CENTER FOR DRUG EVALUATION AND RESEARCH

DRUG DEVELOPMENT TOOL NUMBER:

COA DDT 003

Study Endpoints and Labeling Development (SEALD) Review

SEALD Review
Elektra J. Papadopoulos, MD, MPH
COA DDT 003
Exacerbations of Chronic Pulmonary Disease Tool (EXACT)

SEALD TRACKING NUMBER	2013-055
APPLICATION NUMBER	DDT# 003
LETTER DATE/SUBMISSION NUMBER	April 19, 2013
REVIEW COMPLETION DATE	September 30, 2013
SEALD REVIEWER	Elektra J. Papadopoulos
SEALD DIRECTOR	Laurie B. Burke
SUBMITTER	EXACT-PRO Initiative United BioSource Corporation (Now Evidera)
CLINICAL OUTCOME ASSESSMENT TYPE	Patient-reported outcome (PRO)
CONCEPT OF INTEREST	Symptoms of acute bacterial exacerbation of chronic bronchitis in patients with chronic obstructive pulmonary disease (ABECB-COPD)
MEASURE(S)	Exacerbations of Chronic Pulmonary Disease Tool (EXACT)
INTENDED POPULATION(S)	Outpatients with ABECB-COPD

SEALD Review
Elektra J. Papadopoulos, MD, MPH
COA DDT 003
Exacerbations of Chronic Pulmonary Disease Tool (EXACT)

A. EXECUTIVE SUMMARY

This Clinical Outcome Assessment Qualification Review concludes that the electronically-administered patient-reported outcome (PRO) instrument, *Exacerbations of Chronic Pulmonary Disease Tool (EXACT),* is qualified for use as an endpoint for use in exploratory studies (typically phase 2 trials) for the measurement of symptoms of acute bacterial exacerbation of chronic bronchitis in patients with chronic obstructive pulmonary disease (ABECB-COPD). Ultimately, this tool is intended to be further developed for use as a primary or secondary endpoint in confirmatory clinical trials in patients with ABECB-COPD.

The specific context of use that defines the boundaries of the EXACT's application within the scope of this qualification statement is that of a superiority study in outpatients with a clinician diagnosis of ABECB-COPD as described within the September 2012 FDA guidance for industry: *Acute Bacterial Exacerbation of Chronic Bronchitis in Patients with Chronic Obstructive Pulmonary Disease: Developing Drugs for Treatment[1] (ABECB-COPD Guidance).* Accordingly, this qualification decision does not include the context of use of development of drugs for reduction in incidence of ABECB-COPD events in patients who are not experiencing an exacerbation. It also does not include studies of drugs for the treatment of COPD or COPD exacerbations caused by factors other than bacterial infection.

The submitter has demonstrated the content validity of the EXACT Total Score as a measure of symptoms of ABECB-COPD using the principles of instrument development delineated in the final FDA Guidance for Industry: *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims[2]* (PRO Guidance). Evidence that instrument items are relevant, understandable and complete in relation to the desired claims was well-documented and came from individual patient interviews, focus group sessions and cognitive debriefings conducted in a sample of subjects who adequately represented the targeted patient population of patients with ABECB-COPD. Additionally, a panel that included experts in pulmonary medicine, clinical research, and instrument development and translation served as advisors throughout the development process to ensure the instrument was developed appropriately for use in multinational clinical trials.

The submitter has also demonstrated evidence of reliability (test-retest and internal consistency), construct validity and ability to detect change in the setting of clinical trials.

The next step in the EXACT-PRO development should be quantitative evaluation in longitudinal studies within the qualified context of use to further evaluate construct validity and ability to detect change as well to provide interpretation guidelines for change (improvement or decline) over time. Evaluations should also include estimation of the effect size on EXACT total score of an antibacterial drug versus placebo that would enable appropriate efficacy assumptions and sample size calculations for confirmatory clinical trials.

---

[1]http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070935.pdf
[2] http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf

B. STUDY ENDPOINT REVIEW

1 BACKGROUND

ABECB-COPD is a patient-based experience with the initial detection of an event originating with the patient and the initial and ongoing status reports of severity relying on patient report to the clinician. Therefore, a standardize method for the assessment of exacerbation-related outcomes from the patient perspective is needed for use in international clinical trials.

To develop a PRO measure that would address this unmet need, the *EXACT-PRO Initiative* was formed which included the participation of CDER and multiple other stakeholders under the direction of Dr. Nancy Kline Leidy of United BioSource Corporation (UBC), now Evidera. Initial discussions with FDA took place in 2005 with the official start of the project taking place in early 2006.

The PRO Evidence Dossier describing the EXACT's initial instrument development was submitted to CDER on December 29, 2009. CDER concluded that the content validity of the instrument had been demonstrated based on documentation of literature review, expert input and patient input in the form of qualitative research in the targeted patient population. At that time, an earlier version of the instrument comprising 23 items had been administered within the context of an observational study; however, the final 14-item EXACT (representing a subset of the 23-item instrument) had not been psychometrically evaluated within the context of clinical trials.

On December 13, 2011, the results of psychometric analyses derived from the use of EXACT in the context of three clinical trials were submitted to CDER. Each of these studies was designed to assess the impact of treatment on patients who were enrolled with stable-state disease at baseline, specifically for the reduction of the 'frequency, severity, and/or duration of exacerbations in COPD'. These psychometric analyses were performed according to a statistical analysis plan provided to CDER *a priori.* An additional amendment was received on April 19, 2013 that describing analyses from the use of the EXACT in a phase 3 trial of a drug, aclidinium bromide (NDA 202,450), which was approved by FDA in 2012.

The documents referred to in this review comprise both the December 29, 2009 submission as well as the December 13, 2011 and April 19, 2013 submissions.

2 CONTEXT OF USE

2.1 Disease Definition

The EXACT Total Score measures symptoms of acute bacterial exacerbation of chronic bronchitis in patients with chronic obstructive pulmonary disease (ABECB-COPD).

According to the ABECB-COPD Guidance, ABECB-COPD refers to a clinical diagnosis of presumptive bacterial infection superimposed on the patient's chronic pulmonary condition. The acute component of ABECB-COPD is usually manifested as a worsening of the same symptoms patients experience when they are not experiencing an acute infection.

2.2 Target Population

The EXACT is for use in studies enrolling outpatients with ABECB-COPD and who meet the clinical trial entry criteria as described in the ABECB-COPD Guidance. Accordingly, patients with ABECB-COPD characterized as severe (e.g., requiring hospitalization) are excluded from the intended target population for the EXACT.

The EXACT has been translated for use in multiple language and culture groups as described in the instrument user manual.

2.3 Clinical trial design

The EXACT is for use in a placebo-controlled superiority trial of an antimicrobial therapy intended for use in outpatients enrolled with ABECB-COPD.

2.4 Endpoint positioning

The EXACT is intended for ultimate use as an efficacy endpoint measure (e.g., primary, co-primary or secondary endpoint) in confirmatory clinical trials. Because a responder definition for the targeted patient population is not yet available, we recommend the user evaluate clinically meaningful improvements and decrements in EXACT Total Score prior to using it as an efficacy endpoint measure in confirmatory clinical trials. Therefore, at this time, we are qualifying the EXACT for exploratory purposes (e.g., phase 2 clinical trials).

2.5 Labeling claim(s) based on the COA

The submitter has demonstrated the content validity of the EXACT as a measure of symptoms of ABECB-COPD. Therefore, CDER may grant a claim of improvement of ABECB-COPD symptoms based on EXACT Total score, if a clinically meaningful improvement is demonstrated

in adequate and well-controlled clinical trials that are in alignment with recommendations in the ABECB-COPD Guidance. As the EXACT has not yet been qualified for use as an effectiveness endpoint in confirmatory trials, we recommend that the user discuss plans to use it in this or other contexts with the relevant CDER Review Division.

2.6 Limitations of use

The EXACT was developed as an e-diary instrument and is suitable for administration via hand-held device. Score properties have not been assessed using other modes of administration (e.g., via pen and paper, interactive voice response system (IVRS), interviewer administered).

The EXACT is used to evaluate exacerbations treated in the out-patient setting; its measurement properties have not been assessed when used in acute care, hospital settings.

This qualification does not include use of the EXACT for the following uses:

- Use in patients with bronchiectasis, cystic fibrosis, alpha-1 antitrypsin deficiency, obliterative bronchiolitis or asthma;
- Use for development of drugs for reduction in incidence of ABECB-COPD events in patients who are **not** experiencing an exacerbation; or
- Use in non-inferiority trials.

The qualification does not apply to the subscores of the EXACT, but rather the EXACT Total Score.

Reviewer's comment: The EXACT was developed in the outpatient COPD patient population. While a subset of the focus group and interview participants discussed their symptom experiences leading up to and during hospitalization for COPD exacerbation, there is inadequate qualitative data to support use of the instrument during hospitalization. Questions such as "Were you short of breath today when performing your usual personal care activities like washing or dressing?" are irrelevant in the hospital setting and would suggest inadequate content validity of the instrument for symptom measurement in hospitalized patients.

3 CONCEPT OF INTEREST AND CONCEPTUAL FRAMEWORK

The EXACT Total Score measures symptoms of ABECB-COPD, i.e., an acute, sustained, and worsening of signs and symptoms beyond day-to-day variability.

The instrument's total score is made up of a total of 14 items representing the following domains:
- Breathlessness (5 items),
- Cough and sputum (2 items),
- Chest symptoms (3 items),
- Difficulty bringing up sputum (1 item),
- Tired or weak (1 item),
- Sleep disturbance (1 item), and
- Scared or worried (1 item).

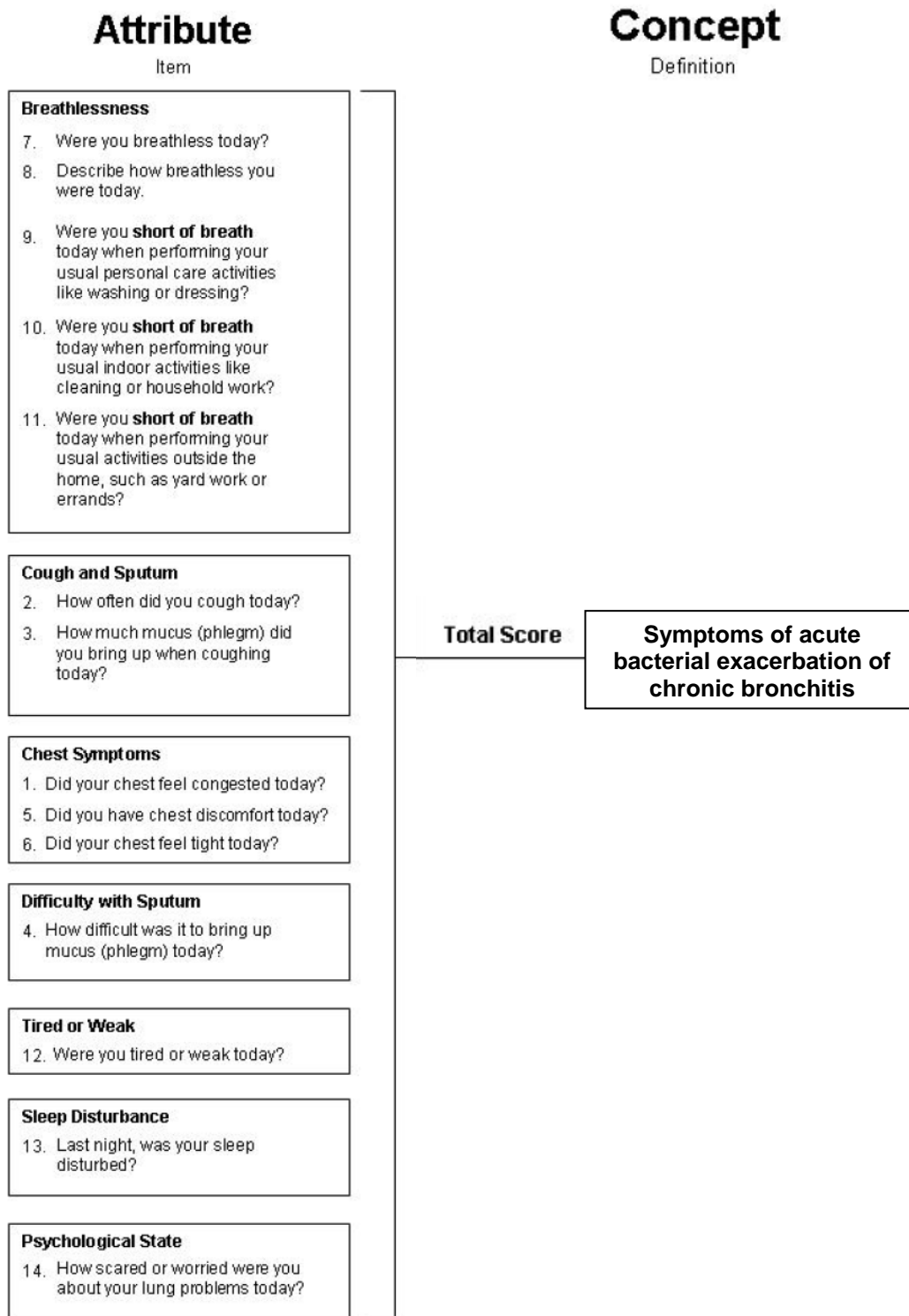The following figure represents the instrument conceptual framework.

Figure 1 EXACT Conceptual Framework

4   METHOD OF ADMINISTRATION

The EXACT is a daily diary, completed each evening before bedtime. The instrument was developed with e-diary administration in mind, with cognitive interviews performed with paper-pen booklet and personal digital assistant (PDA) to document respondent understanding in either mode and user acceptance of the PDA. The screen shots of the instrument are appended to this review.

The Palm® Tungsten E2 was used in this qualitative study and in the initial quantitative validation study. Since then, a variety of other hand-held devices have been used. Therefore, the EXACT user manual includes specifications for administering the EXACT via electronic devices, information related to device selection, and a description of a vendor certification program to maximize consistent use across vendors and devices. This includes formatting and layout specifications. For example, the user manual specifies that each item and its response options should fit on a single screen. If the users wish to use a device where scrolling is necessary or the item and response options do not fit on a single screen, additional usability testing or equivalence studies are recommended.

Additionally, the user manual specifies features to help standardize the use of the EXACT and maximize respondent compliance. For example, the patient enters and updates the bedtime as needed the diary is to be available only during a specified window. The use of alarms is also recommended to maximize compliance. To preserve the integrity of the tool, e-PRO vendors are not licensed or permitted to distribute the EXACT or to modify the EXACT, nor are they permitted to translate the instrument independently of UBC.

Finally, the user manual includes specific advice for training of patients, investigational sites and vendors on the appropriate use of the EXACT.

5   SCORING ALGORITHM

The EXACT Total Score is computed across the 14 items and has a theoretical range of 0 to 100, with higher values indicating a more severe condition.

As the first step in scoring, item responses are converted to item-level raw scores. Some of the items (Items 3, 8, 9, 10, 11, and 14) have unique scoring instructions, whereby certain original item response codes are grouped into a single item-level raw score. Assignment of item-level raw scores can be found in Appendix B. Next, the item-level raw scores of the 14 EXACT items are summed to form the raw summed score. As the final step, an EXACT Total score is computed from the raw summed score using a conversion table for each day of diary collection. The EXACT Total score is based on a logit scoring system and ranges from 0-100.

In addition to the EXACT Total score, there are three domain scores embedded in the measure: Breathlessness, Cough and Sputum, and Chest Symptoms. These scores also range from 0 to 100 with higher scores indicating more severe symptoms. The user manual also provides instructions for computing domain scores. As is shown in Figure 1, items are assigned to domains as follows:

- Breathlessness – Items 7, 8, 9, 10, and 11;
- Cough & Sputum – Items 2 and 3; and
- Chest Symptoms – Items 1, 5, and 6.

Items 4 (difficulty bringing up mucus), 12 (tired or weak), 13 (sleep disturbance), and 14 (scared or worried) do not correspond to a domain score.

Reviewer's comments: As stated previously in this review, this qualification is limited to the EXACT Total Score. Accordingly, it does not include domain scores.

According to the user manual, if the EXACT Total score is 0, it should be changed to "missing". This scoring rule is based on previous validation work demonstrating that moderate to severe COPD patients typically experience symptom(s) each day. Therefore, the submitter has suggested that a score of zero on all 14 EXACT items is likely to represent a situation where in order to complete the diary quickly, the respondent did not accurately report their daily symptom(s).

The EXACT user manual also includes specific guideline regarding monitoring for missing observations. Patients are not allowed to skip items; therefore, no missing data as expected for individual items.

Reviewer's comment: Our review of the qualitative and quantitative evidence concludes that the EXACT Total Score represents the measurement concept reflected in the instrument conceptual framework and is adequate to directly determine the appropriate labeling claim.

6   CONTENT VALIDITY

Content validity is defined as evidence that the instrument measures the concept of interest including evidence from qualitative studies that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population and use. Development and validation procedures for the EXACT were consistent with the methods for PRO instrument development described in the PRO Guidance.

The submitter completed a comprehensive review of the published literature and obtained input from experts in the US and Europe in the areas of pulmonary medicine, clinical research, instrument development and pharmaceutical regulation. These sources informed the development of qualitative study protocols and interview guides for focus groups or interviews to elicit concepts for instrument development and develop the instrument's item content.

The following is an overview of the methods used and the timelines for instrument development:

- Review of the literature and expert input (beginning in 2005);

- Focus groups and structured interviews resulted in a draft instrument comprising 23 items (February/March 2006);

- Cognitive debriefing interviews using draft items (March-June 2006);

- Cognitive debriefings with the final instrument in PDA format (August 2006); and

- Expert panel critique and recommendations, a total of 4 (April 2006-July 2008).

Reviewed in this section are the following items: (1) the methods of item generation; (2) the methods of qualitative analysis; (3) documentation of saturation; (4) development of the item pool which included Rasch methodology in combination with clinical expert advice; (5) description of scoring; (6) description of recall period; (7) evidence of patient understanding; and (8) review of respondent burden.

6.1   Patient Characteristics (Qualitative Research Sample)

Participants were recruited through physicians' offices within the US.

Key inclusion/exclusion criteria from the concept elicitation and cognitive debriefing study protocols were as follows:

Key Inclusion Criteria:
1. ≥ 40 years of age;
2. Smoking for at least 10 pack-years;
3. Current diagnosis of COPD or chronic bronchitis:
    a. COPD as defined by the GOLD Initiative
        i. GOLD – 0 indicates the patient is "at risk". This is characterized by chronic cough and sputum production. Lung function, as measured by spirometry, is still normal.
        ii. GOLD-1 indicates "mild COPD". This stage is characterized by mild airflow limitation and usually, but not always, chronic cough and sputum production. FEV1/FVC < 70%; FEV1 $\geq$ 80% predicted.
        iii. GOLD-2 indicates "moderate COPD". This stage is characterized by worsening airflow and usually the progression of symptoms, with shortness of breath typically developing on exertion. FEV1/FVC < 70%; 50% $\leq$ FEV1 < 80% predicted.
        iv. GOLD-3 indicates "severe COPD". This stage is characterized by further worsening of airflow limitation, increased shortness of breath, and repeated exacerbations. FEV1/FVC < 70%; 30% $\leq$ FEV1 < 50% predicted.
        v. GOLD-4 indicates "very severe COPD". This stage is characterized by severe airflow limitation or the presence of respiratory failure or clinical signs of right heart failure. FEV1/FVC < 70%; FEV1 < 30% predicted or FEV1 < 50% predicted plus chronic respiratory failure.
    b. Bronchitis is chronic if cough and sputum production both occur for at least 3 months in 2 consecutive years with or without airflow obstruction
4. History of acute exacerbations with an acute event in the past 6 months:
    a. Exacerbation defined as a sustained worsening of the patient's condition from the stable state and beyond normal day-to-day variations that is acute in onset and necessitates a change in regular medication in a patient with underlying COPD;
    b. Exacerbation may be associated with a telephone call or an unexpected clinic, emergency department or hospital visit.

Key Exclusion Criteria:
1. Current diagnosis of asthma with no obstructive disease (post bronchodilator >80%; FEV1/FVC $\geq$ 70%) and no chronic bronchitis
2. Current diagnosis of clinically relevant bronchiectasis

A description of the patient demographic and clinical characteristics of participants in the qualitative studies is as follows.

Focus groups and structured patient interviews took place in Phoenix, AZ, Tampa, FL, and Bethesda, MD, in February-March 2006. Patients had a mean age of 68.4 years, 15 (37%) were male, and 40 (100%) were Caucasian. Patients had been diagnosed with COPD 5.8 years and the mean FEV_1 was 49% predicted.

Eight patients who had experienced an exacerbation within the last 10 days were interviewed in 1:1 sessions. Interviews for 7 patients were via telephone and for one patient an in-person interview was conducted. These patients had a mean age of 65.8 years, seven were Caucasian (88%), and two were male (25%). The mean number of years since diagnosis with COPD was 6.1 and the mean FEV1 was 34.7 % predicted.

To increase the ethnic and racial diversity of the qualitative sample, an additional 23 patients were interviewed in San Antonio, TX, Miami, FL and Flint, MI. Patients had a mean age of 58.3 years (10 years younger compared with the previous sample described above), 13 (56.5%) were male, 12 (52%) were Black or African-American, and 11 (48%) were Hispanic/Latino. The mean number of years since diagnosis with COPD was 11.4, 11 (47.8%) were classified as GOLD-IV, and the mean FEV1 was 41% predicted.

The following table summarizes the qualitative research sample demographic and clinical characteristics.

Table 1 Qualitative Research Sample Demographic and Clinical Characteristics

INSTRUMENT DEVELOPMENT PHASE

Characteristics	EXACT Development		Further Development with Debriefing of Draft Items		Final Debriefing	Total Sample (n=83)
	Focus Groups (n=40)	1:1 Interviews (n=8)	1:1 Interviews and Cognitive Debriefing (n=23)	Cognitive Debriefing (n=3)	Cognitive Debriefing with PDA (n=9)	
Demographic						
Age, Mean (SD)	68.4 (7.7)	65.8 (8.0)	58.3 (10.2)	65.3 (6.1)	69.1 (5.9)	65.3 (9.7)
Gender, Male n (%)	15 (38%)	2 (25%)	13 (57%)	1 (33%)	6 (67%)	37 (45%)
Race/Ethnicity n (%)						
White	40 (100%)	7 (88%)	0 (0%)	3 (100%)	9 (100%)	59 (71%)
Hispanic or Latino	0 (0%)	0 (0%)	11 (48%)	0 (0%)	0 (0%)	11 (13%)
Black or African American	0 (0%)	1 (13%)	12 (52%)	0 (0%)	0 (0%)	13 (16%)
Clinical						
COPD Dx – Years, Mean (SD)	5.8 (4.1)	6.1 (2.5)	11.4 (5.9)	18.0 (13.7)	9.8 (4.7)	8.3 (6.4)
PFT[*], Mean (SD)						
FEV_1 (L)	1.2 (0.4)	0.9 (0.2)[1]	1.0 (0.4)	1.0 (0.2)	1.3 (0.5)	1.1 (0.4)[1]
FEV_1 % Predicted	48.6 (14.3)	34.7 (10.9)	41.0 (18.1)	39.3 (9.9)	43.4 (17.1)	44.4 (15.8)
GOLD Stage n (%)						
I ($FEV_1/FVC < 70\%$; $FEV_1 \geq 80\%$ predicted)	2 (5%)	0 (0%)	0 (0%)	0 (0%)	1 (11%)	3 (4%)
II ($FEV_1/FVC < 70\%$; $50\% \leq FEV_1 < 80\%$ predicted)	22 (55%)	1 (12%)	2 (9%)	0 (0%)	0 (0%)	25 (30%)
III ($FEV_1/FVC < 70\%$; $30\% \leq FEV_1 < 50\%$ predicted)	14 (35%)	5 (63%)	10 (43%)	3 (100%)	5 (56%)	37 (44%)
IV ($FEV_1/FVC < 70\%$; $FEV_1 < 30\%$ predicted or $FEV_1 < 50\%$ predicted plus chronic respiratory failure)	2 (5%)	2 (25%)	11 (48%)	0 (0%)	3 (33%)	18 (22%)

[*] Stable state.

12

Reviewer's comments: There was inadequate representation of subjects who were not Caucasian in the first qualitative sample. This was addressed with a second round of 1:1 structured interviews conducted with a sample that included more racial and ethnic diversity, with a specific focus on COPD patients who were African American or Hispanic/Latino. In sum, the subjects who participated in the qualitative research adequately represent those expected to participate in clinical trials of ABECB-COPD in terms of their demographic and clinical characteristics. See also the clinical review by Dr. Joseph Toerner.

6.2   Concept elicitation

The submitter developed a hypothesized conceptual framework for the instrument based upon literature review and expert input. Concept elicitation study protocols and interview guides were developed with the aim to elicit patient descriptions of features and essential elements of an exacerbation.

The interview guide for the focus group interviews included questions to elicit the reasons that patients sought medical attention and how they felt the day before the call to their physician, the day of their call to the physician and the day after. Symptoms were then elicited in an open-ended non-leading manner followed by probing questions on specific concepts. Other questions were aimed at ascertaining what signaled the patient to call their physician and how the symptoms change over time. Special attention was given to the words and phrases that patients used to describe their symptoms.

Audio transcripts were coded and transcribed for systematic analysis of the qualitative data using Atlas.ti version 5.0. Four team members, including the two focus group moderators, reviewed the data and developed a draft coding scheme. Data were then pilot-coded using the draft coding scheme, and revisions were made to the codes for clarity and comprehensive coverage of content. Data were independently coded by two team members. Coded transcripts were compared, and discrepancies were resolved through consensus. The coded data were stratified and examined by gender and disease severity to evaluate consistency across patients.

Patients interviewed comprised a total of 40 patients with clinician-confirmed exacerbations in the previous 6 months and an additional 8 patients who had experienced an exacerbation within the previous 10 days. Among patients who had experienced exacerbation within the previous 6 months, there were a total of four focus groups (34 subjects), two 2:1 structured interviews (4 subjects) and two 1:1 interviews (2 subjects). Among the 8 subjects who had experienced an exacerbation within the previous 10 days, all participated in individual (i.e., 1:1) interviews.

Due to the high proportion of Caucasians in the first qualitative sample, a second round of 1:1 structured interviews was conducted with 23 non-Caucasian patients, with a specific focus on COPD patients of African American or Hispanic/Latino descent. Qualitative data gathered during this second round of 1:1 structured interviews were examined for any new terminology, descriptions and themes that should be used to revise the initial draft item pool.

The submitter reported that the major complaint voiced was difficulty breathing. In the majority of patients interviewed (7 out of 8), other symptoms (e.g., congestion and cough) were mentioned when probed. However, the complaint that led the patients to call their physician was typically breathlessness, which often had an impact on the patients' activities.

Examples of patient expressions from the qualitative research are as follows.

Difficulty breathing:
- *Not getting enough air in*
- *Couldn't grab enough air*
- *Having trouble breathing*
- *Laboring to breathe*
- *Struggling for air*
- *Fighting to get air*
- *Extreme shortness of breath*

Activity-related:
- *I was short of breath doing activities where I wasn't normally short of breath*
- *I had to gasp for air every few steps*
- *When I managed to get [out] of bed, I might get out of breath just trying to do some normal things like empty the dishwasher*
- *Don't want to do my normal activities, like walking, and having a little difficulty breathing*

Patients frequently also described a sensation of "chest tightness" associated with not being able to "take in enough air." When asked about whether sleep was disturbed, some patients referred to the need for using pillows to prop themselves up at night. Coughing was also described as being associated with sleep disruption. Patients also frequently described feeling "scared" and "panicky."

Although the data from the second set of qualitative interviews were consistent with those of the first qualitative sample, the submitter noted a higher proportion of patients in the second set of interviews reporting wheezing (6/23) compared with the first sample (4/48). Because a number of patients in the second sample had a concurrent diagnosis of asthma, the submitter concluded that a reference to wheezing represented poorly controlled asthma rather than an attribute of a COPD exacerbation. Therefore, there are no items on wheezing in the EXACT.

Reviewer's comment: This reviewer reviewed the transcripts as well as the Atlas.ti summaries to ascertain whether there were important elements from the transcripts that were not included in the Atlas.ti summaries and also whether there were any elements from the Atlas.ti output summary that were not supported by the transcripts. In sum, everything from the Atlas.ti summary was found to be supported by the transcripts. There also did not appear to be any major themes found in review of the transcripts that were not captured in the Atlas.ti summaries.

Saturation was defined as two focus group discussions and two 2:1 or 1:1 structured interview groups in which no new concepts were introduced by the participants beyond those identified by previous participants. Saturation is documented in the form of a saturation table. The submitter provided a saturation table (Table 1) demonstrating that saturation was reached for the following concepts: exacerbation terminology and definitions, respiratory symptoms (cough, sputum, chest discomfort and difficulty breathing), fatigue or general malaise, sleep disturbance, anxiety and activity limitation.

Table 2 Saturation Table

	Focus Group 1 n=4	Focus Group 2 n=11	Focus Group 3 n=15	Focus Group 4 n=4	Structured 2:1 and 1:1 Interviews in patients <6 months from exacerbation n=6	Structured 1:1 interviews in patients <10 days from exacerbation n=8	Structured 1:1 interviews with minority patients n=23
Exacerbation Terms							
Event							
Episode		x	x			x	x
Sudden onset	x	x	x	x	x	x	x
Persistent (2-3 days)	x	x		x	x	x	x
Cough							
Cough	x	x	x	x	x	x	x
Chest fullness	x	x	x	x	x	x	x
Congestion	x		x	x	x	x	x
Coughing frequently		x	x	x	x	x	x
Sputum							
Mucus				x		x	x
Phlegm	x	x	x	x	x	x	x
Sputum color change	x	x	x	x	x	x	x
Unusual presence of phlegm/mucus	x	x	x	x	x	x	x
Difficulty coughing phlegm/mucus up	x	x	x	x	x	x	x
Chest discomfort							
Chest hurt		x	x	x	x	x	x
Chest tight		x	x		x	x	x
Chest heavy		x	x			x	x

SEALD Review
Elektra J. Papadopoulos, MD, MPH
COA DDT 003
Exacerbations of Chronic Pulmonary Disease Tool (EXACT)

	Focus Group 1 n=4	Focus Group 2 n=11	Focus Group 3 n=15	Focus Group 4 n=4	Structured 2:1 and 1:1 Interviews in patients <6 months from exacerbation n=6	Structured 1:1 interviews in patients <10 days from exacerbation n=8	Structured 1:1 interviews with minority patients n=23
Difficulty breathing							
Short of breath		x	x	x	x	x	x
Difficulty breathing		x	x		x	x	x
Can't catch breath	x	x	x		x	x	x
Trouble breathing	x	x	x	x	x	x	x
Couldn't breathe	x	x	x	x	x	x	x
Hard to breathe at rest		x		x	x	x	x
Breathless	x						
Out of breath	x		x	x		x	x
Wheezing			x		x	x	x
Activity Impact							
Instrumental ADLs (eating, talking, bathing, brushing hair/teeth, etc)	x	x	x	x	x	x	x
ADLs (cooking, dishes, etc)		x	x	x		x	x
Other indoor activities (housework, laundry, misc indoor)		x	x	x		x	x
Outdoor activities (gardening, mowing lawn, driving, going out)	x	x	x	x	x	x	x
Walking	x	x	x	x	x	x	x
Pace		x	x	x	x	x	x
Emotional State							
Anxious		x		x			x
Close to death	x		x				x
Distress				x		x	
Panic			x			x	x
Scared		x	x		x	x	x
Worried	x						x
Miscellaneous							
Trouble Sleeping	x	x	x	x		x	x
Tired	x	x	x	x	x		x
Weak	x			x		x	x
No energy		x	x		x	x	x

6.3   Item Development

Following concept elicitation and documentation of saturation of concept, the submitter continued the process of instrument development in a manner consistent with the principles in the PRO Guidance. This included generation of new items based upon patient input, drafting the instrument, conducting cognitive interviews to document patient understanding, pilot testing the draft instrument and documenting content validity.

The submitter provided an item tracking matrix in appendix K of the qualitative report (February 19, 2009), which is not reproduced here. This table summarizes changes to the items up to and including the cognitive debriefing interviews through the final version, which resulted from a process of item reduction.

Based on the results of the initial focus groups and structured interviews, a pool of items was drafted and discussed among UBC team members and members of the EXACT-PRO Initiative Expert Panel during the first meeting. Several rounds of item pool revisions were conducted with input from experts in instrument development, PRO translation, and COPD to ensure readability, translatability, and clinical relevance. Adjustments to the draft item pool included grammatical edits and minor changes in wording to facilitate consistency in wording across future translations.

The outcome of this development process was a draft EXACT item pool of 23 items found in the cognitive debriefing protocol dated June 5, 2006 (Appendix C). Cognitive debriefing interviews with patients who had experienced an exacerbation within the previous 6 months were then carried out using this version. Following cognitive interviews, changes were made to the draft item pool. Examples follow:
- The item 'Were you fatigued?' (Item 17) was dropped from the instrument because patients in cognitive interviews did not understand the word 'fatigue'.
- Item 21 in the June 5, 2006 version included the word "anxious" but based on cognitive interviews, patients took 'anxious' to mean a variety of emotions and suggested 'scared or worried' to describe feeling in exacerbation and the item was revised accordingly.
- Additionally, the branch patterns that were used in the initial cognitive interviews (for items 13, 14, and 15) were removed for the subsequent iteration of the instrument. The following figure exemplifies these branch patterns.
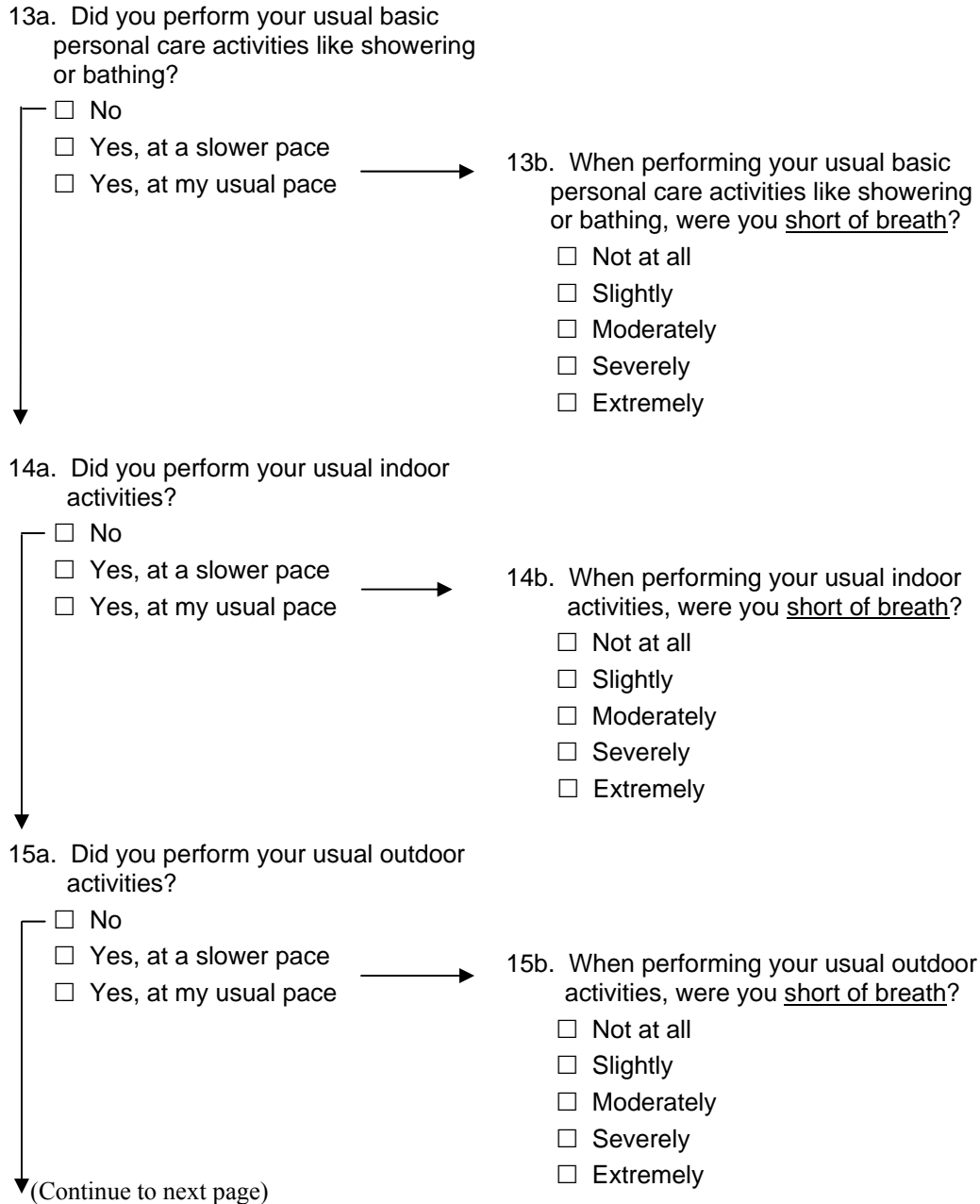
Figure 2 Branch patterns in the June 5, 2006 item pool that were removed in subsequent revisions

13a. Did you perform your usual basic personal care activities like showering or bathing?

- ☐ No
- ☐ Yes, at a slower pace
- ☐ Yes, at my usual pace

13b. When performing your usual basic personal care activities like showering or bathing, were you short of breath?

- ☐ Not at all
- ☐ Slightly
- ☐ Moderately
- ☐ Severely
- ☐ Extremely

14a. Did you perform your usual indoor activities?

- ☐ No
- ☐ Yes, at a slower pace
- ☐ Yes, at my usual pace

14b. When performing your usual indoor activities, were you short of breath?

- ☐ Not at all
- ☐ Slightly
- ☐ Moderately
- ☐ Severely
- ☐ Extremely

15a. Did you perform your usual outdoor activities?

- ☐ No
- ☐ Yes, at a slower pace
- ☐ Yes, at my usual pace

15b. When performing your usual outdoor activities, were you short of breath?

- ☐ Not at all
- ☐ Slightly
- ☐ Moderately
- ☐ Severely
- ☐ Extremely

(Continue to next page)

Additional cognitive debriefing interviews were performed with a new set of patients to evaluate the revised draft item pool and assess the ease or difficulty with which patients used the PDA. The submitter reported that the results from this phase of interviews demonstrated that the instructions and the items comprising the item pool were clear and the content represented the important aspects of their exacerbation experience. No new concepts or items were introduced

as a result of these debriefings. Eight of the nine participants preferred the electronic mode of administration over pen-paper. All of the participants reported that the PDA was easy to use, expected it would take less time to complete than a pen-paper diary, and stated that they would not have a problem using the device in a study. Additional information on documentation of patient understanding based on the cognitive interviews is found in Section 6.4 of this review.

Following the final phase of cognitive interviews, the draft instrument was piloted in an observational study.

A prospective, two-group, observational study was conducted in the US (Jan-May 2007) to reduce the number of items comprising the tool and evaluate the psychometric properties of the EXACT in patients with clinician-confirmed exacerbation (Acute Group) and those considered clinically stable (Stable Group).

The sample included 410 patients with COPD, 222 patients experiencing an exacerbation (Acute Group) and 188 stable patients (Stable Group). The Acute Group completed the EXACT daily via PDA on Days 1-28 and 60-67 (no data were collected between Days 28 and 60). The Stable Group completed the diary for seven days with visits on Days 1 and 7.

A systematic item-reduction process was used to reduce the number of items from the initial item set of 23 items to the 14 items used in the final instrument. A quantitative study summary (Study A2-7031, Feb 20, 2009) included a description of the item-reduction process. Importantly, several conceptually redundant items were included in the pool with the intent of selecting the item with the best empirical performance. For example, there were three distinct items targeting the concept of chest discomfort: 'Did you have chest discomfort today?'; 'Did your chest hurt today?' and 'Did your chest feel tight today?' However, during item reduction one of these items ('Did your chest hurt today?') was dropped and the other two were retained in the final instrument and, appropriately, the concept of chest discomfort is retained in the final instrument conceptual framework.

Item performance was evaluated to inform item reduction, instrument scaling, and scoring according to pre-specified set of analyses. In sum, the 23-item EXACT PRO item pool underwent a step-wise item reduction process through item analysis and Rasch analysis. The item analysis included evaluation of properties such as floor and ceiling effects, inter-item correlation, and item-total correlation.

The submitter used exploratory factor analysis, which demonstrated that the data were unidimensional (which is an assumption of the Rasch model) prior to undertaking Rasch analyses for assessment of overall model fit and individual item fit.

During Rasch analyses, the response option ordering for each item was reviewed by examining the threshold map and categorical probability curves. Individual item model fit and differential item functioning were tested and items were removed through an iterative process. The item-reduction process identified 14 items comprising a single dimension.

According to the quantitative study report, final Rasch analysis results showed an overall chi-square of 149.3 (df=84); the person-separation index was 0.92 indicating good discriminant power. Model fit parameters showed a good fit to a unidimensional model, consistent with a unidimensional concept of exacerbation symptom severity and supporting the use of a single, total score to represent this concept.

The following table summarizes how Rasch modeling as well as descriptive statistics were used together to inform the item-reduction and item-refinement process. Item reduction was also conducted with consideration of the qualitative analysis and with consultation from clinical research experts.

Table 3 EXACT Item Summary Table (Table 2.23)

Item Number and Question	Item deleted	Floor or Ceiling Effect	Inter-item Correlation > 0.8 or Item-total Correlation < 0.2	Exploratory Factor Analysis Loading < 0.3	Rasch Model Analysis (X=non-ordered, XX=mis-fit)	Rationale
Item 1 – Chest congested						
Item 2 – Cough today						
Item 3 – Much mucus when coughing					XX	Rasch model misfit – collapsed response 'a little' and 'some' at Round 8
Item 4 – Difficulty with mucus						
Item 5 – Color mucus	X	X			X	Non-ordered responses – deleted at Round 2
Item 6 – Chest discomfort						
Item 7 – Chest hurt	X					Measurement scale too difficult, rescale by deleting most difficult item – deleted at Round 10
Item 8 – Chest tight						
Item 9 – Breathless today						
Item 10 – How breathless today					X	Non-ordered responses – collapsed responses 'severely' and 'extremely' at Round 2
Item 11 – SOB while sitting	X		X		XX	Rasch model misfit and local independent with Item

Item Number and Question	Item deleted	Floor or Ceiling Effect	Inter-item Correlation > 0.8 or Item-total Correlation < 0.2	Exploratory Factor Analysis Loading < 0.3	Rasch Model Analysis (X=non-ordered, XX=mis-fit)	Rationale
						12 – deleted at Round 9
Item 12 - Difficulty breathing while sitting	X		X			Measurement scale too difficult, rescale by deleting most difficult item – deleted at Round 11
Item 13 - How active	X		X	X		Low correlation with total score - deleted at Round 1
Item 14 - Usual personal care	X	X				Measurement scale too difficult, rescale by deleting most difficult item – deleted at Round 13
Item 15 – SOB with personal care					X	Non-ordered responses – collapsed responses 'extremely' and 'too breathless' at Round 2
Item 16 – Usual indoor	X				XX	Rasch model misfit – deleted at Round 4
Item 17 – SOB with indoor					X	Non-ordered responses – collapsed responses 'extremely' and 'too breathless' at Round 2; collapsed responses 'severely', 'extremely' and 'too breathless' at Round 3;
Item 18 - Usual outdoor	X	X		X	XX	Rasch model misfit – deleted at Round 4
Item 19 – SOB with outdoor					X	Non-ordered responses – collapsed responses 'extremely' and 'too breathless' at Round 2; collapsed responses 'severely', 'extremely' and 'too breathless' at Round 3;
Item 20 – Weak/tired						
Item 21 – Sleep disturbed						
Item 22 - How much sleep	X			X		Low factor loading – deleted at Round 2
Item 23 – Scared/worried					X	Non-ordered responses – collapsed responses 'severely' and 'extremely' at Round 2

Reviewer's comments: Different sources of information guided item pool development, including expert input, input from the FDA, input from patients, and Rasch methodology. An item tracking matrix (record of the development of items used in an instrument) was provided in the qualification package and documented the changes or deletions in items and the reasons for those changes. The methods used to finalize the items were well-documented in the qualification review package and were consistent with the principles of the PRO Guidance and support the content validity of the final EXACT.

6.4 Description of evidence of patient understanding

A total of 35 patients participated in the final phase of cognitive debriefings on the EXACT.

Timeframe
Participants cognitively debriefed on the draft EXACT versions dated June 5, 2006 and August 22, 2006 reported that the questionnaire was to be completed before bed every night, and that they were to reflect on the day's experiences. Patients also said they would have no trouble recalling the events of the day during an exacerbation. To keep patients grounded in a single day, the word "today" was added to each item in the draft EXACT item pool dated August 22, 2006.

Instructions
Patients reported that the instrument was to be completed each evening before going to bed. Participants stated that they understood how to complete the items—i.e., by reflecting on their exacerbation experience that day and marking the response option that most closely matched that experience.

Comment: The EXACT was developed as a daily diary. The recall period of the EXACT is "today." The EXACT's recall period is consistent with the advice provided in the PRO guidance and is appropriate given that symptomatic experience varies from day to day and a daily measure would be needed to capture this variability.

Item Stems:
Item stems were revised or modified based on input from patients and from instrument development, clinical, and translation experts. The majority of revisions occurred after each round of cognitive debriefings.

Response options:
In the first round of cognitive debriefing with the draft EXACT item pool dated June 5, 2006, response options for several items were tested by giving participants a set of cards, each with a different response option.

Participants were asked to place each response option along a 0–100 scale, with higher numbers indicating greater severity.

According to the qualitative study report, more than half of the participants who ordered the intensity response options (shown to participants with Item 1) confirmed that the order was appropriate. Some patients switched "severely" and "extremely."

Comment: The qualitative study report did not specify the percentage of patients who did not confirm the intended order of the response options. The cognitive debriefings indicated that some patients switched the order of 'severely' and 'extremely'. This might be addressed in part by the instrument's scoring algorithm whereby these two response options are assigned the same raw score for some of the items.

In keeping with the intention that the EXACT be administered via electronic diary, the draft EXACT item pool was entered into a PDA system (screenshots dated December 11, 2006) for patient evaluation. Directions for the draft EXACT item pool completed on the PDA instructed patients to tap on the boxes to record a response.

Reviewer's comment: Overall, the evidence submitted indicates that content validity was confirmed with cognitive debriefing or other forms of patient interview and demonstrated that items are comprehensive and that patients understand the questions asked and respond in a way that was intended.

6.5   Review of respondent burden

Patients with COPD who are experiencing acute exacerbation often have difficulty with the basic activities of daily living. According to the PRO dossier, the instrument developers considered twice daily administration, however, to reduce respondent burden, it was decided to administer the EXACT once a day, in the evening prior to bedtime.

The instrument developers also sought to minimize the length of the questionnaire, while preserving the tool's reliability and validity. The final tool contains a total of 14 items.

Comment: The length of the tool and the frequency of administration are important considerations in PRO instrument development. In this case, the instrument developers sought to minimize respondent burden. Studies indicate good overall compliance with the EXACT, which further supports the acceptability of the instrument. However, a decrease in compliance was shown during hospitalization.

7 OTHER MEASUREMENT PROPERTIES (RELIABILITY, CONSTRUCT VALIDITY, ABILITY TO DETECT CHANGE) AND DESCRIPTIVE STATISTICS

The primary dataset used to evaluate the psychometric properties of the EXACT in a clinical trial context was from a trial conducted by Mpex Pharmaceuticals (MPEX-302). EXACT Total score psychometric evaluations were also performed using two other clinical trial datasets AZ Study 12 and AZ Study 20. The final form EXACT was used in each of these clinical trials. The analyses presented in this review using these three clinical trials were received on December 13, 2011.

The submitter provided a statistical analysis plan (SAP) on September 3, 2010 that delineated how the measurement properties of the EXACT would be evaluated. The SAP included a plan for construct validity evaluation using *a priori* hypotheses concerning logical relationships that should exist with other measures of characteristics of patients and patient groups.

In each of the three trials, medically treated exacerbations (MTEs) were defined in terms of patient-report of healthcare resource utilization for exacerbations of COPD, including clinic or emergency room visits with antibiotic and/or systemic corticosteroid treatment or hospitalization for exacerbation of COPD.

The SAP delineated the following objectives.

Table 4 Submitter's description of planned analyses from statistical analysis plan (September 3, 2010)

Part 1:	
Sample and EXACT descriptive statistics	
	Assess patient compliance in completing the EXACT
	Describe sample demographic and clinical characteristics
	Present descriptive statistics for the EXACT
	Describe EXACT-determined event frequency, severity, and duration of exacerbations using the proposed scoring algorithms for these outcomes
Part 2:	
EXACT measurement properties: reliability and validity	
	Assess the reliability (internal consistency, reproducibility) and validity (construct, criterion-related of the EXACT score in the context of a clinical trial
Part 3:	
Comparison of Exacerbations Identified by EXACT and Medically Treated Exacerbations	
	Compare exacerbations based on the EXACT and medically treated exacerbations (MTE) with regard to frequency, severity, and duration
Part 4:	
Efficacy Analysis	

	1. Frequency of exacerbations
	
	   a. Primary: To test whether MP-376 is effective in patients with COPD when compared to placebo, with regard to its effect on the frequency of exacerbations, as defined by the EXACT.
	
	   b. Exploratory: To test whether MP-376 is effective in patients with COPD when compared to placebo, with regard to its effect on the time to first exacerbation, as defined by the EXACT.
	
	   c. Exploratory: To test whether MP-376 is effective in patients with COPD when compared to placebo, with regard to its effect on the proportion of exacerbation-free days, as defined by the EXACT.
	
	2. Duration of exacerbations
	
	   a. Primary: To test whether MP-376 is effective in patients with COPD, when compared to placebo, with regard to its effect on the duration of exacerbations, as defined by the EXACT.
	
	3. Severity of exacerbations: To test whether MP-376 is effective in patients with COPD when compared to placebo, with regard to its effect on the severity
	
	   a. Primary: To test whether MP-376 is effective in patients with COPD, when compared to placebo, with regard to its effect on the severity of exacerbations, as defined by EXACT.
	
	   b. Exploratory: To test whether MP-376 is effective in patients with COPD when compared to placebo, with regard to its effect on the severity of breathlessness, cough and sputum.

On November 22, 2010, the submitter provided an addendum to the SAP that included plans for supplementary Rasch model analyses in order to confirm the properties of the original EXACT scoring algorithm.   It also included further specifications for sensitivity analyses around criteria for defining exacerbation onset and recovery.

MPEX-302 was a phase 2, multi-center, randomized, double-blind, placebo-controlled study to evaluate the safety, tolerability and efficacy of MP-376 (levofloxacin inhalation solution) administered for 5 Days every 28 Days to prevent acute exacerbations in COPD patients at high risk for exacerbation.

Reviewer's comment: It is important to keep in mind that the each of these three clinical trials were designed to evaluate the effect of the drug on the frequency of exacerbation events in COPD patients who are stable at enrollment and prospectively assessed for the development exacerbation events. This is different from the context of use for which the EXACT is qualified, which is for a trial enrolling patients diagnosed by their clinician as having ABECB-COPD. However, the trials provided valuable information on the test-retest reliability of the EXACT Total Score, which necessitates evaluation in a stable patient population. The next step in the development of the EXACT should be an evaluation of its use in the intended application (e.g., a placebo-controlled trial of an antimicrobial treatment in outpatients experiencing an ABECB-COPD) to further evaluate its longitudinal psychometric properties.

The MPEX-302 included the following study design:

- At least 300 COPD patients would be enrolled from 45 sites in the United States and treated for a minimum of 6 months.

- Following a screening period of up to 14 days, patients were and randomized (Day 1, Cycle 1) in a 2:1 ratio to study drug (n=200) or placebo (n=100).

- The study drug was administered twice daily over the first five days of each 28 day-cycle.

- Each participant was followed for a minimum of six and up to twelve 28-day treatment cycles.

- Patients were to meet the following key inclusion criteria:

  o Forced expiratory volume in one second (FEV1) ≤70% of predicted (post-bronchodilator administration) and FEV1/Forced vital capacity (FVC) ≤0.7 (post-bronchodilator)
  o Minimum of two documented acute exacerbation episodes during the preceding 12 months

- No acute exacerbation of COPD that required treatment within 30 days prior to baseline
- Stable on chronic therapy with inhaled long acting bronchodilators and/or inhaled or systemic steroids for preceding 30 days from baseline (if receiving)
- History of COPD with mucopurulent sputum production on most days even when exacerbation-free

A statistical analysis plan (SAP) for the psychometric evaluation was developed *a priori* and included descriptive statistics, reliability and construct validity analyses, and comparisons with medically treated exacerbations (MTEs).

The measurement properties described below are derived using the EXACT in its final form (i.e., the 14-item EXACT).

Analyses performed on the Mpex data were based on data from 235 participants.

In this study, baseline is the seven day period prior to when the patient receives their first dose of study drug on Cycle 1, Day 1 (Day -7 to Day -1).

The **Baseline EXACT Total Score** is the mean within-patient score over this seven day period (Day -7 to Day -1) with a minimum of four daily values required for calculation. If fewer than four days of data are available, then a baseline score was not to be calculated.

For a participant's data to be included in the validation analyses, at least four baseline days of EXACT data and >80% of daily data were required, from baseline to final visit or early termination date.

For participants who did not experience an EXACT-based event during the first 28-day cycle of the study, the participant's baseline was to be reset. The **Stable Reset Baseline** was defined as the mean within-person score for the last 7 days of each 28-day cycle. At least 4 of 7 days were required to reset the baseline EXACT score; otherwise, the previous one was to be used until the ensuing cycle.

7.1 Descriptive Statistics

7.1.1 Compliance

Compliance rates across the three clinical trials are summarized in the following table. Compliance was defined as the total number of actual completed diaries divided by the number of diary entries expected.

Table 5 Summary of EXACT Compliance

	Mpex (N=235)	AZ Study 12 (N=749)	AZ Study 20 (N=597)
Baseline to Final Visit	87%	94%	97%
Baseline to First EXACT Exacerbation	95%	95%	97%
Compliance during hospitalizations (COPD hospitalization ± 7 days)	69%	63%	61%

The summary of compliance demonstrated compliance rates of 87% -97% across the three clinical trials. Compliance dropped during hospitalization to the range of 61% -69% across trials.

Reviewer's comment: A lower compliance rate during hospitalization was evident possibly limiting the utility of the EXACT in hospitalized patients. The qualified context of use is for outpatients with COPD.

7.1.2 Demographics

The demographic characteristics of the clinical trial patient populations are summarized in the following table.

Table 6 Demographic Characteristics

	Mpex (N=235)	AZ Study 12 (N=749)	AZ Study 20 (N=597)
Age, mean (SD) range	63.7 (8.95) 41 – 83	62.3 (8.25) 40 – 80	61.7 (8.27) 41 – 80
Gender, % male	113 (48.1%)	572 (76.4%)	443 (74.2%)
Race/Ethnicity			
Hispanic or Latino	6 (2.6%)	2 (0.3%)	2 (0.3%)
Asian	2 (0.9%)	212 (28.3%)	0 (0.0%)
Black/African American	18 (7.7%)	0 (0.0%)	0 (0.0%)
White	214 (91.1%)	536 (71.6%)	592 (99.2%)

The mean age was 62-64 years of age across the three studies. In the Mpex study, 48% were male. In the AZ studies the majority of patients were male (74%-76%). In all three studies the majority of patients were White, 72%-99% across studies.

A summary of key clinical characteristics is shown below.

Table 7 Key Clinical Characteristics

	Mpex (N=235)	AZ Study 12 (N=749)	AZ Study 20 (N=597)
FEV_1, mean (SD) range	1.2 (0.58) 0.2 – 3.6	1.7 (0.49) 0.6 – 3.3	1.6 (0.52) 0.6 – 3.6
FEV_1% predicted, mean (SD) range	42.2 (18.10) 7.2 – 99.0	58.8 (12.70) 22.8 – 107.3	54.2 (15.36) 22.8 – 105.9
GOLD stage, n (%)			
0	22 (9.4%)	0 (0.0%)	0 (0.0%)
I	2 (0.9%)	23 (3.2%)	22 (3.8%)
II	55 (23.4%)	503 (69.1%)	300 (51.9%)
III	89 (37.9%)	197 (27.1%)	240 (41.5%)
IV	66 (28.1%)	5 (0.7%)	16 (2.8%)
# of acute exacerbations in past 12 months, mean (SD) range	2.1 (0.70) 0.0 – 5.0	1.4 (1.05) 1.0 – 14.0	1.2 (0.50) 1.0 – 4.0
SGRQ Total Score*	57.4 (16.21) 16.7 – 96.7	52.6 (18.36) 4.9 – 100.0	54.9 (17.14) 11.0 – 97.1

*SGRQ-C was used in AZ 12 and 20.

The mean FEV1% predicted ranged from 42%-59% across the three studies. Of note, the FEV1% predicted ranges indicated that a subset of patients enrolled did not meet clinical trial

entry criteria evidenced by a normal FEV1% predicted. In addition, a relatively high percentage of patients were classified as GOLD stage 0 (9%), in the Mpex trials. In both AZ trials, there were no patients enrolled who were classified as GOLD stage 0.

7.2   Cross-sectional measurement properties

7.2.1   Reliability (Test-retest and internal consistency)

The internal consistency reliability of the EXACT Total Score was assessed using data from baseline (Day -1). Estimates were based on Cronbach's alpha and the person-separation index, which is based upon Rasch scoring logits. For each parameter, values can range from 0-1.0, with higher estimates indicating increased reliability.

The test-retest reliability of the EXACT total score was calculated using data from the clinical trial stable run-in period (Day -7 - Day -1), before randomization.

The results for the internal consistency reliability as well as the test-retest reliability are shown in Table 1.

Table 8 Reliability Estimates for EXACT Total Score

Reliability Parameter	Mpex (N=235)[1]	AZ Study 12 (N=735)[1]	AZ Study 20 (N=586)[1]
Internal consistency			
Cronbach's alpha	0.90	0.94	0.94
Person-separation index	0.94	0.94	0.94
Reproducibility			
Intraclass correlation coefficient (ICC)[1]	0.70	0.75	0.77

[1]Random-effects model

Note: Test-retest reliability was calculated using data from the clinical trial run-in period (Day -7 - Day -1)

These results indicate that the EXACT Total Score has acceptable internal consistency reliability and that the score is reproducible in stable conditions. It is important to note that symptoms of COPD can vary from day to day even in patients who are considered clinically stable, which can have an impact on the test-retest reliability observed over the course of one week. In the Mpex study, reproducibility over two consecutive days when patients reported their lung condition was unchanged was also estimated and was found to have an intraclass correlation coefficient (ICC) of 0.84.

7.2.2   Construct validity

Construct validity is defined in the PRO guidance as evidence that relationships among items, domains, and concepts conform to *a priori* hypotheses concerning logical relationships that should exist with other measures or characteristics of patients and patient groups.

Criterion validity is defined as the extent to which the scores of a PRO instrument are related to a known gold standard measure of the same concept.

It is important to note that for many PROs, criterion validity cannot be measured because there is no 'gold standard' measure. There is not a 'gold standard' measure of the concept of symptoms of ABECB-COPD. It was for this reason that the consortium was convened to develop the EXACT.

It was hypothesized that the EXACT Total scores should be related to patient reports of their health status. The relationship between EXACT scores (Baseline, Day 1) and patients' baseline reports of their health status using the SGRQ total and domain scores (Baseline, Day 1) prior to treatment were assessed via Spearman's rank-order correlation coefficient.

It was also hypothesized that the EXACT Total Score should be related to patient symptoms as measured by the Breathlessness, Cough and Sputum Scale (BCSS©)[3]. The BCSS© is a three-item questionnaire that assesses patient's breathlessness, cough and sputum using a Likert scale with zero representing the improvement symptoms and a 4 indicating worsening of symptoms.

---

[3] Leidy NK, Rennard SI, Schmier J, Jones MK, Goldman M. The Breathlessness, Cough, and Sputum Scale: The Development of Emperically Based Guidelines for Interpretation. *Chest.* 2003;124:2182−91.

Table 9 Correlations between Day 1 EXACT Total scores and other clinical variables

Criterion Variable	Mpex (N=225–227)	AZ Study 12 (N=712–731)	AZ Study 20 (N=567–585)
SGRQ[3]	0.62***	0.46***	0.46***
Symptom	0.37***	0.38***	0.42***
Activity	0.54***	0.41***	0.36***
Impact	0.58***	0.42***	0.44***
FEV$_1$% predicted	-0.14*	-0.10*	-0.13**
GOLD stage	0.18**	0.10*	0.15***
BCSS[2]	Not available	0.83***	0.83***
Breathlessness	Not available	0.74***	0.69***
Cough	Not available	0.66***	0.80***
Sputum	Not available	0.68***	0.74***

[1]Spearman's rank-order correlation and p-value; Sample size range a function of missing data.
[2]Visit 2, Day 1; AZ trials only
[3]The AZ trials used the St. George's Respiratory Questionnaire for COPD (SGRQ-C).
*$p<0.05$; **$p<0.01$; ***$p<0.001$.

The above table is reproduced from the sponsor's submission. We do not agree that the variables utilized for construct validation represent 'criterion' variables in the sense that they do not represent "gold standard" measures of the concept of interest.

The submitter presented the correlations between EXACT Total Score and FEV1% predicted and GOLD stage for descriptive purposes. Because the EXACT is not a measure of airflow obstruction, it was expected (and shown) that the EXACT Total score would not show any meaningful relationship with either of these two variables.

In response to an information request, the submitter also provided a summary of changes in domain scores that were observed in patients experiencing an MTE in the Mpex trial. In the majority of patients (56%), an increase in all three domains (Breathlessness, Chest Symptoms, and Cough and Sputum) was observed on the day of the MTE. The Breathlessness domain increased in the majority of patients consistent with the qualitative research findings that showed breathlessness to be the primary symptom alerting patients to seek medical care.

The mean EXACT Total Score at baseline ranged from 42 to 46 across the three studies.

Of the patients exceeding the EXACT Total Score thresholds for exacerbation (i.e., 12 points for 2 days and 9 points for 3 days) in the Mpex study, AZ Study 12 and AZ Study 20, the mean severity level (i.e., highest single score) of the first event was 57 (N=110), 58 (N=241) and 61 (N=173), respectively.

As described in the statistical analysis plan, the EXACT Total Score was evaluated for its concordance with MTEs. In the Mpex study, the majority of MTEs did not have a corresponding EXACT event. Specifically, of the 137 MTEs that were recorded in the study, 60 (44%) also met the definition of an event by EXACT scoring.[4] Approximately half of the MTEs in AZ Study 12 and AZ Study 20 did not have a corresponding EXACT event ±4 day from MTE onset.[5]

It is important to note that EXACT is proposed as a measure of symptom severity and not a diagnostic tool. While the submitter included these thresholds with the intent of documenting onset of exacerbation in patients who were stable (non-exacerbating) at baseline, these thresholds of event onset are not relevant to the qualified context of use in which patients are enrolled with an MTE and their symptom severity as measured by the EXACT Total Score is measured over time.

7.2.3   Ability to detect change

The PRO Guidance states that the Agency will review an instrument's ability to detect change using data that compare change in PRO scores to change in other similar measures that indicate that the patient's state has changed with respect to the concept of interest.

The submitter presents analyses evaluating changes in EXACT scores over time in patients who are stable with respect to COPD symptoms at baseline and who are followed prospectively for the development of medically treated exacerbation events (MTE's).

The EXACT Total scores for the 7 days prior to the MTE onset and for 21 days after MTE onset (Day 1) are presented graphically below for the Mpex study data (Figure 3). Day 1 was the day the patient was seen in the clinic or hospitalized.

A moderate MTE was defined as one where the patient was treated with either antibiotics or corticosteroids, but was not hospitalized. A severe MTE was defined as an event necessitating hospitalization.

---

[4] Performance of the EXACT in a Controlled Clinical Trial: Final Report_Mpex; Table 20. Frequency of Medically Treated Exacerbations (MTEs) with Corresponding EXACT-based Exacerbations
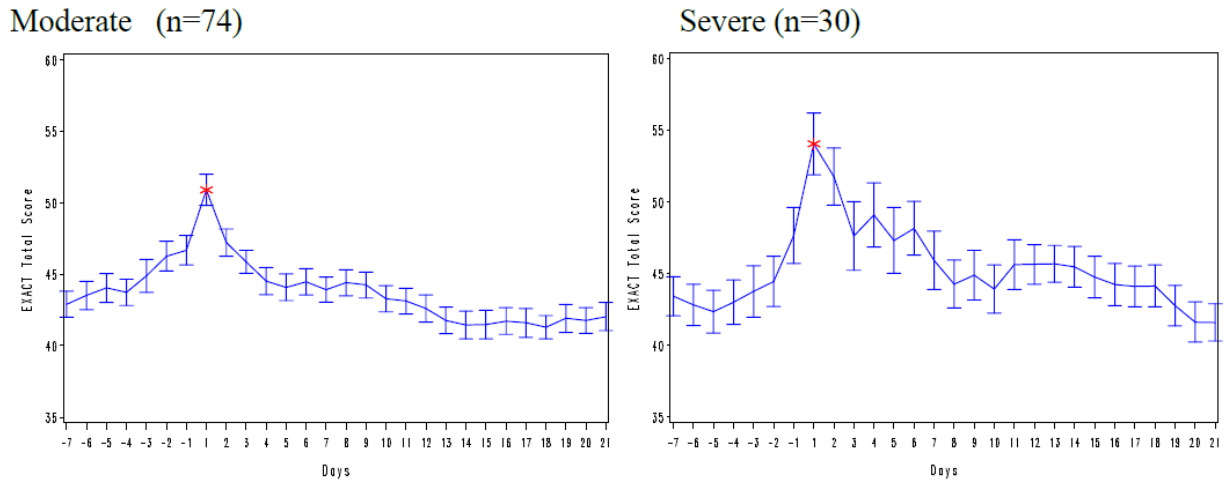[5] Page 35 of Performance of the EXACT in a Controlled Clinical Trial of AZD9668: Final Report

Figure 3 Mean EXACT Total Scores (±SE) Day -7 to Day 21 of a Medically Treated Exacerbation by Severity

*Mpex Study, N= # subjects

These graphs indicate that the EXACT Total scores demonstrate ability to detect change at the group level and track with both moderate as well as severe MTE's. Similar findings were also noted using data from AZ 12 and AZ 20 and were presented in the submission, but are not reproduced here.

In addition, the submitter provided information suggesting that the EXACT can detect differences between treatment groups in a clinical trial (ATTAIN) (Figure 4).

Reviewer's comment: The data presented in Figure 4 below were presented by Paul Jones et al and have not been reviewed in detail by the Agency.
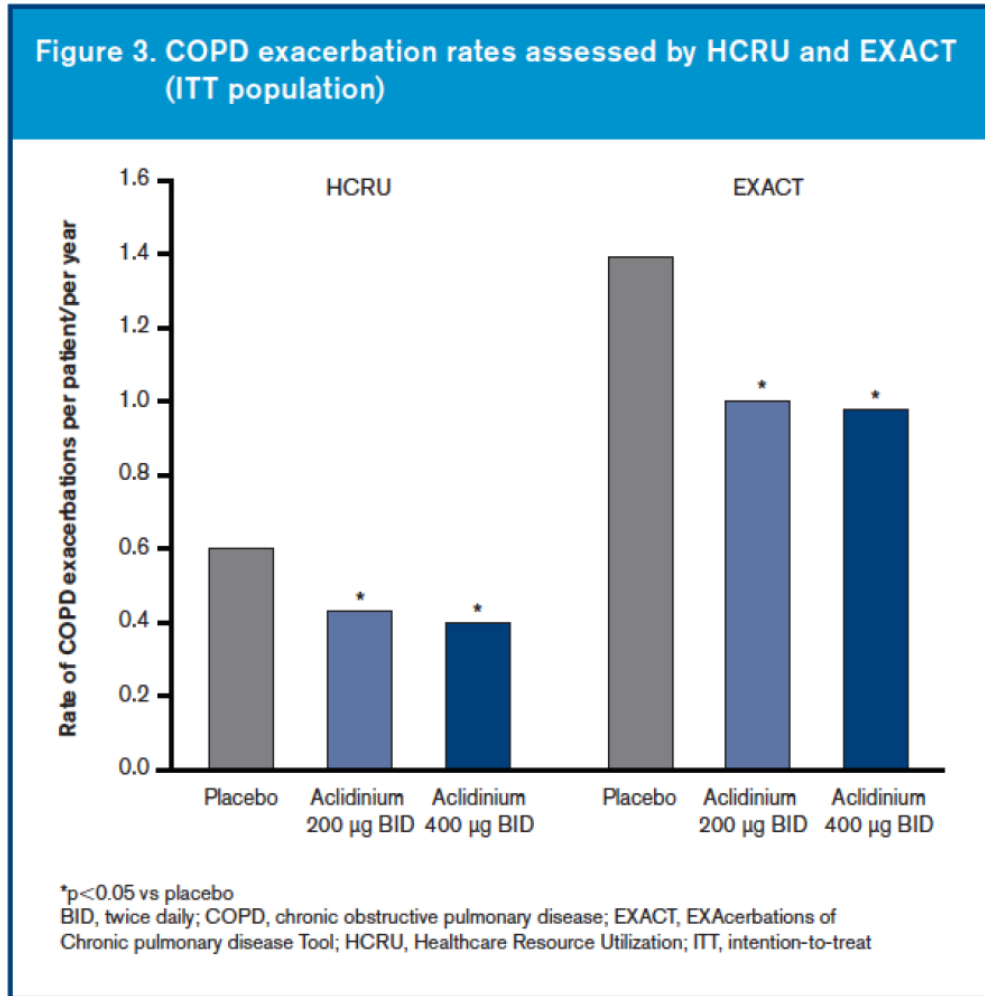
Figure 4 Effect of Aclidinum on Exacerbation Rate

Figure 3. COPD exacerbation rates assessed by HCRU and EXACT (ITT population)

*p<0.05 vs placebo
BID, twice daily; COPD, chronic obstructive pulmonary disease; EXACT, EXAcerbations of Chronic pulmonary disease Tool; HCRU, Healthcare Resource Utilization; ITT, intention-to-treat

Source: Jones P, et al. Aclidinium bromide in patients with chronic obstructive pulmonary disease (COPD): reduction in exacerbations as defined by health-care utilization and the EXACT diary card. Poster presented at the American College of Chest Physicians Annual Congress; October 2011; Honolulu, Hawaii.

8   INTERPRETATION OF SCORES

Planning for interpretation of clinical trial results includes developing a responder definition for the context of use, i.e., the individual patient score change over a predetermined time period that should be interpreted as a treatment benefit.

A responder definition for the targeted patient population is not yet available; this should be evaluated in longitudinal studies and prespecified before use of the EXACT as a primary or secondary endpoint in confirmatory clinical trials.

9 LANGUAGE TRANSLATION AND CULTURAL ADAPTATION

Translation methods for the EXACT follow the Principles of Good Practice for Translation and Cultural Adaptation, an ISPOR Task Force Report, including item definition, dual forward translation; reconciliation; dual back translation; back translation review; harmonization; in-person cognitive testing with COPD patients in each target country using a standardized interview script; analysis of cognitive testing results; clinician review as-needed to verify terminology; finalization; and dual proofreading. This methodology was followed to ensure that the translated versions of the EXACT are both conceptually equivalent to the English source version and easily understood by the target population.

To ensure consistency across translations, UBC developed an item definition document (IDD) which was distributed to all linguistic teams. The IDD provided linguists with the item stems and item response options, as well as the intended meaning/interpretation of terms in the item/response options.

Foreseeable translation issues and points of clarification were also outlined and linguists were provided with a list of acceptable and unacceptable alternative terms or phrases to consider when necessary. Linguists were also instructed that for words and phrases that are repeated throughout the EXACT, it was imperative that a consistent translation be created (e.g., "moderately" was to be translated consistently for each response option), or if this was not possible that the reason be carefully detailed in the report.

A list of all language/culture groups for which the EXACT has been translated is provided at the instrument website (www.exactproinitiative.com) and is updated as new translations become available. Thus far, the EXACT has been linguistically validated in 42 languages, for use in 47 countries.

The translation certificates for each translation are available from the instrument developer upon request.

10 REFORMATTING FOR NEW METHOD OR MODE OF ADMINISTRATION

The EXACT is intended for use as an electronic diary to be completed by the patient each evening at bedtime. The user manual includes specifications concerning the formatting and methods of administration to protect the integrity of the instrument. If the instrument format is modified outside of these specifications, then additional usability testing or equivalence studies are recommended prior to use in a clinical trial.

Appendix A

EXACT Screen Shots: Final Instrument

EXACT

Standardized instruction given to patients with PDA training and with take-home instruction manual:
"Please complete your diary every evening, just before you go to bed."

Daily Diary

As you answer the following questions, please select the option that best describes your experience.

[Exit diary] [Next]

Question 1 of 14

Did your chest feel congested today?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely

[Back] [Next]

Question 2 of 14

How often did you cough today?

○ Not at all
○ Rarely
○ Occasionally
○ Frequently
○ Almost constantly

[Back] [Next]

Question 3 of 14

How much mucus (phlegm) did you bring up when coughing today?

○ None at all
○ A little
○ Some
○ A great deal
○ A very great deal

[Back] [Next]

Question 4 of 14

How difficult was it to bring up mucus (phlegm) today?

○ Not at all
○ Slightly
○ Moderately
○ Quite a bit
○ Extremely

[Back] [Next]

Question 5 of 14

Did you have chest discomfort today?

○ Not at all
○ Slight
○ Moderate
○ Severe
○ Extreme

[Back] [Next]

Question 6 of 14

Did your chest feel tight today?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely

[Back] [Next]

Question 7 of 14

Were you breathless today?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely

[Back] [Next]

Question 8 of 14

Describe how breathless you were today:

○ Unaware of breathlessness
○ Breathless during strenuous activity
○ Breathless during light activity
○ Breathless when washing or dressing
○ Present when resting

[Back] [Next]

Question 9 of 14

Were you short of breath today when performing your usual personal care activities like washing or dressing?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely
○ Too breathless to do these

[Back] [Next]

Question 10 of 14

Were you short of breath today when performing your usual indoor activities like cleaning or household work?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely
○ Too breathless to do these

[Back] [Next]

Question 11 of 14

Were you short of breath today when performing your usual activities outside the home such as yard work or errands?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely
○ Too breathless to do these

[Back] [Next]

Question 12 of 14

Were you tired or weak today?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely

[Back] [Next]

Question 13 of 14

Last night, was your sleep disturbed?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely

[Back] [Next]

Question 14 of 14

How scared or worried were you about your lung problems today?

○ Not at all
○ Slightly
○ Moderately
○ Severely
○ Extremely

[Back] [Next]

Appendix B

Annotated EXACT for Raw Score Assignment

The following annotates the raw item-level score values associated with each text response for the EXACT items. Please take note of items with collapsed response scale scoring, highlighted in **bold.**

1. Did your chest feel congested today?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	4. Extremely
2. How often did you cough today?	0. Not at all
	1. Rarely
	2. Occasionally
	3. Frequently
	4. Almost constantly
3. How much mucus (phlegm) did you bring up when coughing today?	0. None at all
	1. A little
	1. Some
	2. A great deal
	3. A very great deal
	NOTE: Score "a little" and "some" the same.
4. How difficult was it to bring up mucus (phlegm) today?	0. Not at all
	1. Slightly
	2. Moderately
	3. Quite a bit
	4. Extremely
5. Did you have chest discomfort today?	0. Not at all
	1. Slight
	2. Moderate
	3. Severe
	4. Extreme

6. Did your chest feel tight today?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	4. Extremely
7. Were you breathless today?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	4. Extremely
8. Describe how breathless you were today:	0. Unaware of breathlessness
	1. Breathless during strenuous activity
	2. Breathless during light activity
	3. Breathless when washing or dressing
	3. Present when resting
	NOTE: Score "Breathless when washing or dressing" and "Present when resting" the same.
9. Were you short of breath today when performing your usual personal care activities like washing or dressing?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	3. Extremely
	4. Too breathless to do these
	NOTE: Score "severely" and "extremely" the same.
10. Were you short of breath today when performing your usual indoor activities like cleaning or household work?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	3. Extremely
	3. Too breathless to do these
	NOTE: Score "severely", "extremely", and "Too breathless to do these" the same.

11. Were you short of breath today when performing your usual activities outside the home such as yard work or errands?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	3. Extremely
	3. Too breathless to do these
	NOTE: Score "severely", "extremely", and "Too breathless to do these" the same.
12. Were you tired or weak today?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	4. Extremely
13. Last night, was your sleep disturbed?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	4. Extremely
14. How scared or worried were you about your lung problems today?	0. Not at all
	1. Slightly
	2. Moderately
	3. Severely
	3. Extremely
	NOTE: Score "severely" and "extremely" the same.

Appendix C

EXACT DRAFT –Version 4 (June 6, 2006)

This questionnaire is to be completed every evening, just before you go to bed.

Please check the box that best describes your experience **today.**

1. Did your chest feel congested?
   - ☐ Not at all
   - ☐ Slightly
   - ☐ Moderately
   - ☐ Severely
   - ☐ Extremely

2. How often did you cough?
   - ☐ Not at all
   - ☐ Rarely
   - ☐ Occasionally
   - ☐ Frequently
   - ☐ Very Often
   - ☐ Almost constantly

3. How much phlegm/mucus did you bring up when coughing?
   - ☐ None at all
   - ☐ A little
   - ☐ Some
   - ☐ A great deal
   - ☐ A very great deal

4. How difficult was it to bring up phlegm/mucus?
    - ☐ Not at all
    - ☐ Slightly
    - ☐ Moderately
    - ☐ Quite a bit
    - ☐ Extremely

5. What color was your phlegm/mucus?
    - ☐ Clear or White
    - ☐ Tan or Grey
    - ☐ Yellow
    - ☐ Green
    - ☐ Brown

6. Did you have chest discomfort?
    - ☐ Not at all
    - ☐ Slight
    - ☐ Moderate
    - ☐ Severe
    - ☐ Extreme

7. Did your chest hurt?
    - ☐ Not at all
    - ☐ Slightly
    - ☐ Moderately
    - ☐ Severely
    - ☐ Extremely

8. Did your chest feel tight?
    - ☐ Not at all
    - ☐ Slightly
    - ☐ Moderately
    - ☐ Severely
    - ☐ Extremely

9. Did your chest feel heavy?
    ☐ Not at all
    ☐ Slightly
    ☐ Moderately
    ☐ Severely
    ☐ Extremely

10. How breathless were you?
    ☐ None/unaware of difficulty
    ☐ Mild: noticeable during strenuous activity
    ☐ Moderate: noticeable during light activity
    ☐ Marked: noticeable when washing/dressing
    ☐ Severe: almost constant, present when resting

11. Were you short of breath at rest?
    ☐ Not at all
    ☐ Slightly
    ☐ Moderately
    ☐ Severely
    ☐ Extremely

12. Did you have difficulty breathing at rest?
    ☐ Not at all
    ☐ Slight
    ☐ Moderate
    ☐ Severe
    ☐ Extreme

Continue to next page ⟶

48

13a. Did you perform your usual basic personal care activities like showering or bathing?

☐ No
☐ Yes, at a slower pace
☐ Yes, at my usual pace

13b. When performing your usual basic personal care activities like showering or bathing, were you short of breath?

☐ Not at all
☐ Slightly
☐ Moderately
☐ Severely
☐ Extremely

14a. Did you perform your usual indoor activities?

☐ No
☐ Yes, at a slower pace
☐ Yes, at my usual pace

14b. When performing your usual indoor activities, were you short of breath?

☐ Not at all
☐ Slightly
☐ Moderately
☐ Severely
☐ Extremely

15a. Did you perform your usual outdoor activities?

☐ No
☐ Yes, at a slower pace
☐ Yes, at my usual pace

15b. When performing your usual outdoor activities, were you short of breath?

☐ Not at all
☐ Slightly
☐ Moderately
☐ Severely
☐ Extremely

(Continue to next page)

16. Were you tired?
- ☐ Not at all
- ☐ Slightly
- ☐ Moderately
- ☐ Severely
- ☐ Extremely

17. Were you fatigued?
- ☐ Not at all
- ☐ Slightly
- ☐ Moderately
- ☐ Severely
- ☐ Extremely

18. Last night, was your sleep disturbed?
- ☐ Not at all
- ☐ Slightly
- ☐ Moderately
- ☐ Severely
- ☐ Extremely

19. How much did you sleep over the past 24 hours?
- ☐ Much less than usual
- ☐ Less than usual
- ☐ As usual
- ☐ More than usual
- ☐ Much more than usual

20. Last night, did you wake up because of your lung problem?
- ☐ Not at all
- ☐ Yes, once or twice
- ☐ Yes, frequently
- ☐ Awake for most of the night

21. Were you anxious today?
    ☐ Not at all
    ☐ Slightly
    ☐ Moderately
    ☐ Severely
    ☐ Extremely

22. Overall, how did you feel today compared to yesterday?
    ☐ Much better
    ☐ Somewhat better
    ☐ A little better
    ☐ About the same
    ☐ A little worse
    ☐ Somewhat worse
    ☐ Much worse

23. Overall, how was your lung condition today compared to yesterday?
    ☐ Much better
    ☐ Somewhat better
    ☐ A little better
    ☐ About the same
    ☐ A little worse
    ☐ Somewhat worse
    ☐ Much worse