DDT COA QUALIFICATION REVIEW

DDT COA QUALIFICATION NUMBER	00009
SUBMISSION DATE	March 24, 2017
PRIMARY COA REVIEWERS	Nikunj B. Patel, PharmD; Wen-Hung Chen, PhD
PRIMARY CLINICAL REVIEWER	Lauretta Odogwu, MD,
DDT COA QUALIFICATION SCIENTIFIC COORDINATOR	Michelle Campbell, PhD
ASSOCIATE DIRECTOR, COA STAFF	Elektra J. Papadopoulos, MD, MPH
REVIEW DIVISION(S)	Division of Oncology Products 2 (DOP2)
QRT MEMBERS	DOP2: Lauretta Odogwu, MD; Erin Larkins, MD; Patricia Keegan, MD OB: Mallorie Fierro, PhD OCE: Paul Kluetz, MD COA: Nikunj Patel, PharmD; Wen-Hung Chen, PhD; Michelle Campbell, PhD; Elektra Papadopoulos, MD, MPH
REVIEW COMPLETION DATE	February 28, 2017
SUBMITTER	Critical Path Institute Patient-Reported Outcome (PRO) Consortium
CLINICAL OUTCOME ASSESSMENT TYPE	Patient-reported outcome
ENDPOINT(S) CONCEPT(S)	Overall severity of non-small cell lung cancer symptoms
MEASURE(S)	Non-Small Cell Lung Cancer Symptom Assessment Questionnaire (NSCLC-SAQ)
INTENDED POPULATION(S)	Patients (age 18 years or older) with Stage IIIB or IV NSCLC

Wen-hung Chen -S
Digitally signed by Wen-hung Chen -S
DN: c=US, o=U.S. Government, ou=HHS, ou=FDA, ou=People, cn=Wen-hung Chen -S, 0.9.2342.19200300.100.1.1=2001645800
Date: 2019.08.09 17:23:03 -04'00'

Elektra J. Papadopoulos -S
Digitally signed by Elektra J. Papadopoulos -S
DN: c=US, o=U.S. Government, ou=HHS, ou=FDA, ou=People, 0.9.2342.19200300.100.1.1=1300170743, cn=Elektra J. Papadopoulos -S
Date: 2019.08.09 16:49:44 -04'00'

A. EXECUTIVE SUMMARY & RECOMMENDATIONS

This drug development tool (DDT) clinical outcome assessment (COA) review concludes that the Non-Small Cell Lung Cancer Symptom Assessment Questionnaire (NSCLC-SAQ) has adequate evidence of content validity and cross-sectional measurement properties (i.e., internal consistency reliability, test-retest reliability, convergent validity, and known-groups validity) as a measure of overall symptoms of NSCLC in the context of use described below.

The concept of interest of NSCLS-SAS NSCLC is the overall severity of NSCLS symptoms: cough, pain, dyspnea, fatigue, and appetite. The target patient population is adult patients aged 18 years and older with diagnosis of Stage IIIB or IV NSCLC, either treatment naïve or who have received chemotherapy in the last 6 months.

Further evaluation is needed on the instrument's longitudinal measurement properties (i.e., ability to detect change) and the interpretation of clinically meaningful within-patient change in score. It is recommended that this information be obtained in early phase studies in drug development programs.

B. CLINICAL OUTCOME ASSESSMENT REVIEW

Background:

The PRO Consortium's Non-Small Cell Lung Cancer (NSCLC) Working Group at the Critical Path Institute initiated the development and qualification of a new patient-reported outcome (PRO) questionnaire to assess key symptoms of NSCLC as an endpoint measure in clinical trials, referred to as the Non-Small Cell Lung Cancer Symptom Assessment Questionnaire (NSCLC-SAQ. The NSCLC WG (hereafter referred as the Submitter) proposes that the NSCLC-SAQ tool will be used to measure patient self-reported symptom severity in advanced non-small cell lung cancer to be used to evaluate treatment benefit in clinical trials and potentially communicate this treatment benefit in labeling.

The development of the NSCLC-SAQ has included:

- Completion of systematic reviews of the NSCLC literature and existing PRO measures
- The formation of an expert panel of clinical and methodological experts to provide advice during the development process
- Completion of qualitative concept elicitation interviews conducted to identify the NSCLC symptom-related concepts that are most important and relevant to the patients' experience
- A formal item-generation process in which evidence from the concept elicitation interviews, systematic literature reviews, and expert input was used to develop the content of the NSCLC-SAQ

- Qualitative cognitive interviews with participants with NSCLC to evaluate and refine the draft measure, including item reduction
- A translatability assessment conducted concurrently with the early cognitive interview process
- An electronic implementation assessment (by the Electronic Patient-Reported Outcome [ePRO] Consortium's Instrument Migration Subcommittee) to assess the viability for implementation of the PRO measure on all available and appropriate electronic platforms
- Programming for tablet-based data collection and cognitive interviews to assess conceptual equivalence between the paper and electronic formats

1.1 Clinical Trial Design

The Submitter proposes the NSCLC-SAQ use in randomized controlled clinical trials.

Reviewer Comments: The use of NSCLC-SAQ in randomized controlled clinical trials appears reasonable.

1.2 Intended Endpoint Positioning

The Submitter proposes the NSCLC-SAQ is intended to be used as a secondary endpoint measure. The specific endpoint selection, positioning, and measurement approach would be determined by the study sponsor in concert with the appropriate regulatory review agencies.

Reviewer Comments: This reviewer recommends future sponsors to engage the FDA during drug development to discuss appropriate endpoint positioning.

1.3 Labeling or promotional claim(s) based on the COA

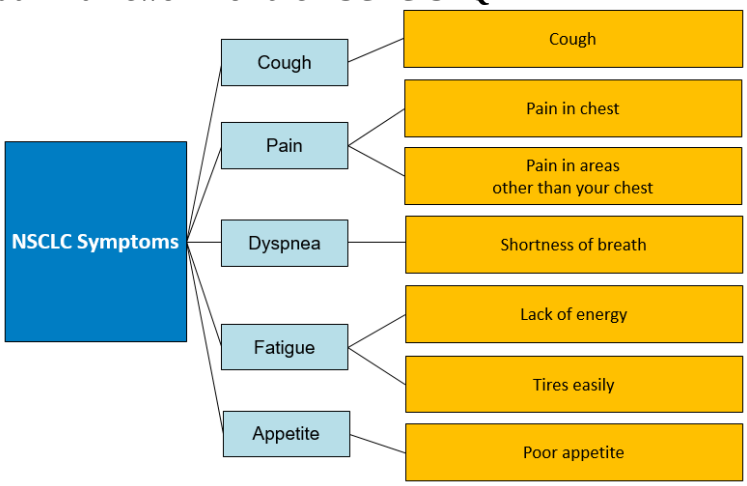
The Submitter proposes that specific labeling will be defined by the clinical trial sponsor in discussion with the FDA.

Reviewer Comments: This reviewer recommends future sponsors to engage the FDA during drug development to discuss the use of NSCLC-SAQ to support labeling claims.

2 CONCEPT OF INTEREST (COI) AND CONCEPTUAL FRAMEWORK

The NSCLC-SAQ assesses patient-reported symptom severity associated with NSCLC. The final seven items of the NSCLC-SAQ address five different symptom concepts that are key to assess for the treatment of NSCLC: cough (one item), pain (two items), dyspnea (one item), fatigue (two items), and appetite (one item).

Figure 1: Conceptual Framework for the *NSCLC-SAQ*



<u>Reviewer Comments:</u> The conceptual framework includes the relevant and important symptoms of advanced NSCLC. The applicant identified these 7 symptoms through interviews with patients as well as through evaluation of the literature and existing lung cancer PRO measures.

3 COA MEASURE OVERVIEW

Non-Small Cell Lung Cancer Symptom Assessment Questionnaire (NSCLC-SAQ): This is a 7-item PRO instrument designed to assess five symptom concepts of NSCLC: cough, pain, dyspnea, fatigue, and appetite. The recall period is one week (worded as "over the last 7 days"). Respondents respond to each of the seven items using a five-point verbal rating scale from either "No <symptom> at All" to "Very severe <symptom>" or from "Never to Always," depending on the item's question structure relative to either intensity or frequency. The NSCLC-SAQ is available in both paper-and-pencil and electronic administration formats. Refer to Appendix A for a copy of the instrument.

User manual: The submitter has developed a provisional user manual.

Timing, data collection method and mode of administration:

- Timing: The NSCLC-SAQ is intended to be administered once weekly at the same time for each assessment.
- Mode of Administration: While the NSCLC-SAQ has been designed specifically for patient self-report, it is possible that in the case of very ill patients, a care provider might need to read the questions verbatim and capture the patient's response on the electronic device or paper-based format.
- Data Collection Method: The NSCLC-SAQ has been designed for electronic data collection on a tablet to be completed directly by the patient. The NSCLC-SAQ was initially developed using a paper-and-pencil format and was later programmed for tablet administration and cognitively evaluated for equivalence between the two formats. The tablet version was used to complete the development and establish the initial measurement properties. Therefore, it is possible, if necessary, to use a paper-and-pencil version of the NSCLC-SAQ, since it was developed, cognitively tested with patients in its early development and shown to be equivalent in terms of patient comprehension to the tablet version, so the measurement properties should be comparable.

Training method/materials: The developer's recommendations on respondent training and instructions for administration by study/clinic staff can be found in the provisional user manual.

4 SCORING ALGORITHM

The provisional scoring algorithm of the NSCLC-SAQ total score is as follows:

- **Cough Domain Score**: score of the cough item, or missing if skipped
- **Fatigue Domain Score**: if both items present, compute mean; or use score from 1 item if the other is missing; or set to missing if both are skipped
- **Pain Domain Score**: if both items present, use most severe of both; or use score from 1 item if the other is missing; or set to missing if both are skipped
- **Dyspnea Domain Score**: score of the shortness of breath item, or missing if skipped
- **Appetite Domain Score**: score of the poor appetite item, or missing if skipped
- **NSCLC-SAQ Total Score**: sum all five domain scores; if any are missing, a total score is not computed. This creates a total score ranging between 0 and 20 with higher scores indicating more severe symptomatology.

Table 1. Scoring the NSCLC-SAQ

Domain	Item		Response
Cough	1. How would you rate your coughing at its worst…?		0, 1, 2, 3, 4
Pain	2. How would you rate the worst pain in your chest…?	*Create a single score by selecting the highest severity (i.e., value) on either item*	0, 1, 2, 3, 4
	3. How would you rate the worst pain in areas other than your chest…?		
Dyspnea	4. How often did you feel short of breath during usual activities…?		0, 1, 2, 3, 4
Fatigue	5. How often did you have low energy…?	*Create a single score by calculating the mean of these 2 items*	0, 1, 2, 3, 4
	6. How often did you tire easily…?		
Appetite	7. How often did you have a poor appetite over the last 7 days?		0, 1, 2, 3, 4
NSCLC-SAQ Total Score (Sum the 5 domains)			**Range 0 to 20**

5 CONTENT VALIDITY

Content validity is defined as evidence that the instrument measures the concept of interest (i.e., disease-specific symptoms) including evidence that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use. The following is an overview of the methods used for instrument development:

- Systematic reviews of the NSCLC literature and existing PRO measures
- Expert panel of clinical and methodological experts to provide advice during the development process
- Qualitative concept elicitation interviews conducted to identify the NSCLC symptom-related concepts that are most important and relevant to the patients' experience
- A formal item-generation process in which evidence from the concept elicitation interviews, systematic literature reviews, and expert input was used to develop the content of the NSCLC-SAQ
- Qualitative cognitive interviews with participants with NSCLC to evaluate and refine the draft measure, including item reduction
- A translatability assessment conducted concurrently with the early cognitive interview process
- An electronic implementation assessment to assess the viability for implementation of the PRO measure on all available and appropriate electronic platforms
- Programming for tablet-based data collection and cognitive interviews to assess conceptual equivalence between the paper and electronic formats

<u>COA Reviewer Comments</u>: The qualitative data submitted for review has demonstrated adequate evidence of content validity with the concurrence of the Review Division during the qualification review team (QRT) meeting on March 21, 2018. At this QRT meeting, the Review Division further

recommended to modify description of concept of Interest as "overall symptom severity of common NSCLC symptoms".

This section focuses on review of the following items:

- Literature & Instrument Review
- Concept Elicitation and Cognitive Interview Studies Patient Population Characteristics
- Concept Elicitation Patient Interview Study
- Item Generation & Development
- Description of Recall Period
- Cognitive Patient Interview Study
- Review of Respondent Burden
- Translatability Assessment

5.1 Literature & Instrument Review

The developer conducted a systematic literature review prior to concept elicitation to evaluate and summarize published research into patient perspectives of NSCLC. The findings from the literature review supported initial development work that informed concept elicitation and item generation/selection process. Separately, the developer also conducted a targeted review of existing NSCLC-focused PRO instruments. This instrument review findings were considered during the item generation process.

5.2 Concept Elicitation and Cognitive Interview Studies Patient Population Characteristics

NSCLC-SAQ content was informed via a review of existing published research studies conducted in NSCLC and findings from open-ended concept elicitation and cognitive patient interview studies in the target patient population. Key eligibility criteria from the concept elicitation and cognitive patient interview studies protocol were as follows:

Inclusion Criteria:
Subjects were eligible for inclusion if each of the following was satisfied at the time of screening:

1. Subject was a male or female at least 18 years of age.
2. Subject had a diagnosis of Stage I-IV NSCLC (at least 85% of patients were required to have histological evidence of either adenocarcinoma or squamous cell carcinoma).
3. Subject had an ECOG Performance Status of 0-2.
4. Subject was diagnosed with Stage I or II cancer and naïve to treatment for NSCLC; and Subject was diagnosed with Stage III or IV cancer and naïve to treatment for NSCLC or had recovered from any prior treatment-related toxicity/adverse events to Common Terminology Criteria for Adverse Events (CTCAE) v4.03 grade 1 (mild) or better.

Concept Elicitation Patient Interview Study Population: A total of 51 participants in the concept elicitation interviews were recruited from clinic sites in six US states (Alabama, Idaho, Illinois, Montana, New York, and North Dakota) while a total of 20 participants in the cognitive debriefing study were recruited from clinical sites in Alabama and Illinois. The Tables 2 and 3 below summarize demographic and clinical characteristics of concept elicitation participants (Source: Qualification Briefing Package, pp22-23).

Table 2: Demographic Characteristics of Concept Elicitation Participants

		Total participants N=51 (100%)
Age (Years):	- Mean (SD)	64.9 (11.2)
	- Median	66
	- Range	46-86
Gender:	- Male	25(49%)
	- Female	26 (51%)
Marital status:	- Married or Living as Married	34 (67%)
	- Widowed	5 (10%)
	- Separated	1 (2%)
	- Divorced	7 (14%)
	- Never Married	4 (8%)
Highest Level of Education Completed:	- Less than High School	3 (6%)
	- High School	25 (49%)
	- Some College	13 (26%)
	- Bachelor's Degree	3 (6%)
	- Graduate or Professional School	7 (14%)
Current Employment Status:	- Employed full-time for wages	8 (16%)
	- Employed part-time for wages	1 (2%)
	- Self-employed	5 (10%)
	- Out of work for more than 1 year	1 (2%)
	- Out of work for less than 1 year	3 (6%)
	- Retired	21 (41%)
	- Unable to work	12 (24%)
Ethnic group:	- Hispanic, Latino or Spanish Origin	2 (4%)
Racial group:	- White	38 (75%)
	- Asian	2 (4%)
	- Black or African American	8 (16%)
	- American Indian/Alaska Native	1 (2%)

Table 3: Clinical and Screening Characteristics for Concept Elicitation Participants

		Total Participants N=51 (100%)
Histological evidence of adenocarcinoma or squamous cell carcinoma	- Unknown	1 (3%)
	- Adenocarcinoma	36 (71%)
	- Squamous cell carcinoma	13 (26%)
	- Adenocarcinoma & Squamous cell carcinoma	1 (2%)
Have documented diagnosis of COPD	- Yes	18 (35%)
	- No	33 (65%)
Stage at initial NSCLC diagnosis	- Stage I	6 (12%)
	- Stage II	1 (2%)
	- Stage III	25 (49%)
	- Stage IV	19 (37%)
Current stage of NSCLC	- Stage I	6 (12%)
	- Stage II	---
	- Stage III	19 (37%)
	- Stage IV	26 (51%)
Current line of NSCLC treatment	- Early	19 (37%)
	- 1st line advanced/metastatic	18 (35%)
	- 2nd line advanced/metastatic	9 (18%)
	- 3rd line advanced/metastatic	3 (6%)
	- Other: Observation, Subsequent	2 (4%)
Current ECOG performance status	- ECOG status 0	17 (33%)
	- ECOG status 1	24 (47%)
	- ECOG status 2	10 (20%)

Cognitive Patient Interview Study Population: The developer conducted cognitive interviews in three waves in a total of 20 adult patients (≥18 years) with NSCLC. Patients were recruited from clinical sites in Alabama and Illinois, United States to participate in this cross-sectional qualitative interview study. The Tables 4 and 5 below summarize demographic and clinical characteristics for the cognitive interview participants (Source: Qualification Briefing Package, pp26-27).

Table 4: Demographic Characteristics for the Cognitive Interview Participants

		Wave 1: Cognitive Interviews N=4 (100%)	Waves 2 and 3: Cognitive Interviews + ePRO Assessment N=16 (100%)
Age (Years):	-Mean (SD)	68.0 (13.3)	64.8 (10.8)
	-Median	68.0	64.5
	-Range	56-80	44-83
Gender:	-Male	1 (25.0%)	11 (68.8%)
	-Female	3 (75.0%)	5 (31.3%)
Marital status:	-Married	1 (25.0%)	11 (68.8%)
	-Widowed	2 (50.0%)	2 (12.5%)
	-Separated	---	1 (6.3%)
	-Divorced	1 (25.0%)	1 (6.3%)
	-Never Married	---	1 (6.3%)
Highest Level of Education Completed:	-Less than High School	1 (25.0%)	---
	-High School	3 (75.0%)	10 (62.5%)
	-Some College	---	6 (37.5%)
Employment outside home:	-Homemaker	---	2 (12.5%)
	-Full-time	1 (25.0%)	3 (18.8%)
	-Part-time	---	---
	-Retired	2 (50.0%)	8 (50.0%)
	-Unable to work	1 (25.0%)	2 (12.5%)
	-Missing	---	1 (6.3%)
Ethnicity	-Hispanic	---	1 (6.3%)
Race	-White	2 (50.0%)	12 (75.0%)
	-Black/African American	2 (50.0%)	3 (18.8%)

Table 5: Clinical Characteristics for the Cognitive Interview Participants

		Wave 1 N=4 (100%)	Waves 2 and 3 N=16 (100%)
Histological evidence of adenocarcinoma or squamous cell carcinoma	-Unknown	---	1 (6.3%)
	-Adenocarcinoma	1 (25.0%)	5 (31.3%)
	-Squamous cell carcinoma	3 (75.0%)	9 (56.3%)
	-Adenocarcinoma & Squamous cell carcinoma	---	1 (6.3%)
Have documented diagnosis of COPD	-Yes	3 (75.0%)	2 (12.5%)
	-No	1 (25.0%)	14 (87.5%)
Stage at initial NSCLC diagnosis	-Stage III	2 (50.0%)	8 (50.0%)
	-Stage IV	2 (50.0%)	8 (50.0%)
Current stage of NSCLC	-Stage I	---	1 (6.3%)
	-Stage III	2 (50.0%)	7 (43.8%)
	-Stage IV	2 (50.0%)	8 (50.0%)
Current line of NSCLC treatment	-Early	1 (25.0%)	3 (18.8%)
	-1st line advanced/metastatic	2 (50.0%)	5 (31.3%)
	-2nd line advanced/metastatic	1 (25.0%)	3 (18.8%)
	-3rd line advanced/metastatic	---	3 (18.8%)
	-Unknown	---	2 (12.5%)
Current ECOG status	-ECOG status 0	2 (50.0%)	5 (31.3%)
	-ECOG status 1	---	11 (68.8%)
	-ECOG status 2	2 (50.0%)	---

Reviewer Comments: The patient population in the qualitative research adequately represents those expected to participate in clinical trials of NSCLC-SAQ in terms of their demographic and clinical characteristics. FDA previously agreed on adequacy of qualitative research patient population.[1] *The majority of the study population consisted patients with advanced forms of NSCLC (Stages III-IV). A smaller number of patients with less severe forms of NSCLC (n=7 Stage I; none from Stage II NSCLC) were included in qualitative research.*

5.3 Concept Elicitation Patient Interview Study

Concept elicitation is a method or process by which concepts (e.g., disease-specific symptoms) that are important to patients are collected through open-ended patient interviews in the target patient population. The developer conducted a cross-sectional concept elicitation patient interview study in a total of 51 adult patients diagnosed with NSCLC across six different geographic locations in the United States (New York, North Dakota, Idaho, Illinois, Alabama, and Montana) to understand the symptoms and effects of NSCLC. Trained interviewers obtained both unprompted and prompted patient input about NSCLC symptoms and their impacts and how the patient feels these factors affect their ability to function. Refer to Section B5.2 for additional background on study population demographic and clinical characteristics.

Saturation of concept, defined as the point at which no new concepts were elicited, was achieved after the review of third transcript groups. Saturation is documented in the form of a saturation table. The developer provided a saturation table (see Table 6 below) demonstrating that saturation was reached for the following concepts: fatigue and tiredness, pain and discomfort, respiratory symptoms, digesting symptoms, and other symptoms. An "X" denotes the group in which a concept was first elicited. Results showed that 93% of all symptom concepts expressed by participants were noted for the first time in the first group of nine interviews. Another 4.7% of concepts newly appeared in the second transcript group, and final 2.3% appeared for the first time in the third transcript group. No new concepts appeared in the final three groups of eight interviews each.

[1] DDT 000009 NSCLC Comments on DDT Submission, December 10, 2017

Table 6: Saturation of Symptom Concepts

Concept Description	Group 1 (N=9 transcripts)	Group 2 (N=9 transcripts)	Group 3 (N=9 transcripts)	Group 4 (N=8 transcripts)	Group 5 (N=8 transcripts)	Group 6 (N=8 transcripts)
Fatigue and Tiredness						
Exhaustion	X					
Fatigue	X					
Low Energy	X					
Low Stamina	X					
Tiredness	X					
Weakness	X					
Pain and Discomfort						
Back Pain	X					
Bone Pain	X					
Chest Pain	X					
General Pain	X					
Muscle Pain	X					
Respiratory Symptoms						
Bronchitis	X					
Coughing Up Blood	X					
Coughing	X					
Difficulty Breathing	X					
Emphysema	X					
Phlegm	X					
Pneumonia	X					
Shortness of Breath	X					
Wheezing	X					
Digestive Symptoms						
Appetite	X					
Diarrhea	X					
Difficulty Swallowing	X					
Nausea	X					
Vomiting	X					
Other Symptoms						
Cognition	X					
Dizziness and Fainting	X					
Headache	X					
Heart Problems	X					
Heat Sensitivity	X					
Hoarseness	X					
Immunity Lessened	X					
Lump		X				
Numbness	X					
Restlessness	X					
Feeling Sick	X					
Skin Change	X					
Sore Throat			X			
Swelling	X					
Taste Change	X					
Twitching	X					
Voice Change		X				
Weight Loss	X					
Number of concepts coded in each group	**40**	**2**	**1**	**0**	**0**	**0**
Percent of relevant concepts coded (N=43)	**93.0%**	**4.7%**	**2.3%**	**0.0%**	**0.0%**	**0.0%**

(Source: Concept Elicitation Summary Report, page 30)

The most predominant symptom concept was *coughing* followed by *shortness of breath* and *difficulty breathing*. These were followed by the symptom concepts of *tiredness* and *fatigue.* When enumerated from specific exercises given during the interview, coughing was shown to be the symptom concept most often mentioned spontaneously, followed by *shortness of breath, chest pain, general pain, tiredness, difficulty breathing, and less appetite.* Additionally, the developer also conducted bothersome rating exercise to show which symptoms of NSCLC are the most bothersome for patients using an 11-point numeric rating scale (NRS) where 0 was anchored with "not bothersome at all" while 10 with "extremely bothersome." Physical symptoms and fatigue /tiredness symptoms were rated the highest for degree of bother (7.8 to

10.0). Respiratory symptoms, appetite, pain and breathing symptoms were rated lower in bothersome (5.6 to 6.9 out of 10). Similar analyses were conducted for the disease-related impacts; however, these were out of scope for this qualification review.

Reviewer Comments: The concept elicitation patient interview study was conducted in adequate patient population in terms of its size and diversity (demographic and clinical). Study protocol, patient interview guide, and analysis plan appeared appropriate. The saturation table above appears to document saturation of most relevant and important concepts and the collection of additional data will not likely add to the understanding of how patients perceive the concept of interest and the items in the PRO instrument. Most predominant symptoms identified through this study were incorporated in the proposed PRO instrument for qualification.

5.4 Item Generation & Development

PRO item generation and development should include input from the target patient population to establish the items that reflect the concept of interest and contribute to its evaluation. An initial 9-item version of the NSCLC-SAQ item content was generated based on:

- Concept elicitation patient interviews
- Literature review
- Review of existing NSCLC-focused PRO instruments
- Input form Clinical and PRO measurement experts

The developer examined 43 symptom concepts from the concept elicitation interview as outlined in the Table 7 during a two-day meeting involving clinical experts, measurement scientists, and members of the C-Path PRO Consortium's NSCLC Working Group. Following key features were considered to guide item selection: (1) the predominance of each symptom concept within the concept elicitation interview transcript database; (2) the number of patients expressing each concept; (3) the percentage of patients who expressed each concept spontaneously, rather than in response to probing by the interviewer; (4) the severity rating given by patients for the symptoms they experienced; and (5) the degree to which patients rated each symptom as bothersome. The concept selection process resulted in the removal of 32 concepts and the retention of 11 concepts for inclusion in PRO measurement (see Table 8).

Table 7 below describes 43 symptom concepts identified from concept elicitation interviews (Source: Item Generation Summary Report, Appendix A).

Concept Summary Table – NSCLC Symptoms

KEY: Gray Shade: No data
Red Text: High end of data group (value above the median observed for all concepts)
Green Bold: Most highly-supported concepts from patient interviews (i.e. concepts with values above the median on 2 or more of the following criteria: number of expressions, number of unique patients expressing concept, percent of patients expressing concept spontaneously, mean severity ratings, and mean bothersome rating).

Sub-Domain	Concept	Predominance in coding		Patient assessment during interview		
		Number of expressions of symptom concept (High if above mean [>29])	Number of transcripts contributing to symptom expression (High if >10 [approx. 30% of sample])	Percent of patients who spontaneously offered symptom (Probed percent if larger) (High if >20%)	Mean Severity Ranking (High if above mean [>6.7])	Mean Bothersome Ranking (High if above mean [>5.7])
Fatigue and Tiredness	Exhaustion	11	6			
	Fatigue	**37**	**13**	19%	**6.9**	**6.3**
	Low Energy	**30**	**10**	10%	**7.0**	**7.8**
	Low Stamina	9	6	7%	**8.0**	**8.0**
	Tiredness	**121**	**26**	**39%**	**6.9**	**6.3**
	Weakness	20	9	13%	4.0	5.0
Pain and Discomfort	**Back Pain**	**30**	5	7%	**8.5**	**7.8**
	Bone Pain	7	4	3%	6.0	**6.0**
	Chest Pain	**42**	9	**26%**	**7.8**	**6.3**
	General Pain	**36**	9	**23%**	**7.9**	**7.1**
	Muscle Pain	24	8	10%	5.5	4.8
Respiratory Symptoms	Bronchitis	14	6	10%	**9.5**	2.5
	Coughing Up Blood	14	5	16%	5.0	3.3
	Coughing	**149**	**27**	**55%**	6.4	5.6
	Difficulty Breathing	**71**	**15**	19%	6.1	**5.9**
	Emphysema	8	6			
	Phlegm	**30**	9	3% (3%)	3.3	3.5
	Pneumonia	**35**	9	16%	**8.8**	2.8
	Shortness of Breath	**75**	**17**	**42%**	6.1	**6.5**
	Wheezing	22	7	13%	6.2	5.2
Digestive Symptoms	**Appetite**	**71**	**17**	**32%**	**7.1**	**6.5**
	Diarrhea	11	5			**10**
	Difficulty Swallowing	14	6	3%	5.0	4.0
	Nausea	**39**	**12**	10%	6.0	**7.3**
	Vomiting	19	6	3%	**7.5**	**6.0**
Other Symptoms	**Cognition**	**30**	8	3%	**8.0**	5.0
	Dizziness and Fainting	24	6	10%	**8.5**	**8.7**
	Headache	8	2	3%	**9.0**	0
	Heart Problems	12	3	3%	5.0	**10**
	Heat Sensitivity	15	5	3%	**7.0**	5.5
	Hoarseness	25	**13**	7% (29%)	5.4	5.3
	Immunity Lessened	12	6	7%	6.0	0.5
	Lump	8	3	3%	4.0	3.0
	Numbness	21	6	7%	**7.0**	**8.0**
	Restlessness	5	2			
	Feeling Sick	8	5			
	Skin Change	4	1	3%	**9.0**	5.0
	Sore Throat	6	2	7%	2.0	1.5
	Swelling	13	8	7%	**9.5**	2.5
	Taste Change	11	4			**7.0**
	Twitching	14	3			
	Voice Change	8	2	3% (16%)	6.3	4.3
	Weight Loss	**36**	**14**	19%	6.0	5.5

Table 8: Conceptual Justification Table for NSCLC Symptoms

Sub-Domain	Concept from patient interviews	Supported by literature	Concept is assessed in other PRO's	Concept is relevant to patients in CE interviews*	Should concept be included for assessing clinically meaningful change in NSCLC? (KEEP or DROP)
Fatigue	Exhaustion		YES		DROP
	Fatigue	YES	YES	YES	**KEEP**
	Low Energy		YES	YES	**KEEP**
	Low Stamina			YES	**KEEP**
	Tiredness	YES	YES	YES	**KEEP**
	Weakness	YES	YES		DROP
Pain and Discomfort	Back Pain			YES	DROP
	Bone Pain				DROP
	Chest Pain		YES	YES	**KEEP**
	General Pain	YES	YES	YES	**KEEP**
	Muscle Pain		YES		DROP
Respiratory Symptoms	Bronchitis	YES			DROP
	Coughing Up Blood	YES	YES		**KEEP**
	Coughing	YES	YES	YES	**KEEP**
	Difficulty Breathing		YES	YES	**KEEP**
	Emphysema				DROP
	Phlegm				DROP
	Pneumonia			YES	DROP
	Shortness of Breath	YES	YES	YES	**KEEP**
	Wheezing				DROP
Digestive Symptoms	Appetite	YES	YES	YES	**KEEP**
	Diarrhea	YES	YES		DROP
	Difficulty Swallowing	YES	YES		DROP
	Nausea	YES	YES	YES	DROP
	Vomiting	YES	YES	YES	DROP
Other Signs and Symptoms	Cognition	YES	YES		DROP
	Dizziness and Fainting		YES		DROP
	Headache				DROP
	Heart Problems				DROP
	Heat Sensitivity				DROP
	Hoarseness	YES			DROP
	Immunity Lessened				DROP
	Lump				DROP
	Numbness	YES	YES	YES	DROP
	Restlessness				DROP
	Feeling Sick		YES		DROP
	Skin Change	YES			DROP
	Sore Throat				DROP
	Swelling				DROP
	Taste Change				DROP
	Twitching				DROP
	Voice Change				DROP
	Weight Loss	YES	YES	YES	DROP

** **Note**: Concepts identified as "relevant to patients" were found to have values above the median for all concepts on 2 or more of the following criteria: number of overall expressions, number of unique patients expressing concept, percent of patients expressing concept spontaneously, severity ratings, and bothersome rating.*

The 11 concepts identified for PRO inclusion were then cross referenced against the content coverage of the existing NSCLC-focused PRO instrument and were reviewed against a set of selection criteria such as item relevance, importance, and sensitivity to change because of a successful treatment. Using the established selection criteria, the developer drafted an initial 9-item instrument with a recall period of 7 day. The 9 NSCLC-related symptom items were cough, coughing up blood (hemoptysis), general pain, chest pain, shortness of breath, difficulty breathing, lack of energy, tires easily, and appetite. The developer considered evaluating both verbal rating scale (VRS) and 11-point numeric rating scale (NRS) as response options for each question during the cognitive interview before confirming most appropriate response format. Refer to Sections C4.5-4.6 for discussion on recall period selection and additional modifications to the initial version of the PRO instrument. The most updated version of the PRO instrument submitted for qualification is a 7-item instrument with verbal rating scale as response options and a recall period of 7 days.

Refer to the item tracking matrix (not reproduced here) in the qualification review package for specific information on individual item generation and development.

Reviewer Comments: The methodologies (concept elicitation, literature/instrument review, expert panel, item tracking matrix) appear reasonable to document evidence of item generation and selection. The developer included input from the target patient population (via concept elicitation) along with literature/instrument review and experts input to generate initial item pool for further testing during cognitive interview study. This reviewer evaluated the item tracking matrix to ensure the item content selection and modifications are appropriate and the reasons for those changes are provided. Refer to Section B5.5 for comments on recall period. Refer to Section B5.6 reviewer comments for concerns related to some item content.

5.5 Description of Recall Period

To select a recall period for the PRO instrument, the developer considered the feasibility of various periods in clinical trial settings, the level of patient burden required by different approaches, and the patient responses during the qualitative interviews indicating the pattern of change in their NSCLC symptoms over time. After considering these factors, the developer considered a recall period of "past 7 days" appropriate.

Reviewer Comments: A recall period of "past 7 days" for the PRO instrument appears reasonable. Patients understood the recall period appropriately based on review of the cognitive interview study results.

5.6 Cognitive Patient Interview Study

Cognitive patient interviews are generally conducted to gain a better understanding of how patients interpret PRO item content (instruction, questions or items, response options, recall period). Cognitive interviews were conducted in three waves to understand how patients with NSCLC comprehend and response to the PRO instrument in addition to inform iterative

modifications of item content to ensure patient understanding. Refer to Section B5.1 for description of patient characteristics of the study population. This reviewer describes key findings below.

Key Findings:

Cognitive Patient Interviews Wave 1 (n=4): During the first wave of interviews, two versions of the 9-item draft instrument were evaluated in 4 patients with NSCLC. There was difficulty expressed by patients with 8 out of the 9 items due to issues around the numeric response scale (NRS). For example, in responding to item 3 ("please rate your cough over the last 7 days"), one patient selected 2 on the NRS scale, but stated that he does not experience coughing. Subsequently, the developer decided to drop the NRS version of the item and decided to procced with the verbal rating for subsequent interviews. Additionally, the developer made the following modifications to the verbal response scale (VRS) version:

- An item asking patients to rate the amount of blood coughed up (hemoptysis) was dropped because of low frequency of experience and expert panel recommendation. This finding confirmed previous supposition by the developer that hemoptysis was infrequently experienced by patients and the item would not be sensitive.
- Both the question stem and the response options of the first 3 items (general pain, chest pain, and cough) were revised to reduce ambiguity around what aspect of the concept was being assessed.
- The position of the phrase "over the past 7 days" was moved to the end of each item stem, to maintain consistency.
- The word "past" was also changed to "last" for the instructions and all items to maintain consistency.
- The coughing item was moved into the first position in sequence so a chest symptom item would start the survey rather than a general pain item.
- The general pain item was replaced by an item assessing pain "in areas other than your chest" to serve as a mutually-exclusive complement to the item assessing pain in the chest.

Cognitive Patient Interviews Wave 2 (n=10): In the second wave, the revised 8-item instrument was evaluated in cognitive interviews with 10 subjects. Participant responses during the Wave 2 interviews supported the overall relevance of the included concepts, provided evidence of conceptual equivalence between the paper and ePRO formats of the instrument, and facilitated refinement of the wording for several items. Specifically, the developer made the following key changes:

- No substantial difficulties were expressed by subjects related to the instructions, recall period, item order, or in the wording presented for the item stems.
- The 3 severity/intensity-focused items (cough, chest pain, non-chest pain) were reworded to assess the peak ("worst") intensity of the symptom.
- The two dyspnea-focused items were combined to result in a single item that assesses "feeling short of breath during usual activities." Because most subjects interpreted "short

of breath" and "difficulty breathing" to be similar, if not interchangeable phrases; and all subjects reporting dyspnea indicated that their breathing symptoms occurred only upon some type of exertion.
- The appetite-focused item was reworded from assessing "good appetite" to "poor appetite" to allow the response options to remain directionally consistent with the other items in the instrument.
- No specific difficulty was reported with the use of the ePRO device, nor did subjects describe any differences in their interpretation of any of the items between the paper and ePRO formats.

Cognitive Patient Interviews Wave 3 (n=6): The revised, 7-item instrument was further evaluated through six additional cognitive interviews (Wave 3). During these interviews, the revised items were confirmed, and additional insight into respondent comprehension of the instrument was gathered. Patients in Wave 3 confirmed the relevance of items, expressed no difficulty with comprehension of the items or response options, and noted no noteworthy differences in meaning or response between the paper and ePRO format of the instrument.

- Subject responses confirmed the clarity of interpretation associated with the "worst" phrasing for the three severity-focused items (cough, chest pain, non-chest pain), and could clearly articulate the meaning of the items with examples from their own experiences with NSCLC.
- For the two pain items, subjects could clearly describe distinct examples of locations where they experience the pain they considered when responding to each item, suggesting the items are understood appropriately by patients.
- Interviewed subjects expressed no difficulty with the revised dyspnea item and confirmed the relevance of assessing shortness of breath occurring with usual activity.
- Subject responses indicated clear conceptual differences between the experience of having low energy and that of tiring easily; and confirmed the relevance of assessing both concepts as part of their overall experience with NSCLC.
- Subject responses confirmed that "poor appetite" was appropriate to their experience with NSCLC and reported no difficulty with understanding or responding to the appetite item.

Refer to the cognitive interview study report for in-depth discussion of study findings.

Reviewer Comments: The developer conducted cognitive interviews in the target patient population to demonstrate evidence of patient understanding of the item content and inform additional iterative modifications to improve item relevance and clarity. The developer also qualitatively evaluated experience using paper vs. electronic version of the PRO questionnaire. This reviewer provides key comments based on review of the cognitive interview study report, corresponding patient transcripts, and relevant parts of quantitative study report.

- *Cognitive interviews were adequately conducted to document evidence of iterative item content modification and patient understanding. Study protocol, patient interview*

guides, and analysis plan were reasonable. Patients understood PRO instruction as intended by the developer.

- *Previously FDA expressed concerns related to patients' ability to respond differentially to chest pain and general pain questions in an earlier version of the instrument.[1] However, patients appeared to comprehend and respond differentially and appropriately to the two pain questions included in the PRO instrument version submitted for qualification. Interestingly, these two items exhibited higher ceiling effects (patients indicating "no pain at all" were 51% and 37% of the sample, respectively, n=152). Also, these two items appeared to be moderately correlated per quantitative study report (r=0.455; see Quantitative Study Report, Table 6, Page 35), while not conclusively but possibly suggesting no measurement redundancy. While patients did not experience difficulties answering the pain questions, this reviewer recommends further evaluation of these two items in clinical trials to ensure they can provide clinically meaningful information on pain experience in the target patient population.*
- *While patients did not have trouble in responding to low energy and tire easily questions, this reviewer cannot conclusively determine whether the low energy and tiredness items are conceptually distinct as asserted by the developer. There were some patients who described both items asking about the same thing (Exemplary quote: "Being tried or low energy are to me the same thing") while some patients considered them two separate things (Exemplary quote: "Tire easily shows more than low energy. Its two different things."). Further, these two items were highly positively correlated per the quantitative study report, indicating possible measurement redundancy (r=0.844; see Quantitative Study Report, Table 6, Page 35). While the developer may retain these two items, this reviewer recommends further testing in clinical trials (particularly multi-national) to elucidate existence of conceptual difference, if any.*
- *While one to two items such as low energy and tire easily may capture some elements of fatigue, this reviewer notes that fatigue is a multi-dimensional concept consisting of both physical and mental aspects of patients' perception of a disease and/or treatment; therefore, may not be comprehensively assessed using one to two questions.*
- *Changes to 3 severity/intensity-focused item (cough, chest pain, and non-chest pain) to assess the peak ("worst") intensity of symptom appears consistent with past FDA advice.*
- *Selection of verbal response scale (VRS) vs. numeric response scale (NRS) as response options for the PRO measure appears appropriate as patients found the VRS much clearer compared to NRS.*
- *Removal of the hemoptysis item appears reasonable because of its infrequent occurrence in the target patient population. Additionally, hemoptysis is not a cancer-specific sign and can often occur in patients with pulmonary diseases such COPD and bronchiectasis/infection.*
- *Combining of two separate dyspnea-focused items to a single item appears reasonable as patients often interpreted them interchangeably. FDA previously flagged these two items as possibly assessing an identical concept.[1]*
- *Rewording of the appetite item appears reasonable to ensure consistent directionality with other items in the instrument. This modification is consistent with past FDA advice.[1]*

Interestingly, treated patients exhibited ceiling effect compared to treatment naïve per quantitative study report (36% of treated patients responded "never" to poor appetite question compared to only 10% of the treatment naïve, n=102 treated patients, n=50 treatment naïve; see Quantitative Study Report, Table 4, Page 32). This observation may possibly suggest, while not categorically, a greater contribution of symptoms from disease compared to treatment.

- *One of the removed concepts, weight loss, was determined to be important for overall measurement of patient status, but was removed from consideration for PRO measurement, as patient's weight will be assessed directly in clinical trials.*
- *Patients appeared to comprehend and use the ePRO version of the questionnaire appropriately.*

Refer to Section B6 for in-depth discussion on other measurement properties and item performance.

5.7 Review of Respondent Burden

Undue physical, emotional, or cognitive strain on patients generally decreases the quality and completeness of PRO data. The NSCLC-SAQ is a 7-item questionnaire developed to present a low burden to respondents. For each of the seven items, the respondent is asked to "please choose the one response that best describes your experience over the last 7 days." Each item concludes with "…over the last 7 days" to remind the respondent to answer the question thinking about the seven days prior to providing the response. The developer did not provide information regarding recommended frequency of PRO administration in a clinical trial.

Reviewer Comments: The NSCLC-SAQ appears to pose minimal respondent burden.

5.8 Translatability Assessment

Parallel with the overall cognitive interview process, experienced PRO linguistics consultants conducted a translatability assessment in five languages (Chinese, Hindi, Japanese, Russian, and Spanish) to evaluate the potential for difficulty in maintaining conceptual equivalence when translating each item.

Experienced PRO measure translation consultants were identified for each of the five languages and asked to review the English text for each item and response option and to rate the level of difficulty they would have in finding a suitable translation for the item that would maintain content equivalency. Difficulties were rated on the following five-point scale: 1=not difficult at all, 2=slightly difficult, 3=moderately difficult, 4=very difficult and 5=extremely difficult. For each item, the consultant also provided a report on any difficulties, suggestions, and explanations for ways to maintain conceptual equivalency if translations were possible.

Findings from the translatability assessment facilitated revisions to items prior to the completion of the cognitive interview phase, and the overall results of the translatability assessment showed

that most of the items in the NSCLC-SAQ can be translated easily in a way that maintains conceptual equivalence.

6 RELIABILITY, CONSTRUCT VALIDITY, ABILITY TO DETECT CHANGE

A quantitative pilot study was conducted to assess item performance, to provide an assessment of the NSCLC-SAQ measurement properties and to develop a provisional scoring approach prior to submission to the FDA for qualification review.

6.1 Design of the Pilot Study

A targeted group of 150 subjects with clinically-diagnosed NSCLC from US-based clinical sites participated in electronic data collection by completing a questionnaire battery using a touchscreen-enabled tablet computer.

Eligible subjects were scheduled for a study visit at their clinic site to complete the questionnaire battery that included demographic information, the NSCLC-SAQ, NCCN/FACT Lung Symptom Index-17 (FLSI-17), and a patient global assessment of severity (PGIS). These subjects were also expected to participate 7 to 10 days later in an additional clinic visit to complete the PGIS, a patient global impression of change (PGIC), and the NSCLC-SAQ to assess test-retest reliability.

Key Eligibility Criteria:

1. Patient was a male or female at least 18 years of age.
2. Patient had a diagnosis of Stage IIIB or IV NSCLC (with at least 85% of subjects having histological evidence of either adenocarcinoma or squamous cell carcinoma).
3. Patient was naïve to treatment (defined as the patient being naive to their current chemotherapy at their current stage of NSCLC and not having received chemotherapy for the past 6 months (from study enrollment) - or - if not treatment naive (received chemotherapy within the past 6 months),has recovered from any prior treatment related toxicities/adverse events to Common Terminology Criteria for Adverse Events (CTCAE) v4.03 grade 1 (mild) or better.

6.2 Pilot Study Results

Demographic and Clinical Characteristics of Pilot Study Participants:

A total of 152 subjects were enrolled and all were included in the NSCLC-SAQ item-level analyses and convergent validity analyses while 90 subjects were included in the retest analyses. Subjects were on average 64 years of age, 57% were female, and 87% were white. More than half (61%) were married or living as married, 16% were divorced, 14% were widowed. The

majority (84%) had a minimum high school education, 49% were retired, 18% were unable to work, and of those that did work, 54% had a household yearly income of $35,000 or higher.

Of the 152 subjects, 126 (83%) patients had Stage IV NSCLC and 26 (17%) had stage III disease; 73% had histological diagnosis of adenocarcinoma, and 23% had histological evidence of squamous cell carcinoma; Time since diagnosis of NSCLC averaged 1.1 years (range 0.1-9.6 years); About one-third (33%) were treatment naïve, 33% had received 1 line of therapy and 17% had received 2 lines, and 17% had received three lines of therapy; 32% were ECOG performance status 0, 51% had an ECOG performance status of 1, and 17% ECOG 2; 65 patients (43%) had a comorbid clinical diagnosis of chronic obstructive pulmonary disease (COPD). At time of enrollment, approximately 59% were on systemic chemotherapy, 7% were receiving chemotherapy with radiation, and 2% were receiving radiation.

Reviewer Comments: The majority (83%) had stage IV NSCLC at the time of diagnosis and a subset of patients (43%) had COPD; more patients had an ECOG of 1. Although Asians in the study were underrepresented, the demographic characteristics in this pilot study is what is generally represented in randomized advanced NSCLC clinical trials. Further psychometric testing in larger and more diverse samples would be informative and is encouraged as the measure is adopted in clinical trials.

Item Descriptive Statistics:

Mean scores for each of the seven items of the NSCLC-SAQ ranged from 0.84 to 2.14 using a response scale between 0 ("*Never*" or "No Pain at All" or "No Coughing at All") to 4 ("*Always*" or "Very Severe Pain" or "Very Severe Coughing"). All items used the full range (0, 1, 2, 3, and 4) of responses. Two items (#2 "How would you rate the worst pain in your chest over the last 7 days?" and #3 "How would you rate the worst pain in areas other than your chest over the last 7 days?") had a ceiling effect of 51% and 37%, respectively. A total of 43 (28%) subjects indicated "No Pain at All" for both pain items. Skipping items was allowed, however, there were no missing item responses in this pilot study.

One pair of items that assessed similar concept, had a large item-to-item correlation (r=0.84) indicating redundancy: #5 "How often did you have low energy?" and #6 "How often did you tire easily?" Another pair of items that assess similar concept (#2 "How would you rate the worst pain in your chest?" and #3 "How would you rate the worst pain in areas other than your chest?") had a correlation of 0.46 indicating no or weak redundancy.

Descriptive statistics were also examined for the *NSCLC-SAQ* items split by the treatment naïve subjects and those that had prior treatment. Scores were slightly higher (worse) for the treatment-naïve subjects but the difference was not statistically significant.

Item Performance using Rasch Model Analysis:

Rasch analyses showed that the item response categories of all seven items were appropriately ordered. That is, each item's response categories reflect an ordered continuum from "No symptom at all" to "Very severe symptom" (items 1-3) or "Never" to "Always" (items 4-7).

To account for conceptual overlap and to avoid over-weighting in the NSCLC-SAQ scoring, the scores for the following items were combined: "low energy" and "tire easily" (represented by the mean of both items), and "pain in chest" and "pain in other areas" (represented by the most severe answer of either item). This scoring approach results in five symptom scores representing each of the five concepts identified in the conceptual framework: cough, pain, dyspnea, fatigue, and appetite.

Reviewer Comments: While the developer may retain these two pairs of items and address the concern of overlapping items using the proposed scoring approach, this reviewer recommends further testing in clinical trials (particularly multi-national) to elucidate existence of conceptual difference, or in the case of finding no conceptual difference remove one item from each of the two pair to further reduce the number of items.

Differential Item Functioning (DIF)

DIF in relation to gender, age, NSCLC stage, treatment status, ECOG performance status and COPD diagnosis was examined within this sample in an effort to test if items were being endorsed differently by men and women, different age groups, subjects in Stage IIIB and IV, being treatment naïve or not, ECOG status groups (0, 1, 2), or by subjects with COPD or not. No DIF was observed for gender, age, stage, or treatment status (i.e., treatment naïve vs. treated). DIF was observed for all *NSCLC-SAQ* variables with respect to subjects with a diagnosis of COPD. DIF was also observed in three *NSCLC-SAQ* variables with regard to ECOG performance status.

Internal Consistency and Test-retest Reliability:

Internal consistency reliability was examined and resulted in a Cronbach's alpha of 0.78, indicating a reliable scale. Test-retest reliability was examined using the intraclass correlation coefficient (ICC) and Pearson's product-moment correlation. Data from stable subjects, defined by providing the same response to the PGIS on Day 1 and Day 8, were used. Of the 148 subjects that completed the Day 8 (retest) data collection, 90 (61%) indicated the same PGIS responses between Day 1 and Day 8. The ICC was 0.87 with a 95% confidence interval of 0.80 to 0.91 and the Pearson's *r* was also 0.87 ($p<0.001$).

Construct Validity:

Convergent validity was assessed by examining the magnitude of correlations between the NCSLC-SAQ items and the FLSI-17 items. All associations hypothesized to have stronger correlations (>0.50) between items of the NSCLC-SAQ and items of the FLSI-17 were met. The NSCLC-SAQ total score was correlated with the FLSI-17 Disease-Related Symptoms-Physical score at 0.87 ($p<0.001$).

Known-groups validity was examined using the PGIS, self-reported health status, and ECOG performance status. The NSCLC-SAQ total score was able to differentiate between levels of: severity (not severe, mildly severe, moderately severe, very/extremely severe [$p<0.001$]), health status (excellent, very good, good, fair, poor [$p<0.001$]), and performance status (ECOG 0, ECOG 1, ECOG 2 [$p<0.001$]).

Ability to Detect Change:

The pilot study was primarily a cross-sectional study with only a sample of 90 subjects participated in the retest. A longitudinal study with sufficient number of patients experience change is needed to evaluate the NSCLC-SAQ Total score's ability to detect change. It is recommended that this information be obtained in early phase studies in drug development programs.

7 INTERPRETATION OF SCORES

Information to support threshold(s) for clinically meaningful within-patient change(s) in the NSCLC-SAQ total score is needed. Further evaluation is needed on the instrument's longitudinal measurement properties and the interpretation of clinically meaningful within-patient change in score. CDER recommends that data to interpret clinically meaningful within-patient change in the NSCLC-SAQ total score be gathered and evaluated in early phase development prior to its use in confirmatory studies.

8 LANGUAGE TRANSLATION AND CULTURAL ADAPTATION

The NSCLC-SAQ was developed in the United States. At the time of qualification review, the developer did not provide copies of translated questionnaires. However, translatability assessment was conducted during the NSCLC-SAQ development (refer to Section C4.8 for additional background).

Reviewer Comments: A complete list of available translations as well as information on methodology of translation should be available from the developer and described in the instrument user manual. FDA recommends careful adherence to good practices for translation and cultural adaptation as described in ISPOR task force report (Wild et al. 2005; Wild et al. 2009), including item definition, dual forward translation; reconciliation; dual back translation review; harmonization; in-person cognitive testing with NSCLC patients in each target country using a standardized interview transcript; analysis of cognitive testing results; clinician review

as-needed to verify terminology; finalization; and dual proofreading. This methodology is to ensure that the translated versions of the PRO instrument are both conceptually equivalent to the source version (in this case U.S. English) and easily understood by the target population.

9 REFORMATTING FOR NEW METHOD OR MODE OF ADMINISTRATION

The NSCLC-SAQ is available in paper and electronic formats. Following the programming of the NSCLC-SAQ onto an electronic tablet format, the 16 cognitive interviews comprising waves 2 and 3 were used to also evaluate the migration of the draft NSCLC-SAQ from paper to electronic format. In these cognitive interviews, the draft NSCLC-SAQ measure was completed by participants on both paper and tablet formats and evaluated through interviewer probing. Feedback from participants' responses during the interviews showed no indication that the understanding of the instructions, items, or response options was affected by the mode of data collection. Therefore, the electronic tablet format of the NSCLC-SAQ was shown to be cognitively equivalent to the paper format originally developed and evaluated.

Reviewer Comments: This reviewer reviewed the cognitive interview study report and corresponding patient transcripts for discussion on administration modes. Patients appeared to comprehend and use the ePRO version of the questionnaire appropriately.

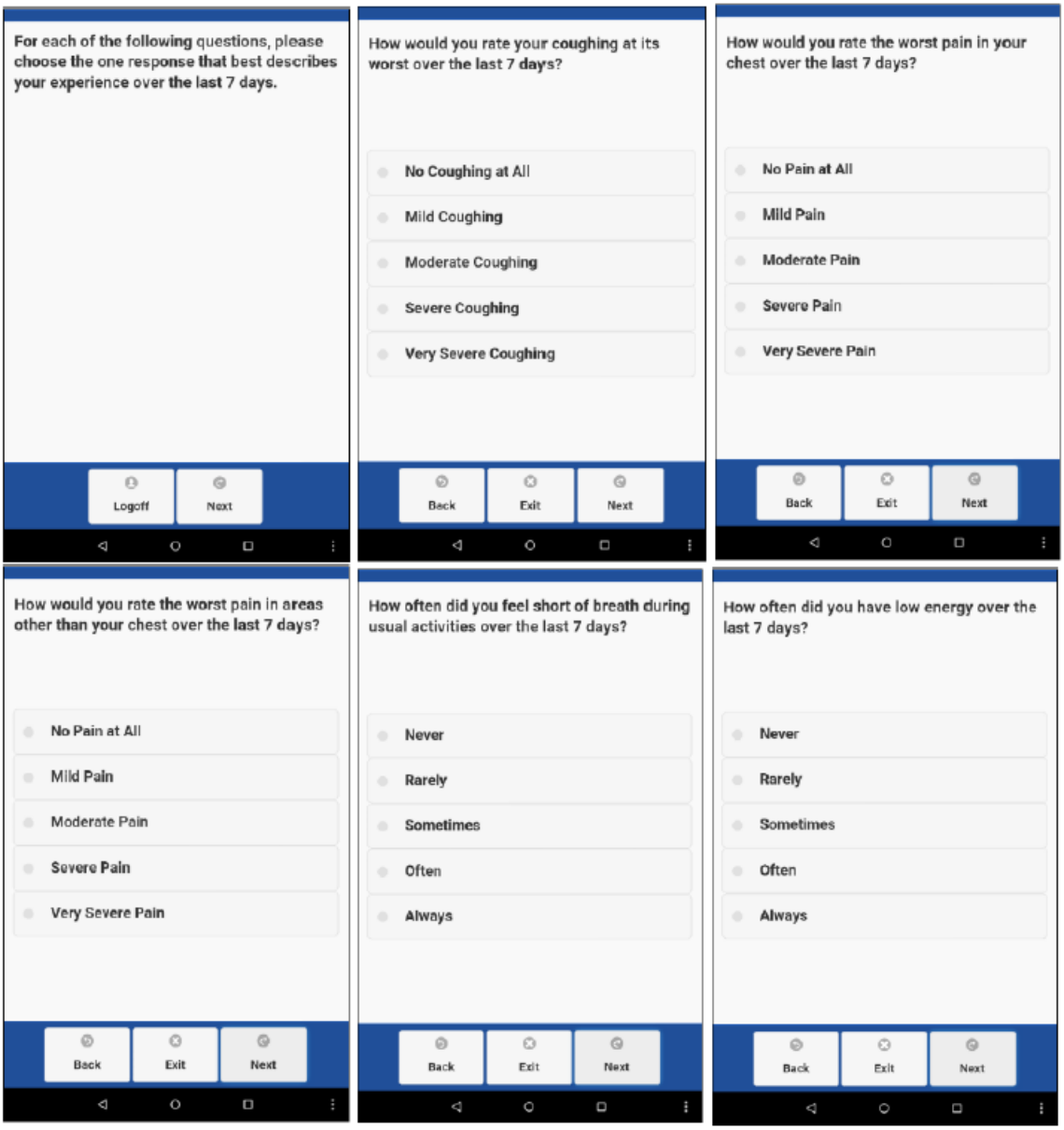
10 REVIEW USER MANUAL

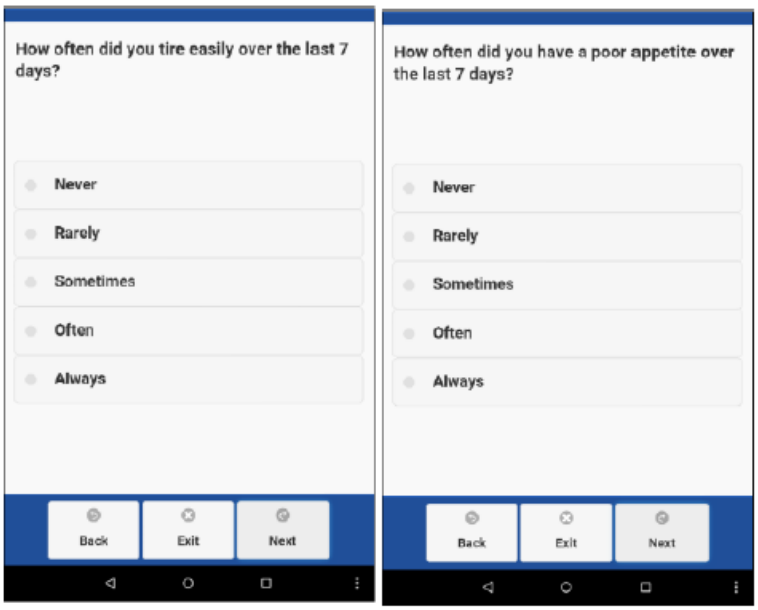
A PRO user manual summarizes the PRO instrument development process and provide instruction on how to incorporate the instrument into a clinical trial in a way that minimizes administrator burden, patient burden, missing data, and poo data quality. A provisional user manual for the NSCLC-SAQ has been developed which outlines information relating to the qualitative and quantitative development and testing of the NSCLC-SAQ Information on the administration procedures, methods and modes are outlined as well as study participant and investigator training processes. Scoring and interpretation procedures are also included to provide guidance to users of the NSCLC-SAQ and to ensure consistent implementation in clinical studies.

Reviewer Comments: The provisional user manual appears consistent with the PRO concepts of interest and context of use as intended by the developer.

C. APPENDICES

Appendix A: NSCLC-SAQ v0.1 Screen Shots




How often did you tire easily over the last 7 days?
Never
Rarely
Sometimes
Often
Always
Back
Exit
Next
How often did you have a poor appetite over the last 7 days?
Never
Rarely
Sometimes
Often
Always
Back
Exit
Next