart disease exists. It would be essential, however, to determine the goal for such a specific questionnaire. If one wishes an estimate of maximal oxygen consumption, something akin to the DASI would be most appropriate. The physical aspects pertinent to quality of life are obviously much more complex than simple maximal exercise capacity. Any questionnaire specific for an ACHD population would first have to probe deeper into what truly constitutes physical quality of life.

Limitations:

Although chosen *a priori*, the strategy of conservatively estimating heath related quality of life by accepting the lower scoring response in cases of multiple responses could have inflated calculated correlations. (See Data handling

section.) This potential problem did not manifest itself in reality, however, because it occurred on two separate instances only. Additionally, for the instance involving the SF-36, numerous question were scored to provide a domain score, which was then in turn summed with multiple other domains for the final summary score. This process would sufficiently dilute any impact a single down graded response has on the final score used for correlation purposes.

This study is also limited in part by its small sample size. The strategy of mailed invitations was only able to achieve a modest response rate of 23%. A more active recruitment strategy may have improved the number of participants.[42] Approaches such as phone solicitation, or monetary reimbursement are considerations for future endeavors.

Small numbers of participants raise the concern of accurate representation by the study group. There is also a preponderance of NYHA Class 1 participants. However, statistical comparison bears out that the study group had similar quality of life scores to the cohort from which the study group was drawn. Additionally, the exercise capacity demonstrated by the study group was comparable to that found by prior investigators. [15,21,22] These points, in conjunction with the random sampling method used, strongly support our belief that, although small in number and well functioning by NYHA criteria, the study population is not biased and is representative of the greater ACHD population.

If greater recruitment of subjects with decreased functional status had occurred, it is likely that a stronger correlation between the physical domain of the quality of life instruments and maximal exercise capacity would have been found. This is suspected by the trends seen with analysis by NYHA Class subgroups and those with $VO2_{peak}$ less than 85% of predicted normal.

Future Directions:

Future opportunities for assessing health-related quality of life in the adult congenital heart disease population abound. If a larger population was studied, a statistically significant difference in how women versus men with congenital heart defects judge their physical quality of life may have been uncovered. It is tempting to speculate that men may place more weight on maximal physical function as a determinant of their quality of life than women. Multiple social and cultural reasons accounting for this can be hypothesized (e.g. male pressure to play organized sports, male tendency to judge worth through success at physical competition).

A larger study group would have also provided a clearer understanding of the impact age imparts on perceived physical quality of life. The trend seen in young versus older age could represent several processes. The trend for differences in baseline quality of life scores, as well as correlation of those scores with maximal exercise measures, could represent youthful denial of any limitations due to congenital heart disease. Or equally as likely, age may provide multiple experiences where physical limitations have been acknowledge and are thus reflected in quality of life scores.

An alternatively explanation is perhaps the most important for the medical community. This trend for difference by age in physical quality of life scores and correlation with $VO2_{peak}$ may represent an improvement in surgical technique. The younger ACHD population has had the benefit of cardiothoracic surgery maturation. Perhaps new surgeries and greater experience applying the established techniques have imparted an improved sense of physical quality of life, albeit without advancing maximal exercise capacity. Research comparing physical function and quality of life would be most fruitful if directed at understanding how differing surgical procedures for the same defect affect long-

term outcomes. This may lend weight to optimizing the initial surgical therapy offered to a child with congenital heart disease.

Also interesting is the potential that a functional threshold for physical quality of life may exist. By the data presented here, this threshold may be defined as an achieved $VO2_{peak}$ less than 85% of predicted normal or a functional class less than NHYA functional Class I. It can be hypothesized that if one has the exercise capacity to consider themselves Class I or exceed 85% of predicted VO2, any increased ability beyond this level is not reflected in the instruments used here. However, if maximal exercise capacity is diminished below a threshold level, one's ability to maximally exert comes into play during routine daily life. The results at hand suggest this could possibly be the case. A larger sample size more evenly distributed between functional class and $VO2_{peak}$ would be required to more firmly establish this point.

An important avenue for future study pertains to the greater question of physical activity and quality of life in general. What physical abilities most influence our perceived quality of life? Assessment of the relationship between quality of life and such things as a six-minute walk or physical strength may add to our understanding of this domain.

Conclusions:

This study in an adult congenital heart disease population demonstrates a significant correlation between the objective endpoints of $VO2_{peak}$, Work and METS and the physical domains of the SF-36, SF-12, and SIP. This moderate correlation provides criterion validation for the physical domains of the health related quality of life instruments tested. Additionally, a moderate and significant correlation exists between the DASI predicted VO2 and measured $VO2_{peak}$.

Despite the multiple formal instruments now in wide research use, it is imperative that the medical establishment does not lose sight of the fact that quality of life remains an elusive entity. At our current understanding it is impossible to tell *exactly* what these instruments are measuring. [9] Our ability to quantify a subjective individual perception by the use of a scored questionnaire is a deceptive simplification. A better understanding of what these instruments measured is essential before wide spread application in a clinical setting can be undertaken.

References:

1. Hoffman J, Christianson R. Congenital heart disease in a cohort of 19,502 births: long-term follow-up. Am J Cardiol, 42: 641-647, 1978.

2. McNamara D, Latson L. Long-term follow-up of patients with malformations for which definitive surgical repair has been available for 25 years or more. Am J Cardiol, 50: 560-568, 1982.

3. Hajiro T, et al., Comparison of discriminative properties among disease specific questionnaires for measuring health related quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Car Med, 153 (3 Prt1): 785-790, March, 1998.

4. Morris C, Menashe V. 25-year mortality after surgical repair of congenital heart defect in childhood; a population based cohort study. JAMA, 266(24): 3447-3452, 1991.

5. Perloff J, Child J. Congenital Heart Disease in Adults, 2nd Ed. Toronto, Saunders Co., p. 149, 1998.

6. Guyatt G, Thompson P, Berman L, et al. How should be measure function patients with chronic heart and lung disease? J Chron Dis. 38(6): 517-524, 1985.

7. Guyatt G, Feeny D, Patrick D, Measuring helath related quality of life. Annl of Int Med. 118: 622-629, 1993.

8. Guyatt G, Veldhuyzen S, Feeny D, Patrick D, Measuring quality of life in clinical trials: a taxonomy and review. CMAJ, 140: 1441-1447, 1989.

9. Leplege A, Hunt S, The problem of quality of life in medicine, JAMA, 278(1): 47-50, 1997.

10. Bergner M, Bobbitt R, et al. The Sickness Impact Profile: development and final revision of a health status measure, Medical Care, 19(8): 787-805, 1981.

11. Tarlov A, Ware J, Greenfield S, et al. The Medical Outcomes Study: an application of methods for monitoring the results of medical care, JAMA, 262(7): 925-930, 1989.

12. Kaplan R, Bush J, Health related quality of life measurement for evaluation research and policy analysis. Health Psych. 1(1): 61-80, 1982.

13. McSweeny A, Creer T, Health related quality of life assessment in medical care. Dis.-a-Month, 41(1):1-71, 1995

14. Gersony W, Hayes C, Driscoll D, et al. Second natural history study of congenital heart defects: quality of life of patients with aortic stenosis, pulmonary stenosis, or ventricular septal defect. Circ. 87 (2): (Supp I) I-52-I-65, 1993.

15. Meijboom F, Szatmair A, Deckers J, et al. Cardiac status and health related quality of life in the long term after surgical repair of Tetralogy of Fallot in infancy and childhood. J. Thoracic and Cardiovasc. Surg. 110(4): 883-891, 1995.

16. Favarato M, Romano B. Heart surgery in childhood. Impact on the quality o life. Arquivos Brasileiros de Cardiol. 62(3): 171-4, 1994.

17. Tomita H. Adolescent congenital heart disease: quality of life in patients not undergoing intracardiac repair. J of Cardiol. 24(5): 405-9, 1994.

18. Harrison D, Liu P, Walters J, et al. Cardiopulmonary function in adults patients late after Fontan repair. JACC, 26(4): 1016-21, 1995.

19. Kaplan J, Foster E, Redberg R, Schiller N. Exercise Doppler echocardiography identifies abnormal hemodynamics in adults with congenital heart disease. Amer. Heart J. 127(6): 1572-1580, 1994.

20. Jonsson H, Torbjorn I, et al. Work capacity and central hemodynamics thirteen to twenty-six years after repair of Tetralogy of Fallot. J of Thorc and Cariovasc Surg, 110(2): 416-426, 1994.

21. Winter U, Fritsch J, et al. Cardiopulmonary capacity of patients with congenital heart defects in childhood, adolescence and adulthood. Z Kardiol. 83(3):131-9, 1994.

22. Rigolin V, Li J, Hanson M, et al. Role of right ventricular and pulmonary functional abnormalities in limiting exercise capacity in adults with congenital heart disease. Am J Cardiol. 80(3): 315-22, 1997.

23. Ware J, Snow K, Kosinski M, Gandek B. SF-36 Health Survey, Manual and Interpretation Guide, The Health Institute, New England Medical Center, Boston, 1993.

24. Stewart A, Sheldon G, Hays R, et al. Functional status and well being of patients with chonic conditions; results from the Medical Outcomes Study. JAMA. 262(7): 907-913, 1989.

25. Guyatt G, Measurement of health related quality of life in heart failure. JACC, 22(4-supp. A): 185A-191A, 1993.

26. McHorney C, Ware J, Raczek A. The MOS 36 item short form heallth survey (SF-36): psychometric and clinical tests of validity in measuring physical and mental health constructs. Medical Care. 31(3): 247-263, 1993.

27. Ware J, Kosinski M, Keller S. A 12-item short form health survey: construction of sclaes and preliminary tests of reliability and validity. Medical Care. 34(3): 220-233, 1996.

28. Ware J. Standards for validating health measures: definition and content. J. Chronic Dis. 40(6): 473-480, 1987.

29. Ware J, Kosinski M, Keller S, SF-36 Physical and Mental Health Summary Scales: A User's Manual, The Health Institute, New England Medical Center, Boston, 1994.

30. Ware J, Kosinski M, Keller S, SF-12, How to score the SF-12 physical and mental health summary scales, 2nd Ed., The Health Institute, New England Medical Center, Boston, 1995.

31. Damiano A, Sickness Impact Profile, User's Manual and Interpretation Guide, John Hopkins University, 1996.

32. Hlatky M, Boineau R, Higginbotham M, et al. A brief self administrered questionnaire to determine functional capacity: the Duke activity status index. Amer. J. Card. 64: 651-654, 1989.

33. Casey F, Craig B, Mulholland H. Quality of life in surgically palliated complex congenital heart disease. Arch. Dis. Child. 70(5): 382-386, 1994.

34. Lachin J. Introduction to sample size determination and power analysis of clinical trials. Controlled Clin Trials, 2: 93-113, 1981.

35. Froelicher, V. Exercise and the Heart, 2nd ed. Chicago, Year Book Medical Publishers, pp. 67-71, 1987.

36. Ross, R. Interpreting Exercise Tests. CSI Software, Houston, pp. 7-19, 1989.

37. Wasserman K, Hanse J, et al., Principles of Exercise Testing and Interpretation, 2nd ed. Williams and Wilkins, Philadelphia, 1994.

38. Wasserman K, Whipp B Koyal S, Beaver W. Anaerobic threshold and respiratory gas exchange during exercise. J of Applied Physio. 35(2): 236-243, 1973.

39. Davis J, Vodak P, Wilmore J, et al. Anaerobic threshold and maximal aerobic power for three modes of exercise. J of Applied Physio. 41(4): 544-550, 1976.

40. Criteria Committee, New York Heart Association, Inc. Diseases of the Heart and Blood Vessels. Nomenclature and Criteria for Diagnosis, 6th ed. Boston, Little, Brown and Company, page 114, 1964.

41. Kleinbaum, Kupper, Muller. Applied Regression Analysis and Other Multivariable Methods, 2nd Ed., Duxbury Press, Belmont CA, 1988.

42. Friedman L, Furberg C, DeMets D, Fundamentals of Clinical Trials. 3rd Ed. Springer, New York, 1998.

43. Foster E, Congenital Heart Disease in Adults, West J Med, 163: 492-498, 1995.

44. Brook R, Avery A, Quality of medical care assessment using outcome measures: executive summary. Santa Monica, Rand Corp. 1976.

45. Fisher L, van Belle G, Biostatistics, John Wiley and Sons, Inc., New York, 1993, pp 596-597.

Appendix A: References for Predicted Normal Values

Reference for Predicted Normal Values used in MedGraphics "BreezeEx 3.0" metabolic cart software:

Spirometry variables

FVC and FEV1:

Knudson RJ, et al., The maximal expiratory flow-volume curve: normal standards, variability, and effects of age. Am Rev Respir Dis, 113: 587-600, 1976.

MVV:

Clinical Pulmonary Function Testing. A Manual of Uniform Laboratory Procedures for the Intermountain Area, 2nd Ed., Salt Lake City, Utah, International Thoracic Society, 1984.

Exercise variables

$VO2_{peak}$:

Wasserman K, et al., Principles of exercise testing and interpretation, 2nd Ed., Williams and Wilkins, Philadelphia, 1994.

Appendix B: New York Heart Association Functional Classification

Class I	No limitations: Ordinary physical activity does not cause fatigue, shortness of breath, or palpitations
Class II	Slight Limitations: You are comfortable at rest, but ordinary physical activity results in fatigue, shortness of breath, palpitations, or chest pain
Class III	Marked limitations: Although you are comfortable at rest, less than ordinary activity will lead to symptoms
Class IV	Severe limitations: You are unable to carry on any physical activity without discomfort, and you even have shortness of breath, fatigue, palpitations, or chest pain at rest

Appendix C: Duke Activity Status Index (DASI)

Duke Activity Status Index [15]

Can You... (please check all that apply)

Yes	No	Weight	
-	-	2.75	take care of yourself, that is eating, dressing, bathing, or using the toilet?
-	-	1.75	walk indoors, such as around the house?
-	-	2.75	walk a block or two on level ground?
-	-	5.5	climb a flight of stairs or walk up a hill?
-	-	8.0	run a short distance?
-	-	2.7	do light work around the house like dusting or washing dishes?
-	-	3.5	do moderate work around the house like vacuuming, sweeping floors, or carrying in groceries?
-	-	8.0	do heavy work around the house like scrubbing floors, or lifting or moving heavy furniture?
-	-	4.5	do yard work like raking leaves, weeding, or pushing a power mower?
-	-	5.25	have sexual relations?
-	-	6.0	participate in moderate recreational activities like golf bowling, dancing, doubles tennis, or throwing a baseball or football?
-	-	7.5	participate in strenuous sports like swimming, singles tennis, football, basketball, or skiing?

End of Questionnaire

Note: the weights listed above are not reproduced on the actual questionnaire.

The Duke Activity Status Index regression equation for prediction of peak VO_2 consumption. [32]

$$VO_2 = (0.43 \times DASI) + 9.6$$

DASI is the sum of the weights of all positively endorsed questions.

Appendix D: The Complete SF 12 Health Survey

Instructions for completing the questionnaire
Please answer every question. Some questions may look like others, but each one is different. Please take the time to read and answer each question carefully by filling in the bubble that best represents your response.

Example

This is for your review. Do not answer this question. The questionnaire begins with the section ***Your Health in General*** below.

For each question you will be asked to fill in a bubble in each line.

1. How strongly do you agree or disagree with each of the following statements?

	Strongly agree	**Agree**	**Uncertain**	**Disagree**	**Strongly disagree**
a)I enjoy listening to music.	O	O	O	O	O
b)I enjoy reading magazines.	O	O	O	O	O

Please begin answering the questions now:

Your Health in General

1. In general, would you say your health is:

Excellent	Very Good	Good	Fair	Poor
O	O	O	O	O

2. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

	Yes, Limited A Lot	Yes, Limited A Little	No, Not Limited At All
a. **Moderate activities,** such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	O	O	O
b. Climbing several flights of stairs	O	O	O

Please turn the page to continue.

3. **During the past 4 weeks,** have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	Yes	No
a. Accomplished less than you would like	O	O
b. Were limited in the kind of work or other activities	O	O

4. During the **past 4 weeks,** have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	Yes	No
a. Accomplished less than you would like	O	O
b. Didn't do work or other activities as carefully as usual	O	O

5. During the past 4 weeks, how much did pain interfere with you normal work (including both work outside the home and housework)?

Not at all	A little bit	Moderately	Extremely	Quite a bit
O	O	O	O	O

6. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling.

	All of the Time	Most of the Time	Some of the Time	A Little of the Time	None of the Time
a. have you felt calm and peaceful?	O	O	O	O	O
b. did you have a lot of energy?	O	O	O	O	O
c. have you felt downhearted and blue?	O	O	O	O	O

7. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with you social activities (like visiting with friends, relatives, etc.)?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
O	O	O	O	O

Thank you **for completing this questionnaire**

SF 12 copyright Medical Outcomes Trust and John Ware, Ph.D. - All rights reserved.

Appendix E: Excerpts from MOS SF36

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

	Yes, Limited A Lot	Yes, Limited A Little	No, not Limited At All
a) **Vigorous activities,** such as running, lifting heavy objects, participating in strenuous sports	O	O	O
b) Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	O	O	O
c) Lifting or carrying groceries	O	O	O
d) Climbing several flights of stairs	O	O	O
e) Climbing one flight of stairs	O	O	O
f) Bending, kneeling, or stooping	O	O	O
g) Walking more than a mile	O	O	O
h) Walking several blocks	O	O	O
i) Walking one block	O	O	O
j) Bathing or dressing yourself	O	O	O

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

		Yes	No
a)	Cut down on the amount of time you spent on work or other activities	O	O
b)	Accomplished less than you would like	O	O
c)	Were limited in the kind of work or other activities	O	O
d)	Had difficulty performing the work or other activities (for example, it took extra time)	O	O

10. These questions are about how you feel and how things have been with you during **the past 4** weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during **the past 4weeks...**

	All of the time	**Most of the time**	**A good bit of the time**	**Some of the time**	**A little of the time**	**None of the time**
a) did you feel full of pep?	O	O	O	O	O	O
b) have you been a very nervous person?	O	O	O	O	O	O
c) Have you felt so down in the dumps nothing could cheer you up?	O	O	O	O	O	O
d) have you felt calm and peaceful?	O	O	O	O	O	O
e) did you have a lot of energy?	O	O	O	O	O	O
f) have you felt downhearted and blue?	O	O	O	O	O	O
g) did you feel worn out?	O	O	O	O	O	O
h) have you been a happy person?	O	O	O	O	O	O
i) did you feel tired?	O	O	O	O	O	O

11. How TRUE or FALSE is <u>each</u> of the following statements for you?

	Definitely True	**Mostly True**	**Don't Known**	**Mostly False**	**Definitely False**
a) I seem to get sick a little easier than other people	O	O	O	O	O
b) I am as healthy as anybody I know	O	O	O	O	O
c) I expect my health to get worse	O	O	O	O	O
d) My health is excellent	O	O	O	O	O

Thank you for completing this questionnaire

SF36 copyrighted Medical Outcomes Trust and John Ware Ph.D. all rights reserved.

Appendix F: Excerpts from Sickness Impact Profile

PLEASE RESPOND TO (CHECK) ONLY THOSE STATMENTS THAT YOU ARE SURE DESCRIBE YOU TODAY AND ARE RELATED TO YOUR STATE OF HEALTH.

1. I walk shorter distances or stop to rest often
2. I do not walk up or down hills
3. I use stairs only with mechanical support, for example, handrail, cane, crutches
4. I walk up or down stairs only with assistance from someone else
5. I get around in a wheelchair
6. I do not walk at all
7. I walk by myself but with some difficulty, for example, limp, wobble, stumble, have still leg
8. I walk only with help from someone
9. I go up and down stairs more slowly, for example, one step at a time, stop often
10. I do not use stairs at all
11. I get around only by using a walker, crutches, cane, walls, or furniture
12. I walk more slowly

CHECK HERE WHEN YOU HAVE READ ALL STATEMENTS ON THIS PAGE ___

PLEASE RESPOND TO (CHECK) ONLY THOSE STATEMENTS THAT YOU ARE SURE DESCRIBE YOU TODAY AND ARE RELATED TO YOUR STATE OF HEALTH.

1. I make difficult moves with help, for example, getting into or out of cars, bathtubs
2. I do not move into or out of bed or chair by myself
   but am moved by a person or mechanical aid
3. I stand only for short periods of time
4. I do not maintain balance
5. I move my hands or fingers with some limitation or difficulty
6. I stand up only with someone's help
7. I kneel, stoop, or bend down only by holding on to something
8. I am in a restricted position all the time
9. I am very clumsy in body movements
10. I get in and out of bed or chair s by grasping something for support or using a cane or walker
11. I stay lying down most of the time
12. I change position frequently
13. I hold on to something to move myself around in bed

(Continued on next page)

(Continued from previous page)

14. I do not bathe myself completely, for example, require assistance with bathing
15. I do not bathe myself at all, but am bathed by someone else
16. I use bedpan with assistance
17. I have trouble getting shoes, socks, or stockings on
18. I do not have control of my bladder
19. I do not fasten my clothing, for example, require assistance with buttons, zippers, shoelaces
20. I spend most of the time partly undressed or in pajamas
21. I do not have control of my bowels
22. I dress myself, but do so very slowly
23. I get dressed only with someone's help

CHECK HERE WHEN YOU HAVE READ ALL STATEMENTS ON THIS PAGE _____

Appendix G: Figures

Figure 1

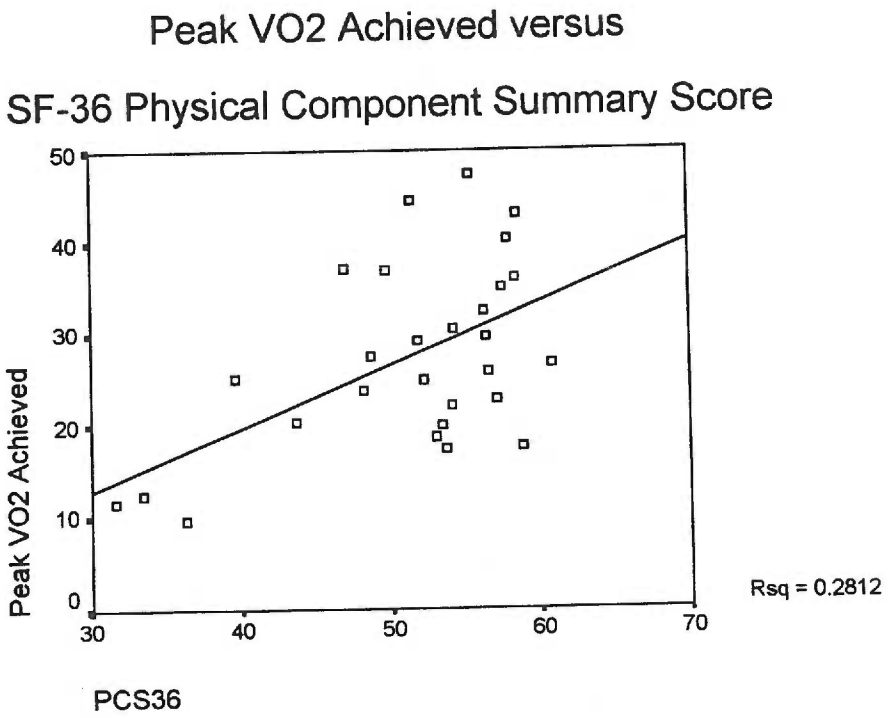


Figure 2

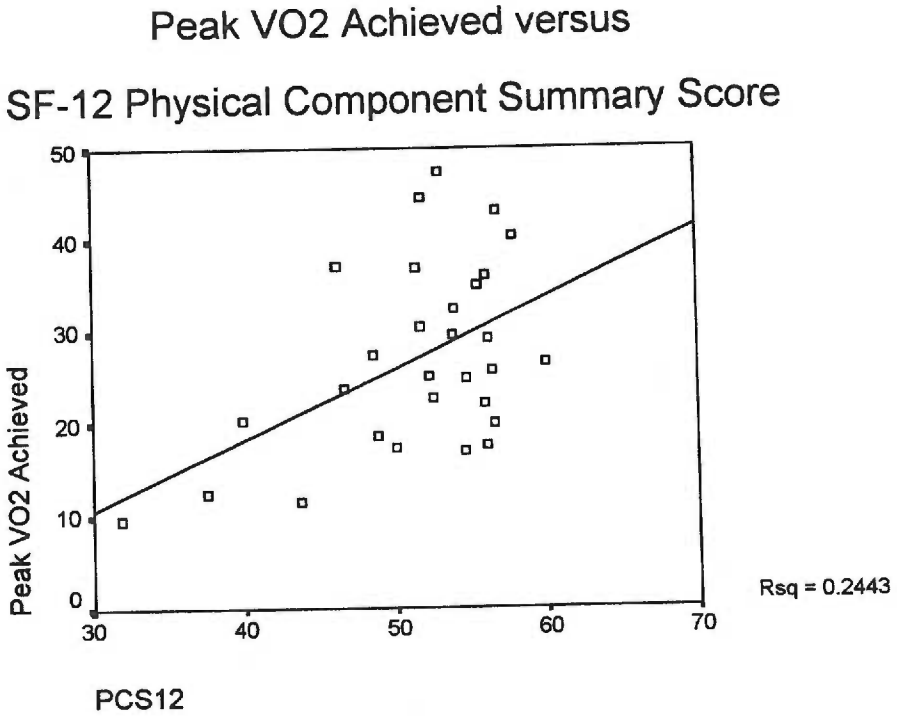


Figure 3

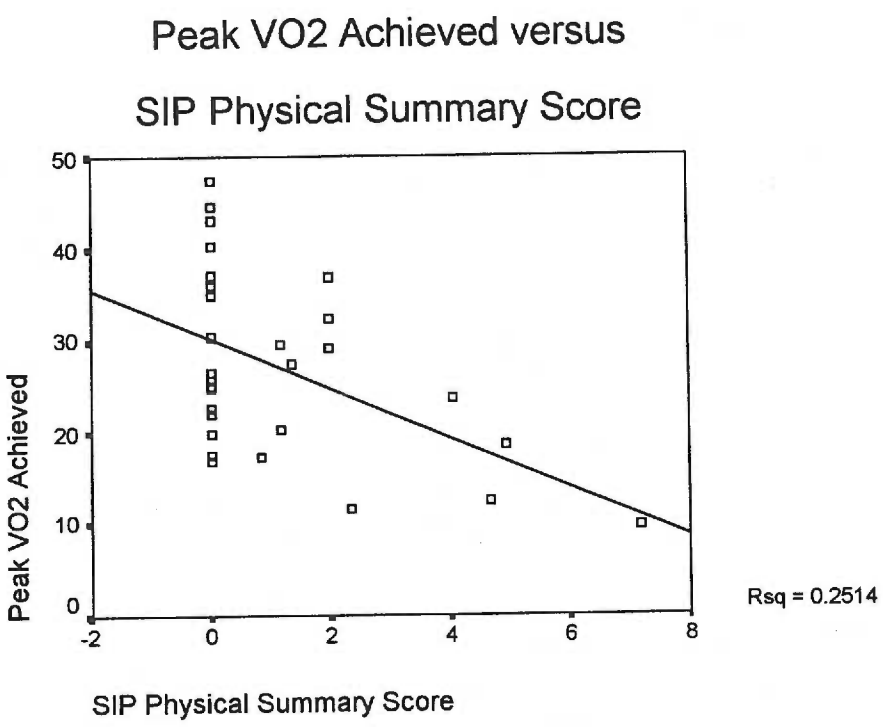


Figure 4

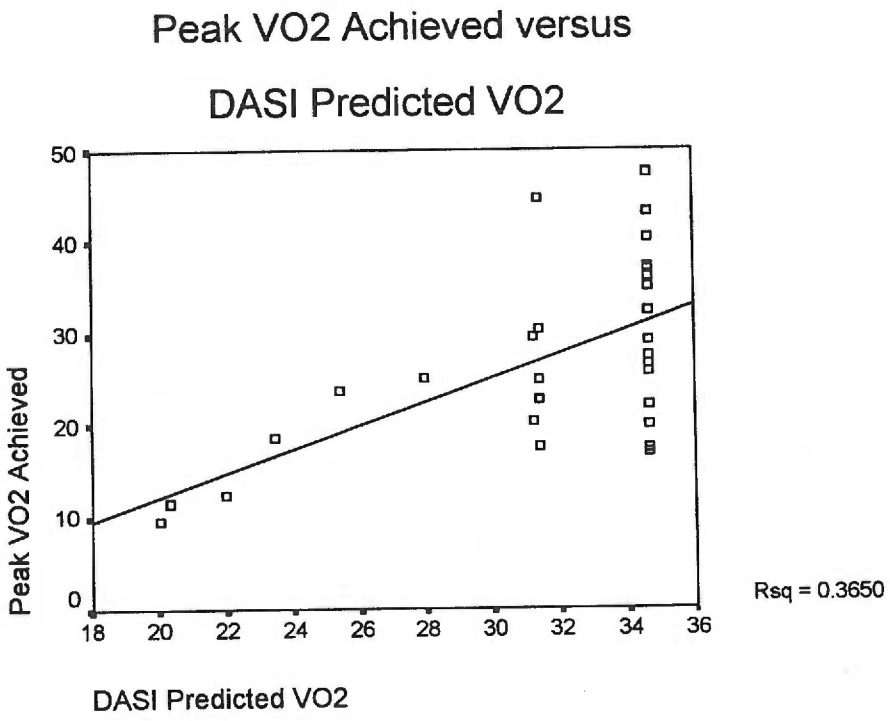


Figure 5

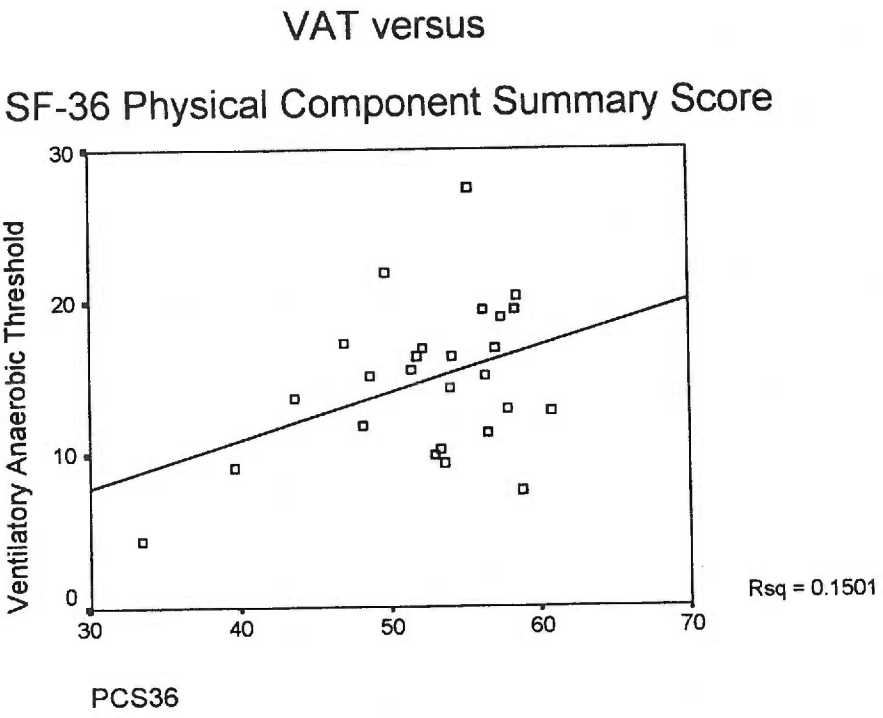


Figure 6

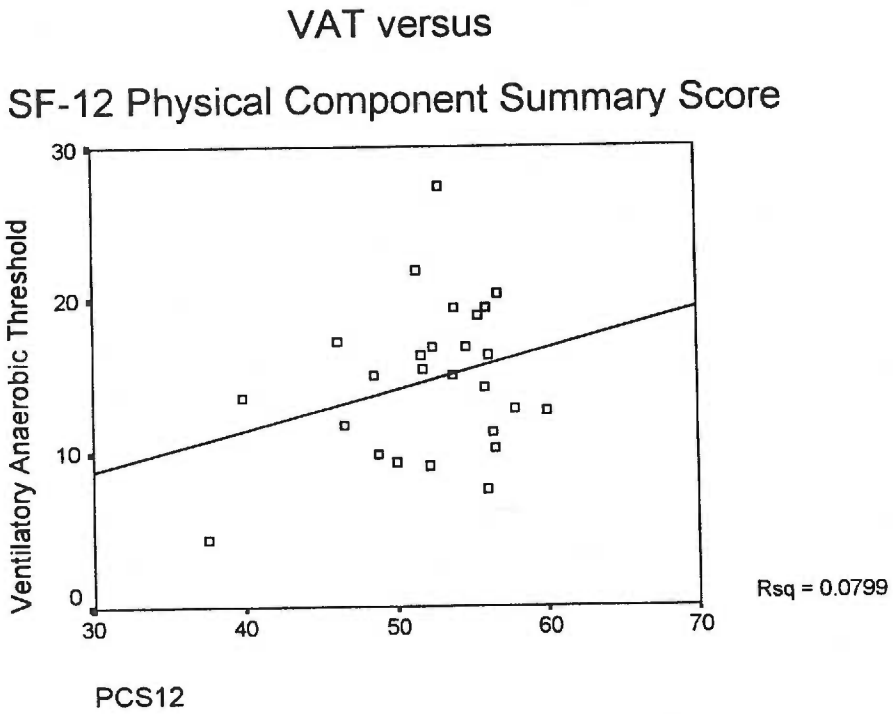


Figure 7

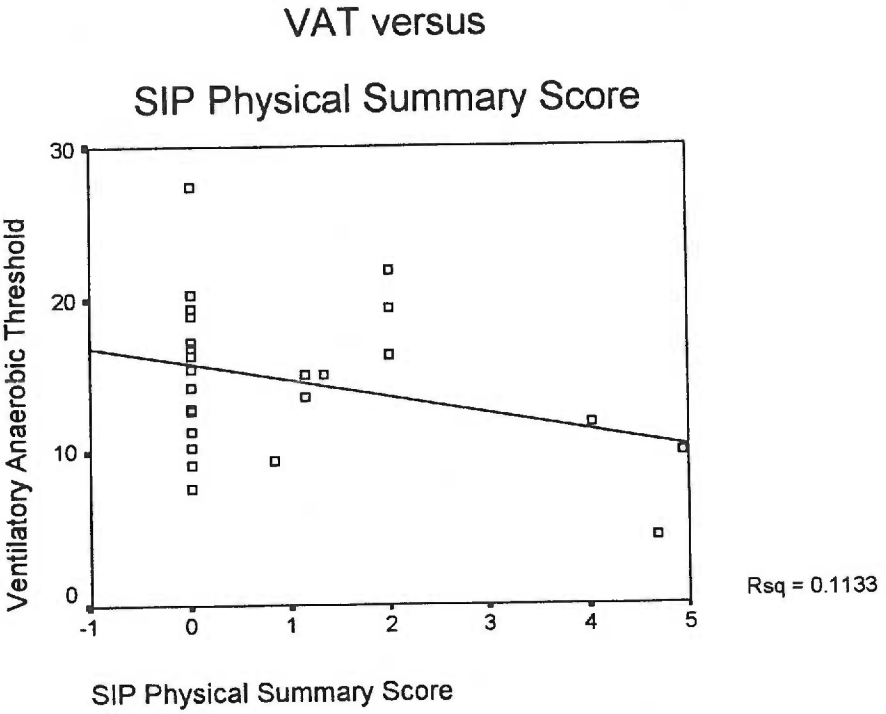


Figure 8

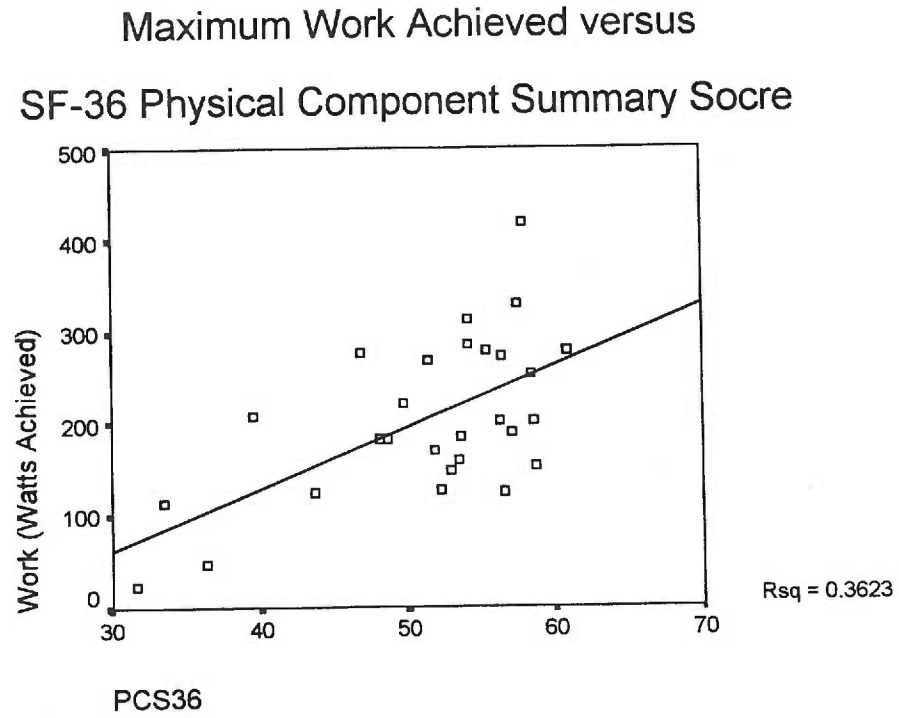


Figure 9

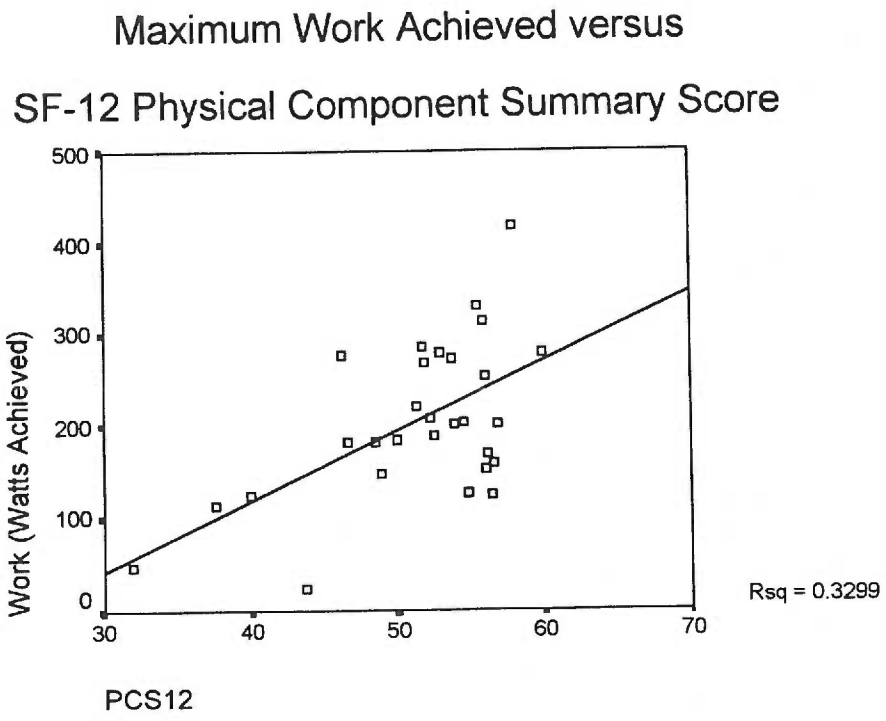


Figure 10

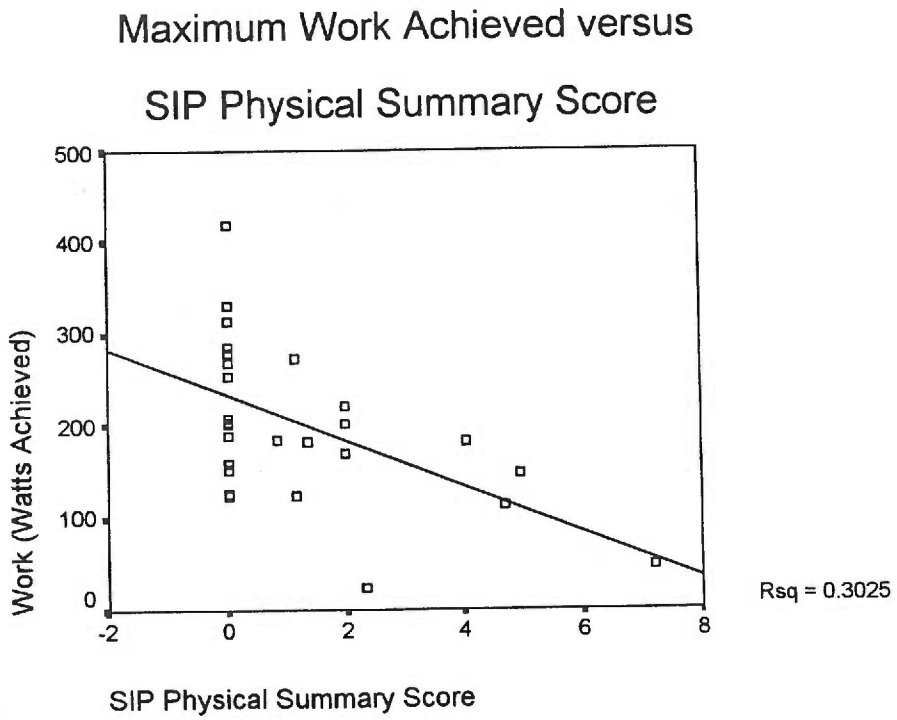


Figure 11

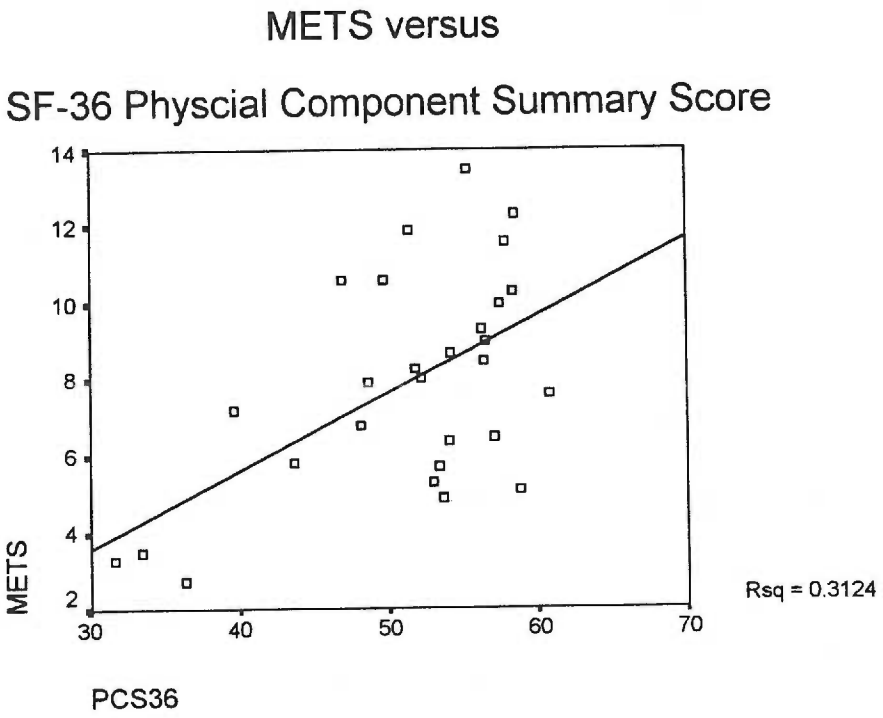


Figure 12

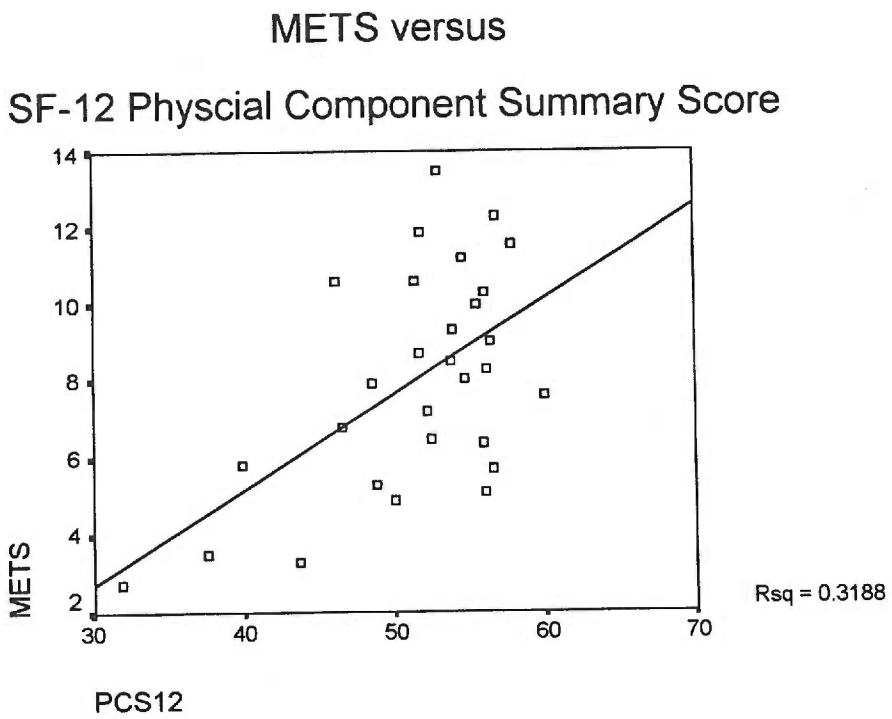
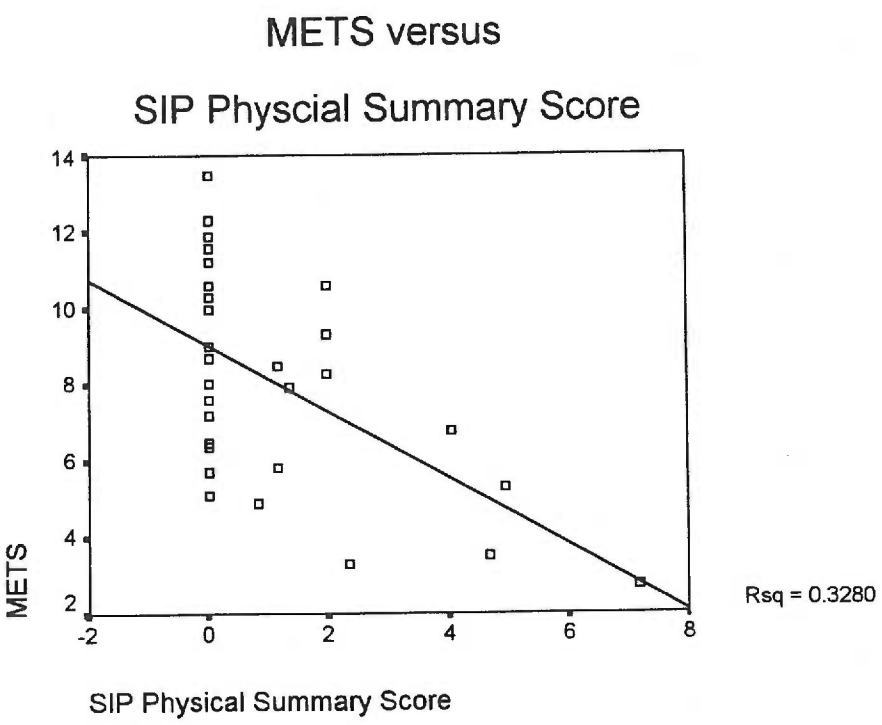


Figure 13


METS versus
SIP Physcial Summary Score
METS
14
12
10
8
6
4
2
-2
0
2
4
6
8
Rsq = 0.3280
SIP Physical Summary Score

Appendix H: Multiple Regression Models

$VO2_{peak}$ as dependent variable, model includes SF -36 quality of life scores.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-48.219	25.541		-1.888	.075	-101.879	5.441
	AGE	.222	.182	.177	1.218	.239	-.161	.606
	NYHA	-.465	3.495	-.031	-.133	.896	-7.808	6.879
	PcpredFEV	-.405	.237	-.674	-1.708	.105	-.903	.093
	PcpredFVC	.653	.249	.983	2.624	.017	.130	1.175
	PcPredMVV	-3.42E-02	.112	-.068	-.305	.764	-.270	.202
	SEX	5.756	3.284	.286	1.753	.097	-1.143	12.655
	MCS36	.236	.159	.262	1.486	.155	-.098	.569
	PCS36	.716	.259	.559	2.764	.013	.172	1.261

a. Dependent Variable: Peak VO2 Achieved

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.853[a]	.727	.605	6.4114

a. Predictors: (Constant), PCS36, PcPredMVV, MCS36, SEX, AGE, PcpredFVC, NYHA, PcpredFEV

$VO2_{peak}$ as dependent variable, model includes SF -12 quality of life scores.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-27.712	25.217		-1.099	.286	-80.492	25.069
	AGE	.188	.188	.155	1.002	.329	-.205	.581
	NYHA	-3.401	3.279	-.223	-1.037	.313	-10.265	3.462
	PcpredFEV	-.329	.265	-.554	-1.240	.230	-.883	.226
	PcpredFVC	.490	.259	.748	1.894	.073	-.051	1.032
	PcPredMVV	-1.07E-02	.123	-.021	-.087	.931	-.267	.246
	SEX	7.868	3.379	.391	2.328	.031	.795	14.941
	PCS12	.585	.301	.374	1.942	.067	-.046	1.215
	MCS12	.129	.181	.122	.713	.485	-.250	.509

a. Dependent Variable: Peak VO2 Achieved

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.813[a]	.661	.518	7.0927

a. Predictors: (Constant), MCS12, PcPredMVV, PCS12, SEX, AGE, PcpredFVC, NYHA, PcpredFEV

$VO2_{peak}$ as dependent variable, model excludes quality of life scores.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	13.590	11.112		1.223	.235	-9.519	36.700
	AGE	9.763E-02	.181	.080	.540	.595	-.278	.473
	NYHA	-7.498	2.494	-.492	-3.007	.007	-12.683	-2.312
	PcpredFEV	-.329	.270	-.555	-1.218	.237	-.892	.233
	PcpredFVC	.413	.266	.630	1.551	.136	-.141	.967
	PcPredMVV	8.563E-02	.112	.170	.768	.451	-.146	.318
	SEX	10.340	3.163	.514	3.269	.004	3.762	16.918

a. Dependent Variable: Peak VO2 Achieved

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.770[a]	.593	.477	7.3868

a. Predictors: (Constant), SEX, PcpredFEV, AGE, NYHA, PcPredMVV, PcpredFVC

PCS36 as dependent variable

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	61.641	9.190		6.708	.000
	Peak VO2 Achieved	.409	.181	.525	2.261	.036
	AGE	-.369	.150	-.377	-2.453	.024
	NYHA	-4.013	2.400	-.343	-1.672	.111
	PcpredFEV	.279	.223	.596	1.255	.225
	PcpredFVC	-.380	.226	-.735	-1.684	.109
	PcPredMVV	5.170E-02	.090	.132	.573	.574
	SEX	-1.300	3.090	-.083	-.421	.679

a. Dependent Variable: PCS36

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.776[a]	.602	.455	5.8747

a. Predictors: (Constant), SEX, AGE, PcpredFEV, NYHA, Peak VO2 Achieved, PcPredMVV, PcpredFVC