

Health-related Quality of Life and Patient-reported Outcomes in Pulmonary Arterial Hypertension

Hubert Chen¹, Darren B. Taichman², and Ramona L. Doyle^{3,4}

¹Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of California San Francisco, San Francisco, California;

²Pulmonary and Critical Care Division, Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; ³Wall Center for Pulmonary Vascular Disease, Division of Pulmonary and Critical Care Medicine, Stanford University, Palo Alto, California; and ⁴Gilead Sciences, Inc., Foster City, California

Advances in our understanding of the basic pathophysiology of pulmonary arterial hypertension (PAH) has led to an expanding number of therapeutic options. The ultimate goals of therapy are to lengthen survival while improving symptoms and quality of life. A wealth of research in other conditions has established health-related quality of life (HRQoL) to be an important clinical endpoint. Until recently, however, little was known about HRQoL in PAH, and how best to measure it. Over the past few years, several studies have begun contributing to this growing area of research. Instruments used to assess HRQoL have varied between studies. The extent to which these instruments are valid in PAH depend on their specific measurement properties. In this article, we provide an overview of the different types of patient-reported outcomes (PROs) used in PAH, focusing in particular on the measurement of HRQoL. In the process, we review the current literature on HRQoL in PAH, summarize the available data from randomized controlled trials, and discuss the implications of these findings on future research. Despite significant progress, the study of HRQoL in PAH remains a nascent field relative to other conditions. As the use of PROs continues to increase, additional work will be needed to begin standardizing the reporting and interpretation of such outcomes in future clinical trials.

Keywords: health status; questionnaire; validation

The development of effective treatments for pulmonary arterial hypertension (PAH) has begun transforming what was once considered a rapidly lethal diagnosis into a condition characterized by chronic dyspnea and progressive functional impairment. The goals of therapy have expanded from increasing survival to improving quality of life, following the paradigm set by other conditions such as congestive heart failure and chronic obstructive pulmonary disease. Although survival remains of central importance, its utility as an endpoint in clinical trials is limited by the fact that patients with PAH are now living significantly longer on available medical therapy, making true placebo-controlled trials challenging (1). Future drug trials are likely to focus on the benefit of “add-on” therapy for which the cumulative effect may be an incremental improvement in quality of life, in spite of only modest changes in survival.

Quality of life represents a broad range of human experiences related to one’s overall well-being and may be influenced by a multitude of nonmedical factors, such as financial status, individual freedom, and one’s own personal environment (2).

(Received in original form February 22, 2008; accepted in final form April 22, 2008)

H.C. is funded by career development grant K23 HL086585. Support for this conference, including travel for D.B.T. and R.L.D., was provided by unrestricted educational grants from Actelion Pharmaceuticals, Pfizer, Gilead Sciences, United Therapeutics, and Lung Rx, Inc.

Correspondence and requests for reprints should be addressed to Hubert Chen, M.D., M.P.H., Pulmonary and Critical Care Division, University of California San Francisco, 350 Parnassus Avenue, Suite 609, San Francisco, CA 94143. E-mail: hubert.chen@ucsf.edu

Proc Am Thorac Soc Vol 5. pp 623–630, 2008

DOI: 10.1513/pats.200802-0205K

Internet address: www.atsjournals.org

The assessment of quality of life in clinical trials, however, is concerned with the more defined concept of *health-related* quality of life (HRQoL), which has been described as “the functional effect of an illness and its consequent therapy upon a patient, as perceived by the patient” (3). HRQoL may be a particularly relevant endpoint in PAH. Existing treatments for PAH often require frequent dosing and monitoring, necessitate the use of specialized drug delivery systems, and can be associated with serious adverse events. Therefore, while new and emerging therapies may improve pulmonary hemodynamics and exercise capacity, such interventions may or may not lead to improved quality of life.

This review focuses primarily on the measurement of HRQoL and its utility as a complementary endpoint in clinical studies of PAH. In the context of HRQoL, we also discuss other types of patient-reported outcome (PRO) measures commonly used in studies of PAH, for which many of the same basic principles apply.

DEFINITION OF A PRO MEASURE

Instruments used to assess HRQoL represent a much broader category of health status measures collectively referred to as PROs. As implied by its name, a PRO is any measurement of a patient’s health status that is directly elicited from the patient (4, 5). Although PROs are commonly associated with HRQoL instruments, a PRO measure can be used to assess any aspect of a patient’s health. Examples range from unidimensional symptom scales, such as the Borg Dyspnea Index (BDI) (6), to complex multidimensional constructs, as in the case the of HRQoL. In most cases, PRO measures are questionnaires, either self-administered or administered by a trained interviewer. In contrast, classification systems that incorporate a provider’s impression of the patient’s response, such as New York Heart Association/World Health Organization (NYHA/WHO) functional class, are not considered true PRO measures.

RATIONALE FOR USING PRO MEASURES

Traditionally, biomedical research has relied on physiologic endpoints to understand the effects of an intervention on a given disease. There is growing recognition, however, that changes in physiologic measures may not always translate into a tangible benefit as perceived by the patient (7). In PAH, for example, it is known that pulmonary hemodynamics do not correlate well with how patients feel and function in their daily lives (8–10). For that reason, regulatory agencies have begun to demand that pivotal trials incorporate endpoints that are both physiologically relevant as well as patient-centered (4).

PROs offer certain advantages over other types of health outcome measures. Most commonly, PROs are used to ascertain treatment effects evident only to the patient (which may otherwise go unrecognized by the physician in an objection evaluation). These might include assessments of symptoms,

such as dyspnea, or broader concerns, such as “quality of life.” As such, PROs are unique in that they directly assess benefits to the patient for which no adequate observable or physical measures exist. Furthermore, PROs are often designed to capture the patient’s perspective, thereby adding another dimension to our understanding of a patient’s response to treatment that cannot be extrapolated from physiologic or clinical endpoints. Finally, PROs are relatively quick and easy to administer, and provide a more formal assessment than outcomes that require a clinical interpretation of the patient’s status. Figure 1 depicts the relationships among various types of endpoints in PAH, and the context in which PRO measures are frequently used.

TYPES OF PRO MEASURES USED IN PAH

The choice of PRO measure depends on its intended purpose. As shown in Figure 1, PROs in PAH are commonly used to measure symptoms, functional status, or HRQoL. Instruments designed to measure symptoms often consist of single-item scales, for example the BDI (6). Such rating scales typically focus on the measurement of a defined construct, the interpretation of which is usually straightforward (e.g., from no shortness of breath to severe dyspnea). Consequently, such measures generally do not require the level of conceptual grounding and psychometric validation expected of more sophisticated health status instruments.

Functional status differs from symptoms in that it refers to the extent to which symptoms interfere with a patient’s ability to perform certain tasks or activities (7). Instruments used to assess functional status include a wide variety of measures. They can range from single-item scales similar to those used to rate symptoms, for example the modified Medical Research Council [MRC] scale (11), to more complex measures that closely resemble HRQoL instruments. Measures of functional status extend beyond the determination of exercise capacity alone in that they incorporate an individual’s ability to perform functional activities, as opposed to merely how far a person can walk in 6 minutes.

The concept of HRQoL encompasses that of both symptoms and functional status (12). In principle, HRQoL instruments are designed to capture not only the level of impairment, but also the impact of that impairment on an individual’s perceived physical, psychological, and social well-being (2). HRQoL is therefore a multidimensional construct by definition. Most HRQoL instruments are composed of multiple domains; however, instruments vary in both scope and content. Some investigators distinguish measures of “health status” from true “quality of life” instruments, which take into account the patient’s own expectations or internal standards (5, 13). To the extent that such instruments reflect those aspects of life valued

most by patients, each may provide further insight into the specific pathways by which PAH leads to HRQoL impairment.

VALIDATION OF HRQOL AND PRO MEASURES

In general, physicians and clinical investigators will agree that HRQoL is important to assess. In everyday clinical practice, physicians often inquire in an informal manner about HRQoL to determine whether a patient with PAH is benefiting from therapy. In clinical trials, however, concern regarding the use of HRQoL as an endpoint centers not on the issue of relevance, but on whether the instruments used to measure it are reliable, valid, and responsive to the effects of treatment (14). Instruments must also be interpretable insofar as they must provide results that represent a meaningful change to the patient. In 2006, the United States Food and Drug Administration (FDA) released a draft guidance document for industry on the appropriate development and use of PRO measures in medical product development (4). The process of instrument development and validation represents a highly specialized discipline that is beyond the scope of this review, and has been described well by others (15). Table 1 provides a brief overview of the methods commonly used to assess the psychometric adequacy of HRQoL and PRO measures.

INSTRUMENTS USED TO ASSESS HRQOL IN PAH

Until a few years ago, very little was known about HRQoL impairment in PAH. Driven by expanding therapeutic options and the ability to focus on endpoints beyond survival, an increasing number of studies have begun to shed light on this previously neglected area of research. Instruments used by investigators have varied from study to study, in large part due to the lack of data on the performance of different measures in PAH. As a result, past investigators have had to either rely on the use of generic instruments or adapt existing measures originally developed for related conditions. Table 2 provides a summary of the various instruments used in studies of HRQoL in PAH.

Generic measures, such as the Medical Outcome Study 36-item Short Form Health Survey (SF-36) (16) and the Nottingham Health Profile (NHP) (17), are advantageous in that they can be applied across a broad spectrum of disease states—even healthy individuals—thereby allowing comparisons with population norms over multiple domains. *Multi-attribute utility measures*, such as the EuroQol (EQ-5D) (18) and the Australian Assessment of Quality of Life (AQoL) (19), also provide a multidimensional assessment of general health, but in addition can be used to derive preference-based “utility” scores that can be applied in economic analyses. Utilities can also be obtained via direct elicitation (e.g., visual analog scales [VAS], standard

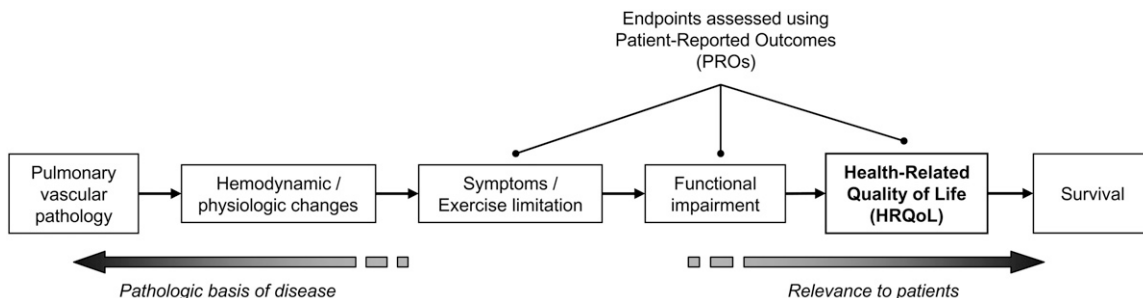


Figure 1. Simplified conceptual model depicting the relationship between different types of endpoints used in studies of pulmonary arterial hypertension.

TABLE 1. METHODS USED TO ASSESS THE PSYCHOMETRIC ADEQUACY OF HRQoL AND PRO MEASURES

Measurement Property	Method of Assessment	Description	Considerations
Reliability	Test-retest reproducibility	Stability of scores upon readministration when no change has occurred in the concept of interest	Most important type of reliability for instruments used in clinical trials
	Internal consistency	Extent to which the items consistently measure the same underlying construct (e.g., Cronbach's coefficient alpha).	Internal consistency alone is not sufficient evidence of reliability
	Inter-interviewer agreement	Agreement between responses when an instrument is administered by different interviewers	Only relevant to instruments administered by an interviewer
Validity	Content, face validity	Completeness, relevance, and comprehensibility of items in assessing the concept of interest	Difficult to assess quantitatively, requiring use of qualitative methods
	Construct validity		
	Convergent validity	Correlation with measures of related constructs in the hypothesized manner, consistent with the proposed conceptual framework	Referent measures may be previously validated PROs or non-PRO measures
	Divergent, discriminant validity	Lack of correlation with measures that are intended to be different or conceptually distinct	
	Known groups validity	Ability to distinguish groups that are expected or known to be different with regard to the concept of interest	
	Predictive validity	Ability to accurately predict future health status or other relevant outcomes	
Responsiveness (longitudinal validity)	Calculation of responsiveness statistic (e.g., effect size)	Ability to detect changes in the measured concept over time, usually in response to a specific intervention or known change in health	Responsiveness of an instrument may depend on the time interval
Interpretability	Minimal important difference (MID)	Smallest difference in score that is considered meaningful to patients and/or clinically relevant	MID must be determined by triangulation of different methods, and may vary for different subpopulations

Definition of abbreviations: HRQoL = health-related quality of life; PRO = patient-reported outcome.

Adapted from FDA Guidance for Industry on Patient-Reported Outcome Measures (4).

gamble), though yield little information regarding HRQoL beyond the overall level of impairment. Due to their broad content and emphasis on functional impairment, generic instruments are sometimes referred to more generally as “health status” measures.

In contrast, *condition-specific measures* are designed to focus on those issues most relevant to a particular group of patients, and therefore may be more sensitive to treatment changes than generic measures. Given the cost and time associated with developing new instruments, it is not uncommon for investigators to modify existing measures for use in less prevalent conditions, as seen in cystic fibrosis, idiopathic pulmonary fibrosis, and sarcoidosis (20–22). In the case of PAH, cardiac- and respiratory-specific instruments have frequently been used given their emphasis on the role of dyspnea and activity limitation in the disablement process (8–10, 23). The validity of such instruments in PAH, however, depends in part on the extent to which those aspects of the disease that are shared in common are considered meaningful and important to patients with PAH.

Studies of HRQoL in PAH

Initial studies specifically assessing HRQoL in PAH were cross-sectional in nature and focused primarily on describing the level of impairment. Shafazand and coworkers studied 53 patients using both generic and cardiac-specific measures (23). Patients reported significant impairment in all domains of the NHP, including energy, emotional reaction, pain, physical mobility, sleep, and social isolation in comparison to population norms. Likewise, HRQoL as measured by the Congestive Heart Failure Questionnaire showed levels of impairment comparable to NYHA/WHO class III-IV left-sided congestive heart failure. Standard gamble derived utilities obtained in the same study indicated that patients with PAH were willing to accept a 29% risk of death to achieve perfect health. Differences in NHP and Congestive Heart Failure Questionnaire scores were observed

for patients treated with intravenous prostacyclin compared with those who were not; however, no difference in utilities were noted, suggesting that such preference-based measures may be less discriminative.

In a similar fashion, Taichman and colleagues studied 155 patients with PAH employing another widely used generic measure, the SF-36, in addition to a popular respiratory-specific measure, the St. George's Respiratory Questionnaire (SGRQ) (8). Both the physical and mental component summary scores of the SF-36 (PCS and MCS) were significantly depressed, demonstrating scores comparable with those of other debilitating and life-threatening conditions such as spinal cord injury and metastatic cancer. All domains were affected, with the greatest impairment observed in the general health, physical functioning, and role-physical and role-emotional domains. The SGRQ, and each of its subscales, also demonstrated evidence of substantial impairment. In a subset of patients, the SF-36 PCS correlated reasonably well with other physical assessments, such as 6-minute walk distance (6MWD) ($r = 0.62$) and the BDI ($r = 0.46$), but not with hemodynamic measurements, providing evidence of both convergent and divergent validity. In addition, the SF-36 PCS was able to discriminate subgroups of patients known to have worse survival based on NYHA/WHO class (III versus II) and PAH etiology (systemic sclerosis-related versus idiopathic).

More recent studies have been longitudinal in design and aimed to assess the measurement properties of existing instruments when applied to patients with PAH. Cenedese and coworkers studied the performance characteristics a German cross-cultural adaptation of the Minnesota Living with Heart Failure Questionnaire (MLHFQ) in 48 patients with either PAH ($n = 26$) or chronic thromboembolic pulmonary hypertension ($n = 22$) (9). The MLHFQ demonstrated high internal consistency ($\alpha = 0.92$), as well as good test re-test reproducibility ($r = 0.94$) in a subset of patients. The total and physical subscores correlated significantly with NYHA/WHO class ($r =$

TABLE 2. HEALTH STATUS MEASURES USED BY STUDIES OF HRQoL IN PAH

Instrument	No. of Items	Response Format	Recall Period	Domains Assessed (Items per Domain)	Scoring Method	Possible Range	Reference
Generic Measures							
Medical Outcome Study 36-item Short Form Health Survey (SF-36)	36	Mixed	Now to 4 weeks	Physical functioning (10) Role physical (4) Bodily pain (2) General health (5) Vitality (4) Social functioning (2) Role emotional (3) Mental health (5)	Weighted	Physical component summary (PCS): 0–100 Mental component summary (MCS): 0–100 Individual domains: 0–100	(8) (10) (26) (30) (25)
Nottingham Health Profile (NHP), Part I	38	Dichotomous (Yes/No)	At the moment	Physical mobility (8) Pain (8) Social isolation (5) Emotional reactions (9) Energy level (3) Sleep (5)	Weighted	Individual domains: 0–100* (No overall score)	(23) (28)
Multi-Attribute and Preference-based Utility Measures							
EuroQol (EQ-5D) utility index	5	3-option response scale	Today	Mobility (1) Self-care (1) Usual activities (1) Pain/Discomfort (1) Anxiety/Depression (1)	Multi-attribute utility theory	0 (death)–1 (ideal health)	(28)
Australian Assessment of Quality of Life (AQoL)	15	4-option response scale	1 week	Illness (3) Independent living (3) Social relationships (3) Physical senses (3) Psychological well-being (3)	Multi-attribute utility theory	0 (death)–1 (ideal health)	(10) (26)
Visual Analogue Scale	1	Continuous scale	Today	NA	NA	0–100	(23) (28)
Standard Gamble	NA	Recursive method	NA	NA	von Neumann-Morgenstern utility theory	0 (death)–1 (ideal health)	(23)
Condition-specific Measures							
Minnesota Living with Heart Failure Questionnaire (MLHFQ)	21	6-point Likert scale (0 to 5)	4 weeks	Physical (8) Emotional (5)	Summative	Total score: 0–105* Physical domain: 0–40* Emotional domain: 0–25*	(10) (9) (25)
Congestive Heart Failure Questionnaire†	20	7-point Likert scale (1 to 7)	2 weeks	Dyspnea (5) Fatigue (4) Emotional function (7) Mastery (4)	Summative	Individual domains: 1–7*	(23)
St. George's Respiratory Questionnaire (SGRQ)	50	Mixed	Now to 4 weeks	Symptoms (8) Activity (16) Impact (26)	Weighted	Total score: 0–100* Individual domains: 0–100*	(8)
PAH-specific Measures							
Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR)	65	Dichotomous and 3-option response scales	Today	Overall symptoms (25) - Energy (10) - Breathlessness (8) - Mood (7) Activity (15) Quality of Life (25)	Summative	Symptom score: 0–25* Activity score: 0–30* Quality of Life score: 0–25*	(28) (30)

Definition of abbreviations: HRQoL = health-related quality of life; NA = not applicable; PAH = pulmonary arterial hypertension.

* Higher scores indicate greater impairment.

† Later adapted to become the Chronic Heart Failure Questionnaire.

0.57–0.61), 6MWD ($r = 0.29–0.42$), and BDI ($r = 0.43–0.51$) in the expected manner, indicating good convergent validity. Among 38 patients treated with vasodilator therapy, MLHFQ scores appeared relatively responsive to improvements in NYHA/WHO class and 6MWD. Effect sizes observed approximated 0.5, consistent with a “moderate” change according to traditional distributional methods (24). In multivariate analyses, using a combined outcome of death, transplant, or pulmonary endarterectomy the total MLHFQ score demonstrated strong predictive validity relative to other noninvasive and invasive measures.

Both Chua and colleagues (10) and Zlupko and coworkers (25) have also studied the MLHFQ in PAH. Using pooled trial

data from 83 patients, Chua and colleagues compared the performance of the MLHFQ with the SF-36 and the AQoL, a multi-attribute utility measure (10). Total scores for all three instruments demonstrated good convergent validity, correlating significantly with 6MWD and NYHA/WHO class in the expected manner. Consistent with previous studies, HRQoL scores for all three instruments correlated poorly with hemodynamic measurements. In general, individual domains of the MLHFQ and SF-36 performed better than those of the AQoL, which appeared to be less sensitive to variation in functional measures. Likewise, within-patient changes in MLHFQ and SF-36 scores showed significant associations with corresponding changes in 6MWD and NYHA/WHO class over time, in contrast to the

AQoL, which was much less responsive. It should be noted, however, that substantially fewer patients completed the SF-36 and AQoL than the MLHFQ, which could have influenced their results. In a larger cohort consisting of 93 patients with PAH, Zlupko and coworkers also administered the MLHFQ and SF-36 and found comparable results (25).

Aside from clinical trials, few studies have prospectively evaluated HRQoL in PAH. In a prospective, open-label study, Keogh and colleagues used the SF-36 and AQoL to assess the effect of bosentan therapy on HRQoL in 177 patients with PAH (26). HRQoL was assessed at baseline and at 3-month intervals after initiation of therapy. HRQoL improved significantly from baseline to 3 months on multiple domains of the SF-36 (physical functioning, role-physical, vitality, social functioning, mental health, and role-emotional), as well as the total AQoL score. According to population-based estimates of variance for the SF-36, the effect sizes observed were in the moderate range (24). Mean change in AQoL score was statistically significant, but was less than the minimal important difference (MID) as defined by other investigators (27). Of interest, improvements in the SF-36 and AQoL persisted out to 6 months. These results must be interpreted with caution, however, as there was a substantial decrease in number of patients beyond 3 months, which may have been related to study cessation before the completion of follow-up or withdrawal due to worsening health status.

Attempting to address the need for a PAH-specific measure, McKenna and coworkers recently developed and validated the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) (28). The CAMPHOR comprises three separate scales designed to assess symptoms, functioning, and quality of life. Quality of life items were defined using a “needs-based” model, which postulates that life gains its quality from the ability and capacity of the individual to satisfy his or her needs (29). In that respect, the CAMPHOR differs from other HRQoL instruments, which generally do not make a distinction among such item content. Items of the CAMPHOR were derived from qualitative interviews conducted among 35 patients with PAH, which were then extensively field tested for face and content validity. Reliability and construct validity of the original instrument was evaluated in the United Kingdom among 91 patients. Each of its scales demonstrated high internal consistency ($\alpha = 0.90$ – 0.92) and good test-retest reproducibility ($r = 0.86$ – 0.92). The CAMPHOR also demonstrated good convergent and divergent validity in relation to the NHP and EQ-5D, and was able to adequately discriminate among patients based on their NYHA/WHO class.

The reliability and validity of the CAMPHOR in a United States population was recently tested by Gomberg-Maitland and colleagues (30). In that study, face and content validity were re-assessed among a subset of patients; no significant modifications to the original instrument were made. Overall, the U.S. CAMPHOR demonstrated good construct validity with respect to the SF-36 and 6MWD. Test-retest reproducibility and known groups validity among its subscales were adequate, though less impressive than originally reported in the United Kingdom. In particular, there appeared to be a possible “ceiling effect” for the symptom subscales (24–37% scoring the minimum), which was less conspicuous when using the total symptom score. Responsiveness and interpretability (i.e., meaningfulness of change) of the CAMPHOR in placebo-controlled trials for PAH remains to be established.

Use of HRQoL Instruments in Clinical Trials

Despite the relative paucity of psychometric data, the use of HRQoL measures in clinical trials has been increasing. Shown in Table 3 are randomized trials in PAH to date that have

included HRQoL as a secondary outcome. The instruments used have varied, thereby making it difficult to compare HRQoL results between studies. The generic measure most commonly used in PAH trials has been the SF-36. Condition-specific measures frequently employed include the MLHFQ and the Chronic Heart Failure Questionnaire (CHQ). Based on available data, domains related to physical functioning appear to be the most responsive to change in the trial setting.

The reporting of HRQoL data itself in clinical trials that have assessed it has generally been poor. Frequently, little information is provided other than whether a statistically significant difference was detected. Seldom are the attributes of the instrument described, or a rationale given for its use. In many cases, the magnitude of changes observed and the specific domains affected are not published in detail. Even when statistically significant differences are present, interpreting the results can be problematic. Additional research is needed to determine the absolute change in score associated with a meaningful difference in the population of interest (i.e., MID). In cases in which instruments have been used extensively for similar conditions, it may be useful for investigators to specify *a priori* what magnitude of change, or effect size, and in what domain(s), treatment effects are anticipated in light of the existing evidence.

Evaluating the responsiveness of HRQoL measures in clinical trials is rarely straightforward (31). Assessments of HRQoL are often treated as secondary endpoints, and therefore studies may be inadequately powered to detect meaningful differences. Case mix—particularly when PAH is associated with other chronic diagnoses—may confound differences in HRQoL when generic measures are used. In addition, the duration of follow-up in trials may be critical when evaluating change in HRQoL. Although the recall period of instruments falls within the time frame of most trials, the benefits associated with a change in therapy may accrue over time, particularly in the case of more distal outcomes, such as HRQoL. For instance, patients may develop a sense of mastery with regard to specialized drug delivery systems or may become accustomed or desensitized to minor side effects. Without studies of longer duration, it is not possible to know whether short-term increases in exercise capacity truly translate to sustained improvements in HRQoL over time.

FUTURE DIRECTIONS

Although significant advances have been made in the development and application of PRO measures in PAH, a number of basic questions still remain. Much of what is known about HRQoL in PAH has been inferred from related conditions in the field of cardiac and respiratory medicine. Consistent with research in other conditions, studies in PAH have shown that measures of physiologic response and exercise capacity, such as 6MWD, account for only a portion of the observed variance in HRQoL (8–10). Greater understanding regarding the disease-specific processes by which HRQoL becomes impaired in PAH is lacking. Emerging data from studies using qualitative techniques indicate that psychosocial factors, such as coping with uncertainty and accommodating medical therapy, may play an influential role (32). Neurocognitive impairments may be important as well (33).

The identification of factors which modify the relationship between treatment and outcomes is crucial to understanding why certain therapies, while efficacious, may not be always be effective. They may further help elucidate discrepancies between improvements in physiological endpoints and HRQoL. Developing a well-grounded conceptual framework is the first

TABLE 3. RANDOMIZED CLINICAL TRIALS IN PAH WITH HRQoL AS SECONDARY OUTCOME

Year	Reference	Intervention	N	Design	Etiology of PAH	NYHA Class	Follow-up Interval	6MWD Change	Instrument(s) Used	HRQoL Findings
1996	(37)	Epoprostenol	81	Randomized, open-label, controlled	IPAH	III, IV	12 wk	113m	NHP, CHQ	Significant improvement in all 4 domains of CHQ, 2 of 6 domains of NHP
2002	(38)	Iloprost	203	Randomized, placebo-controlled	IPAH, CTD, AA, CTEPH	III, IV	12 wk	36m	EQ-5D utility index and VAS, SF-12	Significant improvement in EQ-5D VAS. Nonsignificant trend in EQ-5D utility index. No change in SF-12.
2002	(39)	Treprostinil	470	Randomized, double-blind, placebo-controlled	IPAH, CTD, CHD	II, III, IV	12 wk	16m	MLHFQ	Significant improvement in physical dimension score of MLHFQ. Nonsignificant trend in total MLHFQ score.
2003	(40)	Beraprost	116	Randomized, double-blind, placebo-controlled	IPAH, CTD, CHD	II, III	3, 6 mo	22-31m	MLHFQ	No significant improvement in total MLHFQ or subscores.
2004	(41)	Sildenafil	22	Randomized, double-blind, crossover	IPAH	II, III	6 wk	44%↑ exercise time	CHQ	Significant improvement in dyspnea and fatigue domains of CHQ. Nonsignificant trend in emotional domain.
2004	(42)	Sitaxsentan	178	Randomized, double-blind, placebo-controlled	IPAH, CTD, CHD	II, III, IV	12 wk	33-35m	SF-36	No significant differences in SF-36.
2004	(43)	Treprostinil	90	Post hoc analysis of RCT	CTD	II, III, IV	12 wk	21m	MLHFQ	Nonsignificant trend toward improvement in physical dimension score of MLHFQ.
2005	(44)	Ambrisentan	64	Randomized, double-blind, dose-ranging	IPAH, CTD, HIV, AA	II, III	12 wk	36m	VAS	Significant improvement in VAS compared to baseline.
2005	(45)	Sildenafil v. Bosentan	26	Randomized, double-blind	IPAH, CTD	III	16 wk	55m	KCCQ	No significant difference in HRQoL scores between treatments by intent-to-treat. Significant difference between treatments by per-protocol analysis.
2006	(46)	Exercise training	30	Randomized, controlled	PAH, CTEPH	II, III, IV	15 wk	111m	SF-36	Significant improvement in SF-36 PCS, MCS, as well as physical functioning, role-physical, social functioning, mental health, and vitality domains.
2007	(47)	Sitaxsentan	42	Post hoc analysis of RCT	CTD	II, III	12 wk	58m	SF-36	Significant improvement in physical functioning and role-physical domains.

Definition of abbreviations: 6MWD = six-minute walk distance; AA = anorexigen-associated pulmonary arterial hypertension; CHD = pulmonary arterial hypertension associated with congenital heart disease; CHQ = Chronic Heart Failure Questionnaire; CTD = pulmonary arterial hypertension associated with connective tissue disease; CTEPH = chronic thromboembolic pulmonary hypertension; EQ-5D = EuroQol; HIV = pulmonary arterial hypertension associated with human immunodeficiency virus; IPAH = idiopathic pulmonary arterial hypertension; KCCQ = Kansas City Cardiomyopathy Questionnaire; MLHFQ = Minnesota Living with Heart Failure Questionnaire; NHP = Nottingham Health Profile; NYHA = New York Heart Association; PAH = Pulmonary arterial hypertension; PCS = Physical Component Summary; RCT = randomized controlled trial; SF-36 = Medical Outcome Study 36-item Short Form Health Survey; VAS = visual analog score.

step toward designing new PROs tailored to target particular aspects of the disease or treatment effects. Determining the path by which certain factors lead to HRQoL impairment in PAH (above and beyond dyspnea) may also point toward new areas for intervention. In addition, they may help inform decisions regarding the choice of available medical therapies, as well as the appropriate timing of lung transplant.

As discussed, questions regarding the responsiveness and interpretation of HRQoL measures in PAH also remain. Assessing the responsiveness of instruments in clinical studies will rely in part on future trial designs and the thoroughness with which HRQoL outcomes are reported. Proof of statistical significance alone is no longer sufficient. Additional research is needed to begin establishing MID estimates for important PRO measures. Generally, this requires triangulation of results using different types of methods (e.g., distribution- versus anchor-based) (34). Alternative approaches to reporting of PRO results should also be considered. For example, defining results in terms of the number of responders may be more directly

interpretable than reporting an absolute change in score based on an unfamiliar metric (4, 35). Such approaches may be controversial, however, depending on how a responder is defined. When PROs are used to support a labeling claim, extensive pre-testing of instruments in Phase II trials is often necessary before inclusion in pivotal Phase III trials. Open discussion with the FDA is also strongly advised to pre-specify regulatory requirements. Establishing partnerships between academia and industry may facilitate further research by providing access to valuable PRO data from placebo-controlled trials.

Until there is consensus regarding which PROs are best and for which purpose, direct comparisons of the different measures and their performance characteristics in PAH will be essential. The translation and cultural adaptation of existing instruments for use in other languages and countries presents another area of much-needed research. Finally, there is growing evidence from other conditions that PROs can be a useful tool in medical decision making and may facilitate physician-patient commu-

nication (36). Whether PROs have the potential to serve a similar role in the management of PAH remains open for investigation.

CONCLUSIONS

With the continued development of new therapies for PAH, PROs measures are likely to play a greater role in future clinical trial designs. HRQoL, in particular, has emerged as an important clinical endpoint in PAH. HRQoL measures provide complementary information on treatment effects that may be missed by intermediary outcomes. To date, only a limited number of instruments have been adequately evaluated, although our experience with these measures continues to grow. Generic instruments such as the SF-36 are useful when assessing the global impact of an intervention and comparing outcomes to those of other conditions. Studies using the SF-36 support its general validity in PAH; however, its performance in clinical trials suggests that it is only modestly responsive to changes in health status. Of condition-specific measures used for PAH, the MLHFQ has been the most thoroughly investigated. Studies support its reliability and construct validity in PAH, but evidence of its responsiveness in clinical trials remains limited. Despite data supporting the measurement properties of the MLHFQ, the appropriateness of its item content in patients with PAH has not been well studied. The CAMPHOR, in contrast, was specifically developed for use in patients with PAH. Its reliability and validity, both in the United Kingdom and the United States, have been established. As with other instruments, its responsiveness in the clinical trial setting has yet to be determined. Until now, lack of consensus on which measures to use has made it difficult to compare HRQoL results across different trials. The CAMPHOR therefore holds significant promise. The expectation, however, that any single PRO measure will suffice may be unrealistic, given that different instruments are suited for different needs, depending the study design, nature of the intervention, and the target population. Regardless, further work is ultimately needed to begin standardizing the reporting and interpretation of HRQoL in clinical trials for PAH.

Conflict of Interest Statement: H.C. serves as a consultant to United Therapeutics Corp. D.B.T. received \$130,000 as research grants from Actelion for participation in multicentered clinical studies. R.L.D. is an employee of Gilead Sciences and received \$8,000 from Actelion in 2007; \$3,000 from Encysive in 2007, and \$6,000 in 2006; and \$1,500 from Gilead in 2007.

Acknowledgment: The authors thank Patricia P. Katz (Institute for Health Policy Studies, University of California, San Francisco) for her thorough review of the manuscript.

References

- Hoepfer MM, Oudiz RJ, Peacock A, Tapson VF, Haworth SG, Frost AE, Torbicki A. End points and clinical trial designs in pulmonary arterial hypertension: clinical and regulatory perspectives. *J Am Coll Cardiol* 2004;43:48S–55S.
- Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, Rothman M. Recommendations on health-related quality of life research to support labeling and promotional claims in the united states. *Qual Life Res* 2000;9:887–900.
- Schipper H, Clinch J, Powell V. Definitions and conceptual issues. In: Spilker B, editor. *Quality of life assessments in clinical trials*. New York: Raven Press; 1990. pp. 11–24.
- US Food and Drug Administration. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. Draft Guidance. 2006 [accessed 2007 Sep 2]. Available from: <http://www.fda.gov/cder/guidance/5460dft.pdf>
- Doward LC, McKenna SP. Defining patient-reported outcomes. *Value Health* 2004;7:S4–S8.
- Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 1982;14:377–381.
- Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: a conceptual model of patient outcomes. *JAMA* 1995; 273:59–65.
- Taichman DB, Shin J, Hud L, Archer-Chicko C, Kaplan S, Sager JS, Gallop R, Christie J, Hansen-Flaschen J, Palevsky H. Health-related quality of life in patients with pulmonary arterial hypertension. *Respir Res* 2005;6:92.
- Cenedese E, Speich R, Dorschner L, Ulrich S, Maggiorini M, Jenni R, Fischler M. Measurement of quality of life in pulmonary hypertension and its significance. *Eur Respir J* 2006;28:808–815.
- Chua R, Keogh AM, Byth K, O'Loughlin A. Comparison and validation of three measures of quality of life in patients with pulmonary hypertension. *Intern Med J* 2006;36:705–710.
- Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. *Chest* 1988;93:580–586.
- Curtis JR, Martin DP, Martin TR. Patient-assessed health outcomes in chronic lung disease: what are they, how do they help us, and where do we go from here? *Am J Respir Crit Care Med* 1997;156:1032–1039.
- Jones PW. Health status: what does it mean for payers and patients? *Proc Am Thorac Soc* 2006;3:222–226.
- Guyatt GH, Jaeschke R, Feeney DH, Patrick DL. Measurement in clinical trials: choosing the right approach. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*, 2nd edition. Philadelphia: Lippincott-Raven Publishers; 1996. pp. 41–48.
- Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*, 2nd edition. Philadelphia: Lippincott-Raven Publishers; 1996. pp. 49–56.
- Ware JE Jr, Sherbourne CD. The mos 36-item short-form health survey (sf-36). I. Conceptual framework and item selection. *Med Care* 1992; 30:473–483.
- Hunt SM, McEwen J. The development of a subjective health indicator. *Sociol Health Illn* 1980;2:231–246.
- The euroqol group. Euroqol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- Hawthorne G, Richardson J, Osborne R. The assessment of quality of life (aqol) instrument: a psychometric measure of health-related quality of life. *Qual Life Res* 1999;8:209–224.
- Abbott J, Hart A. Measuring and reporting quality of life outcomes in clinical trials in cystic fibrosis: a critical review. *Health Qual Life Outcomes* 2005;3:19.
- Swigris JJ, Kuschner WG, Jacobs SS, Wilson SR, Gould MK. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax* 2005;60:588–594.
- De Vries J, Drent M. Quality of life and health status in sarcoidosis: a review. *Semin Respir Crit Care Med* 2007;28:121–127.
- Shafazand S, Goldstein MK, Doyle RL, Hlatky MA, Gould MK. Health-related quality of life in patients with pulmonary arterial hypertension. *Chest* 2004;126:1452–1459.
- Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum Associates; 1988.
- Zlupko M, Harhay MO, Gallop R, Shin J, Archer-Chicko C, Patel R, Palevsky HI, Taichman DB. Evaluation of disease-specific health-related quality of life in patients with pulmonary arterial hypertension. *Respir Med* (In Press).
- Keogh AM, McNeil KD, Wlodarczyk J, Gabbay E, Williams TJ. Quality of life in pulmonary arterial hypertension: improvement and maintenance with bosentan. *J Heart Lung Transplant* 2007;26:181–187.
- Hawthorne G, Osborne R. Population norms and meaningful differences for the assessment of quality of life (aqol) measure. *Aust N Z J Public Health* 2005;29:136–142.
- McKenna SP, Doughty N, Meads DM, Doward LC, Pepke-Zaba J. The cambridge pulmonary hypertension outcome review (camphor): a measure of health-related quality of life and quality of life for patients with pulmonary hypertension. *Qual Life Res* 2006;15:103–115.
- McKenna SP, Doward LC. The needs-based approach to quality of life assessment. *Value Health* 2004;7:S1–S3.
- Gomberg-Maitland M, Thenappan T, Rizvi K, Chandra S, Meads DM, McKenna SP. United states validation of the cambridge pulmonary hypertension outcome review (camphor). *J Heart Lung Transplant* 2008;27:124–130.
- Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003;12:349–362.

32. Flattery MP, Pinson JM, Savage L, Salyer J. Living with pulmonary artery hypertension: patients' experiences. *Heart Lung* 2005;34:99-107.
33. White J, Hopkins RO, Glissmeyer EW, Kitterman N, Elliott CG. Cognitive, emotional, and quality of life outcomes in patients with pulmonary arterial hypertension. *Respir Res* 2006;7:55.
34. Lydick E, Epstein RS. Interpretation of quality of life changes. *Qual Life Res* 1993;2:221-226.
35. Schunemann HJ, Guyatt GH. Interpreting the results of quality-of-life measures in clinical trials: the clinician's perspective. In: Lenderking WR, Revicki DA, editors. *Advancing health outcomes research methods and clinical applications*. McLean, VA: Degnon Associates, Inc., 2005. pp. 201-212.
36. National asthma education and prevention program. Expert panel report 3 full report 2007. Guidelines for the diagnosis and management of asthma: measures of asthma assessment and monitoring. National heart, lung, and blood institute. 2007 [cited 2007 Sep 2]. Available from: http://www.nhlbi.nih.gov/guidelines/asthma/04_sec3_comp.pdf.
37. Barst RJ, Rubin LJ, Long WA, McGoon MD, Rich S, Badesch DB, Groves BM, Tapson VF, Bourge RC, Brundage BH, et al. A comparison of continuous intravenous epoprostenol (prostacyclin) with conventional therapy for primary pulmonary hypertension: the primary pulmonary hypertension study group. *N Engl J Med* 1996; 334:296-302.
38. Olschewski H, Simonneau G, Galie N, Higenbottam T, Naeije R, Rubin LJ, Nikkho S, Speich R, Hoeper MM, Behr J, et al. Inhaled iloprost for severe pulmonary hypertension. *N Engl J Med* 2002;347:322-329.
39. Simonneau G, Barst RJ, Galie N, Naeije R, Rich S, Bourge RC, Keogh A, Oudiz R, Frost A, Blackburn SD, et al. Continuous subcutaneous infusion of treprostinil, a prostacyclin analogue, in patients with pulmonary arterial hypertension: a double-blind, randomized, placebo-controlled trial. *Am J Respir Crit Care Med* 2002;165:800-804.
40. Barst RJ, McGoon M, McLaughlin V, Tapson V, Rich S, Rubin L, Wasserman K, Oudiz R, Shapiro S, Robbins IM, et al. Beraprost therapy for pulmonary arterial hypertension. *J Am Coll Cardiol* 2003; 41:2119-2125.
41. Sastry BK, Narasimhan C, Reddy NK, Raju BS. Clinical efficacy of sildenafil in primary pulmonary hypertension: a randomized, placebo-controlled, double-blind, crossover study. *J Am Coll Cardiol* 2004;43: 1149-1153.
42. Barst RJ, Langleben D, Frost A, Horn EM, Oudiz R, Shapiro S, McLaughlin V, Hill N, Tapson VF, Robbins IM, et al. Sitaxsentan therapy for pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2004;169:441-447.
43. Oudiz RJ, Schilz RJ, Barst RJ, Galie N, Rich S, Rubin LJ, Simonneau G. Treprostinil, a prostacyclin analogue, in pulmonary arterial hypertension associated with connective tissue disease. *Chest* 2004;126:420-427.
44. Galie N, Badesch D, Oudiz R, Simonneau G, McGoon MD, Keogh AM, Frost AE, Zwicke D, Naeije R, Shapiro S, et al. Ambrisentan therapy for pulmonary arterial hypertension. *J Am Coll Cardiol* 2005;46:529-535.
45. Wilkins MR, Paul GA, Strange JW, Tunariu N, Gin-Sing W, Banya WA, Westwood MA, Stefanidis A, Ng LL, Pennell DJ, et al. Sildenafil versus endothelin receptor antagonist for pulmonary hypertension (SERAPH) study. *Am J Respir Crit Care Med* 2005;171: 1292-1297.
46. Mereles D, Ehlken N, Kreuzer S, Ghofrani S, Hoeper MM, Halank M, Meyer FJ, Karger G, Buss J, Juenger J, et al. Exercise and respiratory training improve exercise capacity and quality of life in patients with severe chronic pulmonary hypertension. *Circulation* 2006;114:1482-1489.
47. Girgis RE, Frost AE, Hill NS, Horn EM, Langleben D, McLaughlin VV, Oudiz RJ, Robbins IM, Seibold JR, Shapiro S, et al. Selective endothelin a receptor antagonism with sitaxsentan for pulmonary arterial hypertension associated with connective tissue disease. *Ann Rheum Dis* 2007; 66:1467-1472.