ORIGINAL RESEARCH

Psychometric Validation of the EQ-5D-3L in Patients with Nontuberculous Mycobacterial (NTM) Lung Disease Caused by *Mycobacterium avium* Complex (MAC)

Anuj Shah[1]
Xinyi Ng [1]
Ruchit Shah [1]
Caitlyn Solem [1]
Ping Wang[2]
Marko Obradovic[3]

[1]Patient Centered Outcomes, Pharmerit – An OPEN Health Company, Bethesda, MD, USA; [2]Health Economics and Outcomes Research, Insmed Incorporated, Bridgewater Township, NJ, USA; [3]Health Economics and Outcomes Research, Insmed Germany GmbH, Frankfurt am Main, Germany

Purpose: This analysis evaluated the psychometric properties of the EQ-5D-3L among patients with *Mycobacterium avium* complex lung disease (MAC-LD).

Methods: Data from the Phase III CONVERT trial were analyzed. Study measures including the EQ-5D-3L, St. George's Respiratory Questionnaire (SGRQ), the 6-minute walk test (6MWT), and percent predicted forced expiratory volume in 1 second (ppFEV$_1$) were collected at baseline, month 3 (M3), month 4 (M4; only 6MWT), and month 6 (M6). Item characteristics and validity were evaluated at baseline. Test–retest reliability was assessed using intraclass correlation coefficients (ICC) and a weighted kappa statistic among a subgroup of stable patients. Validity was tested by evaluating correlations of the EQ-5D-3L index/visual analogue scale (VAS) scores with SGRQ and 6MWT and comparing mean index/VAS scores across known groups defined using 6MWT and ppFEV$_1$. Responsiveness of the EQ-5D-3L was assessed using 6MWT, SGRQ, ppFEV$_1$, and culture conversion as anchors.

Results: The index score was subject to ceiling effects, with 32.6% of patients reporting perfect health at baseline. ICCs for the index (0.80) and VAS (0.85) scores and weighted kappas for the domains (0.5–0.72) indicated adequate test–retest reliability. Correlations between the index/VAS scores and related domains of the SGRQ and 6MWT were as hypothesized (0.31–0.62), and the mean index/VAS scores were significantly different between the 6MWT and ppFEV$_1$ known groups (p<0.05), supporting the validity of the EQ-5D-3L. No evidence was found supporting the responsiveness of the EQ-5D-3L to changes in any of the anchors.

Conclusion: EQ-5D-3L exhibited evidence of validity and reliability but poor responsiveness to clinically meaningful changes in patients with MAC-LD.

Keywords: psychometric validation, NTM lung disease, EQ-5D, reliability, validity, responsiveness

Introduction

Nontuberculous mycobacterial (NTM) lung disease (NTM-LD) is a rare but emerging global health concern.[1,2] Although more than 200 species and subspecies of NTM have been identified, the most predominant species causing NTM-LD is *Mycobacterium avium* complex (MAC).[3,4] MAC lung disease (MAC-LD) can become a chronically progressive and debilitating disease. It can lead to severe lung injury and cavitary lesions and be associated with poor survival prognosis.[5,6] Patients with MAC-LD experience diminished

Correspondence: Marko Obradovic
Senior Director, HEOR Europe, Insmed Germany GmbH, The Square 12, am Flughafen, Frankfurt (Frankfurt am Main), 60549, Germany
Email marko.obradovic@insmed.com

pulmonary function,[7] which could lead to a reduction in patients' health-related quality of life (HRQoL), self-perceived health status, and functional abilities.[8,9] Additionally, treatment of MAC-LD necessitates the use of a combination of antibiotics, often given over 18 months or more; thus the risk of treatment-related toxicities is relatively high, which can further worsen patients' HRQoL.[5,10]

To accurately assess a disease's impact on patients' HRQoL, it is imperative to select a valid and reliable HRQoL instrument. In the absence of any MAC-LD-specific HRQoL instrument, previous studies have used generic measures, such as EuroQol 5-dimensional 3-level (EQ-5D-3L), to assess the HRQoL of MAC-LD patients.[8,9,11] However, limited information is available regarding the psychometric properties of EQ-5D-3L in MAC-LD. In order to build on this gap in evidence, the current study aimed to investigate the reliability, validity, and responsiveness of the EQ-5D-3L among patients with MAC-LD who were refractory to guideline-based treatment (GBT).

Methods
Study Design
This analysis includes data from CONVERT (INS-212; NCT02344004), a Phase III, randomized, open-label clinical trial that studied the effectiveness of adding amikacin liposome inhalation suspension (ALIS) to GBT in subjects with MAC-LD who were refractory to GBT.[12] This post hoc analysis uses the HRQoL and relevant clinical data collected from baseline to month 6 to evaluate the psychometric properties of the EQ-5D-3L among patients enrolled in the CONVERT study.

The CONVERT study enrolled adults with active MAC-LD, according to the American Thoracic Society and the Infectious Diseases Society of America criteria. Patients were required to be MAC-positive while on GBT ≥6 months and to have been on GBT within 12 months before screening.[12] Patients were randomized 2:1 to receive ALIS and GBT or GBT alone. Details of the CONVERT trial and the results have been published elsewhere.[12] All patients in the intent-to-treat (ITT) population who had a baseline assessment of the EQ-5D-3L in the trial were included in this analysis.

Study Measures
Clinical Measures
Three clinical measures recorded in the trial were used in this analysis: sputum culture conversion, six-minute walk test (6MWT), and percent predicted forced expiratory

volume in 1 second (ppFEV$_1$). The proportion of patients achieving sputum culture conversion within the first 6 months was the primary endpoint assessed in the CONVERT trial, and was based on patient sputum cultures collected monthly from baseline through month 6.[12] Culture conversion was achieved if patients had 3 consecutive negative sputum cultures. The 6MWT provided a measure of the patient's exertional capacity by requiring the patient to walk a prescribed course on a flat surface while recording the distance covered in 6 minutes in meters, with higher scores representing better capacity.[13] The 6MWT was administered at baseline, month 4, and month 6. The FEV$_1$ is a lung function test that measures the "maximal volume of air exhaled in the first second of a forced expiration from a position of full inspiration."[14] The ppFEV$_1$ (predicted for age, race, gender, and height) was generated using the prediction equations from Hankinson et al for baseline and month 6.[15]

Health-Related Quality of Life (HRQoL) Measures
The 2 HRQoL measures used in this analysis were the EQ-5D-3L and the St. George Respiratory Questionnaire (SGRQ). The EQ-5D-3L is frequently used for both clinical and economic appraisals and allows for comparisons between different disease states.[16] It is a self-administered questionnaire that includes 2 components: a descriptive profile and a visual analogue scale (VAS).[17] The descriptive profile measures 5 single-item health dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 3 levels of response: no problems, some problems, and extreme problems. The responses to each item were used to generate a population-normed index summary score, anchored at 0 for death and 1 for perfect health, based on the published United Kingdom time trade-off value set.[18] The VAS assesses a respondent's self-perceived current general health status on a 100-point scale, with 0 representing "worst imaginable health state" and 100 representing "best imaginable health state." The EQ-5D-3L was administered at baseline, month 3, and month 6. The SGRQ is a self-administered 50-item questionnaire evaluating 3 health domains: symptoms, activity, and impact.[19] Each domain is scored from 0 to 100, with a lower score indicating better HRQoL, and the domain scores are combined into a weighted total score ranging from 0 to 100. The SGRQ was administered at baseline, month 3, and month 6.

Statistical Analysis

Item Distribution, Floor and Ceiling Effects

Descriptive statistics (N, mean, standard deviation [SD], median, 25th and 75th percentile, and range) were reported for the EQ-5D-3L index and VAS scores at baseline, month 3, and month 6. The distribution of responses and missing data for each item was reported at baseline in terms of frequencies and percentages. Additionally, the presence of floor and ceiling effects was examined at baseline for each item. An item was considered to demonstrate a floor or ceiling effect if more than 33% (100/the number of response options on the item, ie =100/3) of responses were in the lowest or highest response categories.[20] The proportion of patients reporting the lowest and maximum possible values on the EQ-5D-3L index and VAS scores at baseline were also reported.

Test–Retest Reliability

Test–retest reliability is the ability of an instrument to reliably replicate the results more than once (ie, at two or more different time points) in the same population.[21] Test–retest reliability of the EQ-5D-3L in this analysis was evaluated by comparing the EQ-5D-3L responses at month 3 and month 6 among stable patients. A patient was considered to be stable if they met all of the following criteria: (1) change of <25 meters in the absolute 6MWT score between month 4 and month 6, (2) an absolute change of <4 points in the composite score of the SGRQ between month 3 and month 6, and (3) no change in culture converter status between month 3 and month 6 (ie, converted patients who do not have a relapse or non-converters who do not convert). Intraclass correlation coefficients (ICCs) along

with 95% confidence intervals (CIs) were calculated for the EQ-5D-3L index and VAS scores between the 2 time points (ie, month 3 and month 6). The reliability of single-item dimensions of the EQ-5D-3L were evaluated using the weighted kappa statistic between month 3 and month 6. ICC values >0.75 and weighted kappa values >0.4 were considered to indicate acceptable test–retest reliability.[22]

Validity

Known-groups validity is the ability of an instrument to differentiate among patients based on their disease severity.[23] Known-groups validity was assessed for the EQ-5D-3L index and VAS scores, grouping patients based on their ppFEV$_1$ (≤80%, >80%) and separately based on their 6MWT (≥350 m, <350 m).[19,24] Differences between known groups were examined using the Mann–Whitney U-test.

Construct validity is a measure of how well a measure correlates with a previously validated measure.[21] Construct validity (convergent and divergent validity) was tested by examining the strength and direction of the correlation between the EQ-5D-3L individual dimensions, index, and VAS scores with the SGRQ activity scale, SGRQ impact scale, SGRQ symptom scale, SGRQ total score, and the 6MWT. Pearson's correlation was calculated for VAS and index scores, while Spearman's rank correlation was used for the individual dimensions. Correlations between 0.1 and <0.3 were defined as weak, 0.3 and <0.5 as moderate, and ≥0.5 as strong.[25] The hypothesized strength and direction of each of the correlations are depicted in Table 1.

Table 1 Hypothesized Direction and Strength of Correlations Between EQ-5D-3L, SGRQ, and 6MWT

Measure	EQ-5D-3L Dimensions					EQ-5D-3L Score	
	Mobility	**Self-Care**	**Usual Activities**	**Pain/ Discomfort**	**Anxiety/ Depression**	**VAS**	**Index**
SGRQ-Activity	Moderate positive	Weak positive	Moderate positive	Weak positive	Weak positive	Moderate negative	Moderate negative
SGRQ-Impact	Weak positive	Weak positive	Moderate positive	Weak positive	Moderate positive	Moderate negative	Moderate negative
SGRQ-Symptom	Weak positive	Weak positive	Weak positive	Weak positive	Moderate positive	Moderate negative	Moderate negative
SGRQ Total Score	Weak positive	Weak positive	Moderate positive	Weak positive	Weak positive	Moderate negative	Moderate negative
6MWT	Moderate negative	Weak negative	Weak negative	Weak negative	Weak negative	Moderate positive	Moderate positive

Abbreviations: 6MWT, six-minute walk test; EQ-5D-3L, EuroQol 5-dimensional 3-level; SGRQ, St. George Respiratory Questionnaire; VAS, visual analogue scale.

Responsiveness to Change

Responsiveness to change is an instrument's ability to detect a clinically meaningful change over time. In this analysis, meaningful change was defined based on 4 separate anchors: (1) ppFEV$_1$: relative change of >10% versus <5%,[26] (2) SGRQ total score: absolute change of >4 versus <4,[27] (3) 6MWT: absolute change of >25 meters versus <25 meters,[28] and (4) culture conversion status: patients who achieved culture conversion on or before the month 6 visit without relapse versus those who did not achieve culture conversion. Change in anchors and the EQ-5D-3L scores was calculated as the difference between the baseline and month 6 values on the respective measures. The responsiveness of the EQ-5D-3L was assessed by calculating the Guyatt responsiveness statistic and evaluating point biserial correlations. The Guyatt statistic was calculated as mean change in responder groups divided by the SD of change in the non-responder group. The point biserial correlations quantified the correlations between change in the EQ-5D-3L scores and responder categories (responders vs non-responders) based on respective anchors.[29] The Guyatt responsiveness statistic thresholds for weak, moderate, and strong responsiveness were <0.2, 0.2–0.5, and >0.8, respectively,[30] and the point biserial correlations' thresholds for weak, moderate, and strong correlations were 0.1 and <0.3, 0.3 and <0.5, and ≥0.5, respectively.[25] Cumulative distribution function (CDF) plots of change in EQ-5D-3L index scores and VAS scores, stratified on meaningful change for each anchor (conversion status, ppFEV$_1$, SGRQ, and 6WMT), were also plotted to examine the degree of separation between

responders and non-responders at various score change values.

Patients with missing values for a measure at a particular assessment time point were excluded from the analyses involving that measure at that particular time point. For example, if a patient had missing data for the 6MWT but not for the SGRQ at month 6, the patient was excluded for analyses involving 6MWT data at month 6 but not for analyses involving SGRQ data at month 6. All analyses were performed using SAS 9.4.

Results

Descriptive Analysis

Out of the ITT population (N=336) for the CONVERT trial, only 1 patient did not have a baseline EQ-5D-3L assessment and hence was excluded in this analysis. The baseline demographic and clinical characteristics of the ITT population have been described in detail elsewhere.[12] Among the 335 patients with a baseline EQ-5D-3L assessment, 4 patients had missing data on 1 or more of the single-item dimensions, thereby preventing the calculation of an index score.

Minimal changes in the mean index and VAS scores from baseline (Index: 0.82; VAS: 65.88) to month 6 (Index: 0.83; VAS: 68.19) were observed (Table 2).

The distribution of responses to each single-item dimension at baseline is shown in Figure 1. No floor effects were observed among any of the EQ-5D-3L items. However, substantial ceiling effects were seen for all the items, with >50% of the respondents reporting "No

Table 2 EQ-5D-3L Index and VAS Scores at Baseline, Month 3, and Month 6

Visit	Parameter	EQ-5D-3L Index Score	EQ-5D-3L VAS
Baseline	N Mean (SD) Median (Q1 to Q3) Range	331 0.82 (0.17) 0.80 (0.80 to 1.00) 0.00 to 1.00	335 65.9 (21.4) 70.0 (52.0 to 80.0) 4.0 to 100.0
Month 3	N Mean (SD) Median (Q1 to Q3) Range	299 0.82 (0.17) 0.80 (0.80 to 1.00) 0.20 to 1.00	300 67.7 (19.4) 70.0 (56.5 to 80.0) 6.0 to 100.0
Month 6	N Mean (SD) Median (Q1 to Q3) Range	275 0.83 (0.17) 0.80 (0.80 to 1.00) 0.20 to 1.00	275 68.2 (19.9) 70.0 (55.0 to 83.0) 4.0 to 100.0

Abbreviations: EQ-5D-3L, EuroQol 5-dimensional 3-level; N, number of patients; SD, standard deviation; Q1, lower quartile; Q3, upper quartile; VAS, visual analogue scale.
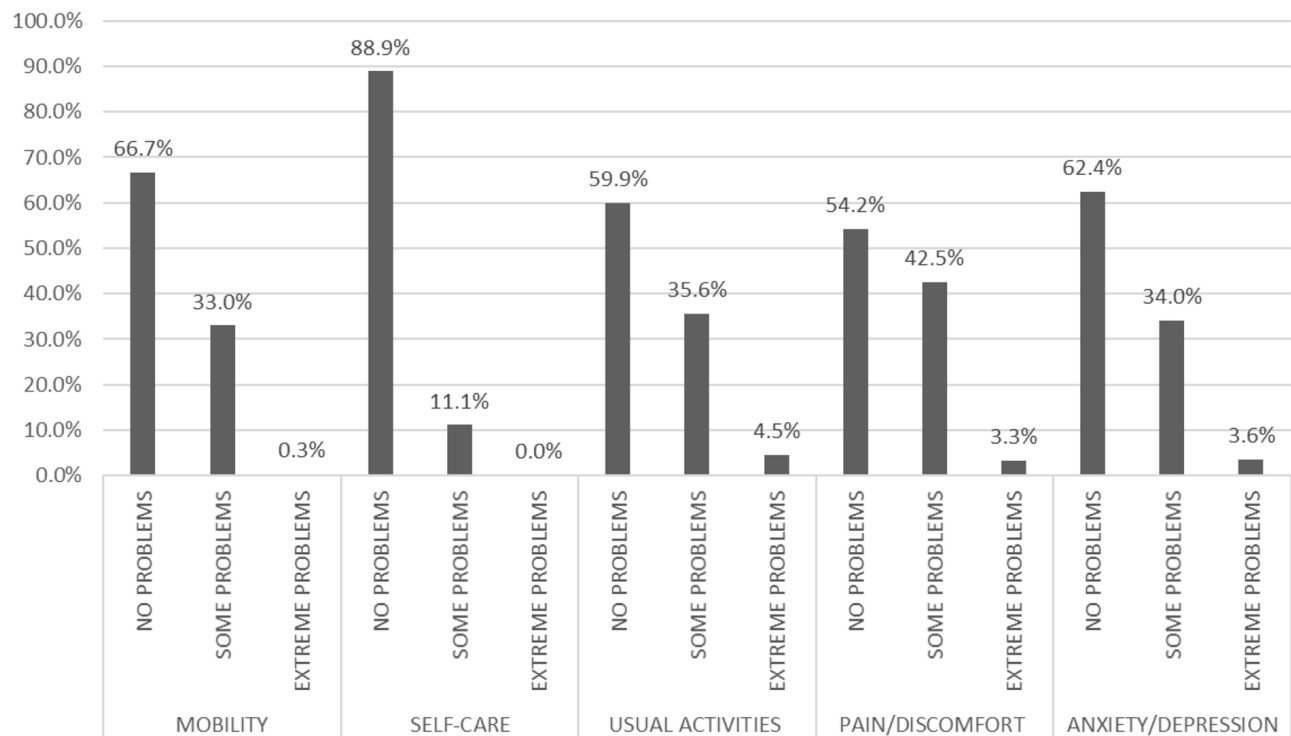
Figure 1 Distribution of responses to each of the EQ-5D-3L items at baseline.
Abbreviation: EQ-5D-3L, EuroQol 5-dimensional 3-level.

problems" for each item. None of the respondents had the lowest possible index or VAS score, but 32.6% of respondents had the maximum possible index score of 1, representing perfect health. Minimal ceiling effects were seen for the VAS, with only 2.4% of respondents reporting a perfect score of 100.

Test–Retest Reliability

50 patients were classified as stable, according to our criteria using culture conversion, 6MWT, and SGRQ. Among the stable patients, the ICCs for both the index (0.802) and VAS (0.852) demonstrated adequate test–retest reliability.[22] Likewise, across all items, the kappa statistic was >0.4, indicating acceptable test–retest reliability (Table 3).

Validity

The index scores were able to differentiate between the hypothesized groups (p<0.05) based on 6MWT (Figure 2A) and ppFEV$_1$ (Figure 2B), thereby suggesting adequate known-groups validity of index scores.

Similarly, statistically significant differences in the VAS scores were also observed between the known groups defined by 6MWT (<350m: 57.82; ≥350m: 68.79; p<0.001) and ppFEV$_1$ (≤80%: 60.38; >80%: 69.23; p<0.001).

Correlations of the EQ-5D-3L single-item domains, index, and VAS scores with the SGRQ domains, SGRQ total score, and the 6MWT were largely as hypothesized, supporting the instrument's construct validity (Table 4).

Responsiveness to Change

Point biserial correlations between change over time in the response categories and the EQ-5D-3L index were small (Table 5). Similarly, based on the estimated Guyatt statistics, EQ-5D-3L index scores did not appear responsive to

Table 3 Test–Retest Reliability of the EQ-5D-3L Items, Index Score, and VAS Score

| EQ-5D-3L Item	N	Weighted Kappa	95% CI
Mobility	50	0.54	0.28–0.81
Self-Care	50	1	1–1
Usual Activities	50	0.72	0.50–0.94
Pain/Discomfort	50	0.50	0.26–0.74
Anxiety/Depression	50	0.60	0.35–0.85
EQ-5D-3L Scores	**N**	**ICC**	**95% CI**
EQ-5D-3L Index	50	0.80	0.68–0.88
EQ-5D-3L VAS	50	0.85	0.75–0.91

Abbreviations: CI, confidence interval; EQ-5D-3L, EuroQol 5-dimensional 3-level; ICC, intraclass correlation coefficient; VAS, visual analogue scale.

A EQ-5D-3L index scores across 6MWT groups
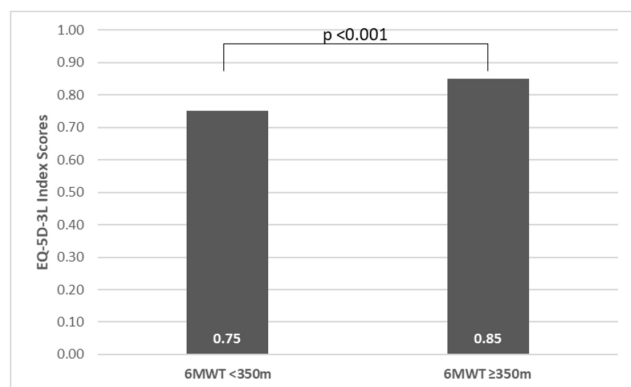
B EQ-5D-3L index scores across ppFEV^1 groups
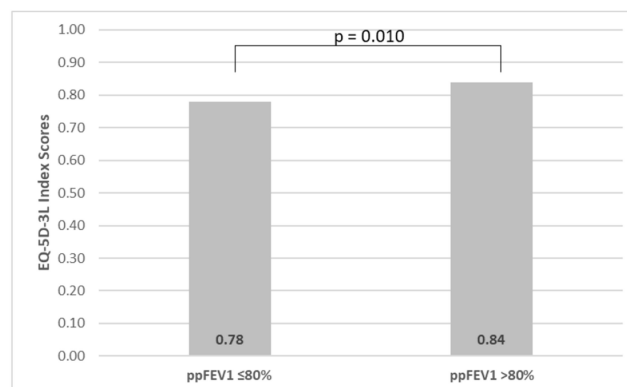
Figure 2 EQ-5D-3L index scores across known groups defined based on (**A**) 6MWT and (**B**) ppFEV$_1$.
Abbreviations: 6MWT, six-minute walk test; EQ-5D-3L, EuroQol 5-dimensional 3-level; ppFEV$_1$, percent predicted forced expiratory volume in 1 second.

meaningful changes defined by culture conversion status, ppFEV$_1$, or 6MWT (Table 5). The EQ-5D VAS scores (results in Appendix Table 1) also displayed small to moderate responsiveness to change in culture conversion, 6MWT, and SGRQ.

CDF plots of change in EQ-5D-3L index scores, stratified on meaningful change for each anchor (conversion status, ppFEV$_1$, SGRQ, and 6WMT) are shown in Appendix Figure 1. The EQ-5D-3L index scores could not discriminate well between patients who had a meaningful change on the anchors versus those that did not, as indicated by the large degree of overlap in the curves. The median change in the EQ-5D-3L index scores was 0 for both groups, further showing the non-responsiveness of the index scores to meaningful changes in the anchors. The CDF plots for change in EQ-5D-3L VAS scores also showed that the change scores were similar among responders and non-responders for all anchors (Appendix Figure 2).

Discussion

In this study, we investigated the psychometric properties of the EQ-5D-3L among patients with MAC-LD who are refractory to GBT. Item and summary score distributions showed that the EQ-5D-3L VAS score had limited ceiling effects, but the individual domains and utility scores were subject to substantial ceiling effects. These findings complement a recent study by Hong et al, which also found ceiling effects in EQ-5D-3L domains and utility scores, with >70% of MAC-LD patients reporting no problems on individual domains and utility scores ranging from 0.91 to 1.00.[11] Additionally, we found that while the EQ-5D-3L is a valid and reliable instrument, it may not be responsive to clinically important changes among MAC-LD patients refractory to GBT.

We tested the reliability of the domains using the weighted kappa statistic and the index and VAS scores using the ICC. Both measures of reliability exceeded the thresholds for good reliability, with the weighted kappa for

Table 4 Spearman Correlations Between EQ-5D-3L, SGRQ, and 6MWT

Measure	EQ-5D-3L Single Item Dimensions					EQ-5D-3L Score	
	Mobility	Self-Care	Usual Activities	Pain/Discomfort	Anxiety/Depression	VAS	Index Score
SGRQ-Activity	0.36	0.19	0.35	0.25	0.17	−0.43	−0.55
SGRQ-Impact	0.28	0.13	0.29	0.26	0.19	−0.48	−0.61
SGRQ-Symptom	0.26	0.10	0.26	0.24	0.13	−0.39	−0.44
SGRQTotal Score	0.34	0.17	0.34	0.29	0.20	−0.50	−0.62
6MWT	−0.23	−0.14	−0.28	−0.19	−0.13	0.31	0.39

Notes: Higher scores on SGRQ and EQ-5D-3L items indicate worse symptoms. Higher 6MWT and EQ-5D index and VAS scores indicate better health-related quality of life. All correlations were statistically significant at α = 0.05.
Abbreviations: 6MWT, 6-minute walk test; EQ-5D-3L, EuroQol 5-dimensional 3-level; SGRQ, St. George's Respiratory Questionnaire; VAS, visual analogue scale.

Table 5 Point Biserial Correlations and Guyatt Responsiveness Statistics for EQ-5D-3L Index Scores by Each of the Anchors

Anchor	Point Biserial Correlations	Mean Change in Responders' Scores	SD of Change in Non-Responders' Scores	Guyatt Statistic
Culture Conversion	0.04	0.00	0.13	−0.01
ppFEV$_1$	0.03	0.00	0.09	0.03
6MWT	0.11	0.01	0.10	0.08
SGRQ	0.21	0.03	0.10	0.29

Abbreviations: 6MWT, 6-minute walk test; SD, standard deviation; SGRQ, St. George's Respiratory Questionnaire; ppFEV$_1$, percent predicted forced expiratory volume in 1 second.

domains ranging from 0.50 to 1.00 and the ICCs for index and VAS scores being >0.8. These findings support the reliability of the EQ-5D-3L and suggest that it has good reproducibility and is unlikely to be subject to measurement error.[31] Similar performance of the EQ-5D-3L has been noted in previous studies among patients with other chronic respiratory diseases, such as chronic obstructive pulmonary disorder and asthma.[32]

We studied the construct validity of the EQ-5D-3L instrument by testing its concurrent and known-groups validity. The concurrent validity was supported with the correlations between the index and VAS scores and 6MWT and SGRQ falling in hypothesized ranges. The correlations reported in our study were similar to the estimates previously reported in the COPD population.[33,34] Specifically, Lin et al found a correlation of 0.29 and 0.45 of 6MWT scores with the EQ-5D-3L VAS, and Nolan et al reported correlations between SGRQ sub-scales and total scores, with EQ-5D-3L index and VAS scores ranging from 0.3 to 0.6.[33,34] Our results also demonstrate evidence supporting the known-groups validity of the EQ-5D-3L, with the index and VAS scores successfully differentiating between patient groups defined using 6WMT and ppFEV$_1$ thresholds. This finding is in agreement with previous studies which have shown that the EQ-5D-3L index and VAS scores were able to discriminate COPD patients based on their ppFEV$_1$ values.[35,36]

However, the EQ-5D-3L index or VAS scores did not demonstrate good responsiveness in the MAC-LD population. The low point biserial correlations as well as weak Guyatt responsiveness statistics showed that detectable changes in EQ-5D-3L index or VAS scores were not observed for clinically significant changes over time detected on key study anchors. Hence, the EQ-5D-3L may not be a suitable instrument to detect longitudinal changes in HRQoL among MAC-LD patients refractory to GBT. The lack of responsiveness to changes in the ppFEV$_1$ is not surprising considering that longitudinal changes in HRQoL and FEV$_1$ have a poor relationship in other chronic respiratory conditions as well.[37,38] The lack of responsiveness of the EQ-5D-3L scores to sputum culture status is noteworthy, given that sputum culture samples are regarded as the primary biomarkers for assessing treatment efficacy in this population.[12,39] These findings are in line with Hong et al, who also reported that the EQ-5D-3L index scores were unable to distinguish between MAC-LD patients and healthy controls.[11] Previous studies assessing the responsiveness of EQ-5D-3L in other respiratory conditions, such as COPD and asthma, also conclude that the EQ-5D-3L is better suited to detecting change in moderate-to-severe patients, with limited responsiveness in mild disease states.[40–42]

The lack of responsiveness of the EQ-5D-3L index to changes in the anchors may be partially attributable to the high ceiling effects (~33% of patients reported perfect health at baseline), which prevented us from detecting further improvements in these patients. Another potential reason for the lack of responsiveness could be that the range of responses for each dimension of the EQ-5D-3L may be too narrow to capture changes in HRQoL in the MAC-LD population.[43] This problem could be circumvented in future studies by opting to use the EQ-5D-5L in favor of the 3L. Nonetheless, a recent study by Nolan et al among COPD patients showed weak correlations between changes in EQ-5D-5L utility scores, with clinically significant changes in these patients.[34] Hoogendoorn et al proposed another potential solution for future consideration, which involved the addition of a respiratory dimension to the EQ-5D (3L or 5L).[44] They found that addition of a "breathing problem" dimension significantly improved the responsiveness of the EQ-5D to clinically significant changes in chronic lung disease patients.[44] However, such an approach would require valuation of the new dimension in order to obtain preference-based utility scores, which can be burdensome in terms of time and resources. Additionally, the comparability of the

utilities obtained using this instrument with established national tariffs may also be questionable.

The results of this study are subject to some limitations. There is no "gold standard" measure of HRQoL in MAC-LD patients; hence we used a range of HRQoL and clinical measures available to test the psychometric properties of the EQ-5D-3L. The minimal clinically important difference thresholds for each anchor used in the responsiveness evaluation have been established among COPD and asthma patients and may not be applicable to MAC-LD patients. Another limitation is the limited duration of the study (6 months). Improvements in HRQoL as a result of culture conversion may take longer than 6 months to be realized. However, since CONVERT patients who did not achieve culture conversion by month 6 were either transitioned to a single-arm study or lost to follow-up, we were unable to test responsiveness using longer-term data. Lastly, our population of MAC-LD patients was restricted to those who were refractory to GBT, indicating that they had had MAC-LD for a considerably long duration (5.7 years on average). Previous studies have reported that patients with chronic disease tend to adapt to their condition,[45] which may have resulted in the high ceiling effects at baseline we observed in this study. Additionally, since our study population included only MAC-LD patients refractory to GBT, our findings may not be generalizable to the overall MAC-LD population.

Conclusions

Within this cohort of MAC-LD patients refractory to GBT, the EQ-5D-3L exhibited evidence of validity and reliability but may not be best suited to studying changes in HRQoL over time. The lack of responsiveness of the EQ-5D-3L may partially be attributable to high ceiling effects at baseline. Overall, this study highlights the need to develop HRQoL measures that can accurately capture and correlate with clinical improvements in the MAC-LD population. With the growing importance of HRQoL measures in the context of drug development and regulation,[14] evaluating existing measures as well as identifying or developing new measures that can capture the HRQoL changes in this population is an important unmet need.

Data Sharing Statement

Analytic datasets can be made available upon a reasonable request to the corresponding author.

Ethical Approval

All procedures performed in the CONVERT trial involving human participants were in accordance with Good Clinical Practice, following local regulations and ethical principles described in the Declaration of Helsinki. The CONVERT trial was conducted in 18 countries at a total of 127 sites. The study protocol, informed consent form (ICF), and all other appropriate, related materials were reviewed and approved by an Independent Ethics Committee (IEC) or Institutional Review Board (IRB) for each site. A list of all IECs/IRBs has been included as Appendix Table 2.

Informed Consent

Informed consent was obtained from all individual participants included in the CONVERT trial.

Disclosure

AS, XN, RS, and CS are full-time employees of Pharmerit, the institution which received funding from Insmed Incorporated for conducting the study. PW and MO are full-time employees of Insmed Incorporated. The authors report no other conflicts of interest in this work.

References

1. Griffith DE, Aksamit TR. Therapy of refractory nontuberculous mycobacterial lung disease. *Curr Opin Infect Dis*. 2012;25(2):218–227. doi:10.1097/QCO.0b013e3283511a64
2. Prevots DR, Marras TK. Epidemiology of human pulmonary infection with nontuberculous mycobacteria: a review. *Clin Chest Med*. 2015;36 (1):13–34. doi:10.1016/j.ccm.2014.10.002
3. Stout JE, Koh WJ, Yew WW. Update on pulmonary disease due to non-tuberculous mycobacteria. *Int j Infect Dis*. 2016;45:123–134.
4. Hoefsloot W, van Ingen J, Andrejak C, et al. The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: an NTM-NET collaborative study. *Eur Respir J*. 2013;42(6):1604–1613.
5. Johnson MM, Odell JA. Nontuberculous mycobacterial pulmonary infections. *J Thorac Dis*. 2014;6(3):210–220.
6. Diel R, Lipman M, Hoefsloot W. High mortality in patients with Mycobacterium avium complex lung disease: a systematic review. *BMC Infect Dis*. 2018;18(1):206.

7. Park HY, Jeong BH, Chon HR, Jeon K, Daley CL, Koh WJ. Lung Function Decline According to Clinical Course in Nontuberculous Mycobacterial Lung Disease. *Chest*. 2016;150(6):1222–1232.

8. Mehta M, Marras TK. Impaired health-related quality of life in pulmonary nontuberculous mycobacterial disease. *Respir Med*. 2015;105(11):1718–1725.

9. Asakura T, Funatsu Y, Ishii M, et al. Health-related quality of life is inversely correlated with C-reactive protein and age in Mycobacterium avium complex lung disease: a cross-sectional analysis of 235 patients. *Respir Res*. 2015;16:145.

10. Griffith DE, Aksamit T, Brown-Elliott BA, et al. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med*. 2007;175 (4):367–416. doi:10.1164/rccm.200604-571ST

11. Hong JY, Lee SA, Kim SY, et al. Factors associated with quality of life measured by EQ-5D in patients with nontuberculous mycobacterial pulmonary disease. *Quality Life Res*. 2014;23(10):2735–2741. doi:10.1007/s11136-014-0727-3

12. Griffith DE, Eagle G, Thomson R, et al. Amikacin Liposome Inhalation Suspension for Treatment-Refractory Lung Disease Caused by Mycobacterium avium Complex (CONVERT). A Prospective, Open-Label, Randomized Study. *Am J Respir Crit Care Med*. 2018;198 (12):1559–1569. doi:10.1164/rccm.201807-1318OC

13. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002;166(1):111–117. doi:10.1164/ajrccm.166.1.at1102

14. Miller MR, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319–338. doi:10.1183/09031936.05.00034805

15. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med*. 1999;159(1):179–187. doi:10.1164/ajrccm.159.1.9712108

16. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208. doi:10.1016/0168-8510(90)90421-9

17. American Society for Radiation Oncology (ASTRO). Coding Tip: gynecologic Brachytherapy Procedures. Available from: https://www.astro.org/Daily-Practice/Coding/Coding-Guidance/FAQ-Coding-Tips. Accessed July 4, 2019.

18. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108. doi:10.1097/00005650-199711000-00002

19. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation: the St. george's Respiratory Questionnaire. *Am Rev Respir Dis*. 1992;145 (6):1321–1327. doi:10.1164/ajrccm/145.6.1321

20. Pompilus F, Burgess S, Hudgens S, Banderas B, Daniels S. Development and validation of a novel patient-reported treatment satisfaction measure for hyperfunctional facial lines: facial line satisfaction questionnaire. *J Cosmet Dermatol*. 2015;14(4):274–285. doi:10.1111/jocd.12166

21. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health*. 2007;10(Suppl 2):S94–s105. doi:10.1111/j.1524-4733.2007.00272.x

22. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health*. 2008;11(2):322–333. doi:10.1111/j.1524-4733.2007.00231.x

23. Encyclopedia of Quality of Life and Well-Being Research: Known-Groups Validity. Available from: https://link.springer.com/referenceworkentry/10.1007%2F978-94-007-0753-5_1581. Accessed September 3, 2019.

24. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Eng J Med*. 2004;350(10):1005–1012. doi:10.1056/NEJMoa021322

25. Cohen J. Statistical Power Analysis for the Behavioral Sciences. *Academic Press*. 1988;2:254.

26. Cazzola M, MacNee W, Martinez FJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J*. 2008;31(2):416–469. doi:10.1183/09031936.00099306

27. Jones PW. St. George's Respiratory Questionnaire: MCID. *COPD*. 2005;2(1):75–79. doi:10.1081/COPD-200050513

28. Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the Six Minute Walk test in chronic lung disease patients.. *Am J Respir Crit Care Med*. 1997;155 (4):1278–1282. doi:10.1164/ajrccm.155.4.9105067

29. de Morton NA, Davidson M, Keating JL. Validity, responsiveness and the minimal clinically important difference for the de Morton Mobility Index (DEMMI) in an older acute medical population. *BMC Geriatr*. 2010;10(1):72. doi:10.1186/1471-2318-10-72

30. Hays RD. D R. Reliability and validity (including responsiveness). In: *Assessing Quality of Life in Clinical Trials: Methods and Practice USA*. Oxford University Press; 2005:35.

31. Roach KE. Measurement of Health Outcomes: Reliability, Validity and Responsiveness. *JPO*. 2006;18(6):P8–P12.

32. Pickard AS, Wilke C, Jung E, Patel S, Stavem K, Lee TA. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. *Respir Med*. 2008;102(4):519–536.

33. Lin FJ, Pickard AS, Krishnan JA, et al. Measuring health-related quality of life in chronic obstructive pulmonary disease: properties of the EQ-5D-5L and PROMIS-43 short form. *BMC Med Res Methodol*. 2014;14:78.

34. Nolan CM, Longworth L, Lord J, et al. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. *Thorax*. 2016;71(6):493–500.

35. Stahl E, Lindberg A, Jansson SA, et al. Health-related quality of life is related to COPD disease severity. *Health Qual Life Outcomes*. 2005;3:56.

36. Rutten-van Molken MP, Oostenbrink JB, Tashkin DP, Burkhart D, Monz BU. Does quality of life of COPD patients as measured by the generic EuroQol five-dimension questionnaire differentiate between COPD severity stages? *Chest*. 2006;130(4):1117–1128.

37. Solem CT, Vera Llonch M, Liu S, et al. Responsiveness of the Eq-5D index and visual analog scale to changes in lung function in patients with cystic fibrosis. *Value in Health*. 2014;17(3):A178.

38. Jones PW, Agusti AG. Outcomes and markers in the assessment of chronic obstructive pulmonary disease. *Eur Respir J*. 2006;27(4):822–832.

39. Vinnard C, Mezochow A, Oakland H, Klingsberg R, Hansen-Flaschen J, Hamilton K. Assessing response to therapy for nontuberculous mycobacterial lung disease: quo vadis? *Front Microbiol*. 2018;9:2813.

40. Payakachat N, Ali MM, Tilford JM, EQ- CT. 5D detect meaningful change? A systematic review. *PharmacoEconomics*. 2015;33 (11):1137–1154.

41. Szende A, Leidy NK, Stahl E, Svensson K. Estimating health utilities in patients with asthma and COPD: evidence on the performance of EQ-5D and SF-6D. *Quality Life Res*. 2009;18(2):267–272.

42. Sullivan PW, Ghushchyan VH, Campbell JD, Globe G, Bender B, Magid DJ. Measurement of utility in asthma: evidence indicating that generic instruments may miss clinically important changes. *Quality Life Res*. 2016;25(12):3017–3026.

43. Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. *Value Health*. 2015;18(1):52–60.

44. Hoogendoorn M, Oppe M, Boland MRS, Goossens LMA, Stolk EA, Rutten-van Molken M. Exploring the impact of adding a respiratory dimension to the EQ-5D-5L. *Med Decision Making*. 2019;39(4):393–404.

45. Gandhi M, Thumboo J, Luo N, Wee HL, Cheung YB. Do chronic disease patients value generic health states differently from individuals with no chronic disease? A case of a multicultural Asian population. *Health Qual Life Outcomes*. 2015;13:8.

Patient Related Outcome Measures   **Dove**press

Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: http://www.dovepress.com/patient-related-outcome-measures-journal