U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

DDT Qualification	The Exacerbation of Chronic Pulmonary Disease Tool Patient Reported Outcome (EXACT-PRO) Instrument
Context of Use:	Measure of symptoms of acute bacterial exacerbation of chronic bronchitis in patients with chronic obstructive pulmonary disease (ABECB-COPD) for use in Phase 2 trials
Developer:	Evidera
Biometrics Division:	DBIV
Statistical Reviewer:	Scott Komo, Dr.P.H.
Concurring Reviewers:	Lisa Kammerman, Ph.D.
Medical Division:	Division of Anti-Infective Products
Qualification Review Team:	Joseph Toerner, M.D., M.P.H. Elektra Papadopoulos, M.D.

Keywords: qualification, patient reported outcomes

1. Executive Summary

The Developer submitted the EXAcerbations of Chronic Pulmonary Disease Tool (EXACT™) patient-reported outcome (PRO) instrument for qualification as a Drug Development Tool (DDT). The intent of EXACT is to quantify frequency, severity, and duration of acute exacerbations in clinical trials of chronic obstructive pulmonary disease (COPD) including those with chronic bronchitis. The EXACT is designed as an electronic diary made up of fourteen items to be completed by the patient each evening just prior to bedtime.

The Developer submitted qualitative data collected from focus groups, cognitive debriefings, and expert panel review. They also submitted quantitative data from an observational validation study of 410 COPD patients. Two hundred twenty-two (222) patients were experiencing an acute exacerbation at the time of enrollment and one hundred eighty-eight (188) patients were in a stable state at the time of enrollment. In addition, the Developer also submitted data from three prevention trials in COPD patients who were stable at baseline. This review will focus on the quantitative data.

Dr. Elektra Papadopoulos, the SEALD, reviewer, has concluded that the developer has demonstrated the content validity of the EXACT as a measure of symptoms of acute bacterial exacerbations of COPD (ABECB-COPD), see her review for details.

The prospective observational study provided evidence of the cross-sectional measurement properties of reliability and construct validity for the instrument. For COPD patients who are experiencing an acute exacerbation at baseline, the EXACT has been studied only in the observational study using an earlier version of the instrument, which contained twenty-three items. However, the similarity of the results from the observational study and the three prevention trials gives some confidence that that patients responded similarly to the 14- and 23-item versions of the instrument

There was some evidence of responsiveness provided in the observational study for acute patients wherein the mean EXACT total scores decreased as the exacerbation improved. However, this finding was tempered by the results from the prevention trials where a significant proportion of the patients who were experiencing a medically treated exacerbation (MTE) did not have a marked increase in their EXACT scores beyond the normal day-to-day variability of five points seen in the observational study of stable patients.

A major limitation is that the EXACT has not been studied in a treatment trial of patients who are experiencing an acute exacerbation at baseline. As a result, questions remain on the responsiveness of the instrument and on its ability to discriminate between and effective and ineffective treatment. Furthermore, we will have difficulty interpreting the results of the instrument because currently we do not know what constitutes a clinically meaningful change. Also we do not have the information needed to size a treatment trial. In addition, the poor concordance between medically treated exacerbations and EXACT-based exacerbations seen in the prevention trials is an additional concern. In the treatment setting, this poses a problem because patients will be enrolled based on clinical criteria not their EXACT scores. As a result, many subjects will have a baseline EXACT score that is similar to that seen when they are not

experiencing an exacerbation and it will likely be more difficult to demonstrate a treatment effect.

Further work is needed to investigate the longitudinal measurement properties of the instrument outlined above. These properties should be studied in a Phase 2 trial prior to using the instrument as a primary or secondary endpoint in a Phase 3 trial.

2. Introduction

The Developer submitted the EXAcerbations of Chronic Pulmonary Disease Tool (EXACT™) patient-reported outcome (PRO) instrument for qualification as a Drug Development Tool (DDT). The intent of EXACT is to quantify frequency, severity, and duration of acute exacerbations in clinical trials of chronic obstructive pulmonary disease (COPD) including those with chronic bronchitis. Exacerbations of COPD are events defined by an acute, sustained worsening of the patient's underlying condition of COPD from the stable state and beyond normal day-to-day variability which may require a change in treatment. The instrument is designed to be administered daily in hopes that it will capture the underlying day-to-day variability of the patient's COPD and detect worsening indicative of the presence of exacerbation. The EXACT is designed as an electronic diary made up of fourteen items to be completed by the patient each evening just prior to bedtime. The instrument includes assessments of Breathlessness (5 items), Cough and Sputum (2 items), Chest Symptoms (3 items), and four additional items (Difficulty with Sputum, Tired or Weak, Sleep Disturbance, and Psychological State). Each item is measured on a 5- or 6-point scale. The 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.

The Developer submitted qualitative data collected from individual patient interviews, focus groups, cognitive debriefings, and expert panel review. Dr. Elektra Papadopoulos, the SEALD, reviewer, has concluded that the developer has demonstrated the content validity of the EXACT as a measure of symptoms of acute bacterial exacerbations of COPD (ABECB-COPD), see her review for details.

The Developer also submitted quantitative data from an observational validation study of COPD patients, a portion of whom were experiencing an acute exacerbation at the time of enrollment and the rest were stable at the time of enrollment. In addition, the Developer also submitted data from three prevention trials in COPD patients. This review will focus on the quantitative data.

2.1 Proposed Context of Use for Qualification

This review will focus on the following context of use: the EXACT can be used as a primary, co-primary, or secondary endpoint in trials evaluating therapies to treat acute bacterial exacerbations of COPD (ABECB-COPD).

The recent Agency guidance (Guidance for Industry — Acute Bacterial Exacerbations of Chronic Bronchitis in Patients with Chronic Obstructive Pulmonary Disease: Developing Antimicrobial Drugs for Treatment) describes the disease as a clinical diagnosis of presumptive bacterial infection superimposed on a chronic pulmonary condition. The guidance points out that

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 Developer's Proposed Claim Language

Therapies to treat acute exacerbations of COPD:

- Mitigates/attenuates/reduces the severity of exacerbations
- Improves the symptoms of exacerbation
- Reduces/speeds time to recovery

2.3 Instrument Scoring

The EXACT total score is based on the sum of the fourteen items. The sum of the fourteen items is converted to a Rasch logit score. The Rasch logit score is then linearly transformed to a 0 – 100 scale to calculate the EXACT total score. The Developer also stated that daily EXACT total scores of 0 are set to missing.

The following definitions were used by the Developer:

- Three-Day Rolling Average
  Assessment of exacerbation duration utilizes a 3-day rolling of the average EXACT total score (including the day before and day after) with computation beginning on Day 2 of the event. Only one of the 3 data points need be present for this computation.
- Exacerbation Frequency
  Exacerbation frequency is defined as the number of events during a given period of time.
  - An exacerbation was defined as an increase of greater than or equal to nine points from baseline for at least three days or an increase of greater than or equal to twelve points from baseline for at least two days
  - Stable Baseline Variability is the patient's variability (SD) over a period of 7 days (minimum of 4 days) during which the patient is on maintenance therapy and rates his/her condition as stable or usual state.
  - Onset is the first day of the recorded increase in EXACT total score. Also referred to as Day 1 of the event.
- Exacerbation Severity
  Exacerbation severity is defined as the EXACT total score at the time of clinic visit or on the worst day of the exacerbation, or alternatively the area under the curve from day of onset to the day of improvement.
- Exacerbation Duration
  Exacerbation duration is defined as the length of time (days) between day of onset and day of improvement.
- Improvement
  Improvement is meant to capture meaningful and sustained decrease in EXACT total score. The Developer defined improvement (decline) in EXACT total score as a greater than 20% decrease from exacerbation onset (Day 1) for at least three consecutive days based on:
  - A 3-day rolling average, with computation beginning on Day 2 of the event.
  - The first of the 3 consecutive days of improvement is designated as the day of improvement

3. Development process

A 23-item EXACT was developed using qualitative methods, including focus groups, 1:1 interviews, and cognitive debriefing interviews involving 83 patients with COPD, with input from a team of experts in pulmonary medicine, clinical research, instrument development, and PRO translation methodology. Rasch analysis was used in an observational validation study to assess the psychometric properties of the instrument. Based on these analyses, fourteen items were selected for the final instrument with collapsing of categories for some of the items.

3.1 Observational validation study

After the qualitative development was finished, quantitative testing of reliability, validity, and responsiveness of EXACT was performed on the 23-item EXACT in an observational validation study of 410 COPD patients. Two hundred twenty-two (222) patients were enrolled during a clinic visit for acute exacerbation of COPD and completed the EXACT on Days 1-28 and again on Days 60-67. One hundred eighty-eight (188) patients were enrolled during a stable state and completed the EXACT on Days 1-7. In addition to collecting the EXACT diary data, clinical characteristics and demographic information were also collected, which included clinical history, pulmonary function (stable state within the previous 6 months to characterize underlying disease severity); St. George's Respiratory Questionnaire-COPD (SGRQC); Modified Medical Research Council Grading System (MMRC); physician assessment of patient's exacerbation manifestations (Acute Group); and patient and clinician global assessments of exacerbation severity (Acute Group), The majority (39%, 161/410) of the patients were classified as GOLD Stage II indicating that they had moderate COPD.

Using the data from the observational validation study, a Rasch analysis found that nine items should be deleted and that five items should have some of their response categories collapsed. These changes were implemented in the 14-item final instrument

It is important to note that the study was observational in nature and that patients completed the 23-item instrument even though only fourteen items were used in the calculation of the EXACT total score. There is a question whether patients would provide similar responses for the fourteen items used to calculate the EXACT total score if only these items were asked, rather than all 23 items.

3.2 Prevention clinical trials

In addition to the observational validation study, the submission also contains the results on the EXACT's (14-item) performance in three randomized, double-blind, placebo-controlled prevention trials of patients enrolled during a stable, non-exacerbating state. Exacerbations in these clinical trials were based on healthcare utilization, specifically receipt of antibiotics and/or systemic corticosteroids and/or hospitalization for the exacerbation. The performance of the EXACT was assessed in these trials by comparing exacerbations defined by healthcare utilization to those defined by changes in the EXACT total score based on the EXACT-defined exacerbation thresholds. Although the trials were for a different context of use, some information on the EXACT for the proposed context of use could be still obtained.

The 3 clinical trials are:

- Mpex Pharmaceuticals, Inc., Protocol MPEX-302
  A randomized, double-blind, placebo-controlled Phase 2 study to evaluate the safety, tolerability and efficacy of MP-376 inhalation solution administered for 5 days every 28 days to prevent acute exacerbations in high risk COPD patients. Patients were enrolled for a minimum of six 28-day cycles and up to a maximum of twelve 28-day cycles.
- AstraZeneca (AZ), Protocol D0520C00012
  A 12-week, randomized, double-blind, placebo-controlled, parallel group, multinational, Phase IIb dose range finding study to evaluate the efficacy and safety of AZD9668 administered orally at 3 dose levels to patients with COPD on treatment with tiotropium.
- AstraZeneca (AZ) Protocol D0520C00020
  A 12-Week, randomized, double-blind, placebo-controlled, parallel group, multinational, Phase IIb study to evaluate the efficacy and safety of 60 mg AZD9668 administered orally twice daily to subjects with COPD on treatment with budesonide/formoterol.

4. Psychometric Properties

The results of the psychometric properties for the observational study will be summarized below along with results from the prevention clinical trials that are relevant to the proposed context of use.

4.1 Reliability

In the observational study, reliability was assessed in the subset of observations from patients considered to be stable. The subset included data from Days 1-7 in the Stable group and Days 60- 67 in the Acute group of patients who were considered stable based on clinical-reported exacerbation state (completely resolved) and patient-reported exacerbation state (much better or returned to health).

Test-retest reliability in the Stable group (n=171) was 0.77 and 0.75 in the Acute group of patients (n=36) who were considered stable. In addition, the EXACT was found to have high internal consistency based on the assessment of the person separation index (0.92) for the Total score and Cronbach's alpha of 0.91.

In the three prevention trials, the reliability of the instrument was assessed using data from the run-in period (Day -7 - Day -1) prior to randomization (see Table 1). Reliability in the prevention trials was similar to that seen in the observational study. This gives some assurance that the reliability seen in the observational study, which used the 23-item instrument, would carry forward to a clinical trial setting that uses the 14-item instrument.

Table 1: Reliability measured during baseline period (Day -7 to -1) in the prevention trials

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

Source: Table 9, 12/2011 submission

4.2 Construct Validity

In the observational study, to assess construct validity, the Developer examined the correlation between the EXACT scores and selected patient-reported and clinical assessments of COPD severity including SGRQ-C, FEV1% predicted, MMRC score, rescue medication use and other clinical tests such as chest findings and both systemic and vital signs. The analyses were conducted in the pooled group of patients at a cross-sectional time point.

The findings in the prevention trials were similar to the observational study where the correlation of the Day 1 EXACT with the SGRQ (for the M-PEX trial) and the SGRQ-C was 0.62, 0.46, and 0.46, respectively.

Relatively large and statistically significant correlations were found between the SGRQ-C and EXACT Total scores on Day 1 ($r=0.64$, $p<0.0001$), where higher scores on both instruments indicate a worse health state.

There was very little correlation ($r=-0.07$) between EXACT Total scores and stable state FEV1% predicted. The Developer argued that this finding was expected, given both the historically low correlation between the individual attributes comprising the EXACT and FEV1 and the presence of Acute and Stable patients in the pooled sample.

The finding of little correlation between the Day 1 EXACT score and the FEV1% predicted was also found in the three prevention trials, where the correlation was -0.14, -0.10, and -0.13 for the MPEX, AZ 12, and AZ 20 trials, respectively.

In the Acute Group, there was no relationship between EXACT total score and stable state FEV_1% predicted: -0.08, which was not an unexpected finding. Similarly, in the Stable Group, the EXACT scores were not correlated with stable state FEV1% predicted: ($r=-0.07$).

4.3 Discriminant validity

In the observational study, the EXACT demonstrated discriminant validity by demonstrating a statistically significant difference ($F=36.7$, $p<0.0001$) in the Day 1 EXACT scores between groups classified by the clinician rating of exacerbation severity (stable, mild, moderate, and

severe) in an ANOVA analysis. A similar ANOVA with groups classified by a patient rating of exacerbation severity also demonstrated a statistically significant difference between groups ($F=45.0$, $p<0.001$).

Mean total (SD) score for the Stable Group was 36 (13); Mean Total (SD) score for Acute Patients rated as mild was 43 (9); moderate was 48 (9); and severe was 55 (9).

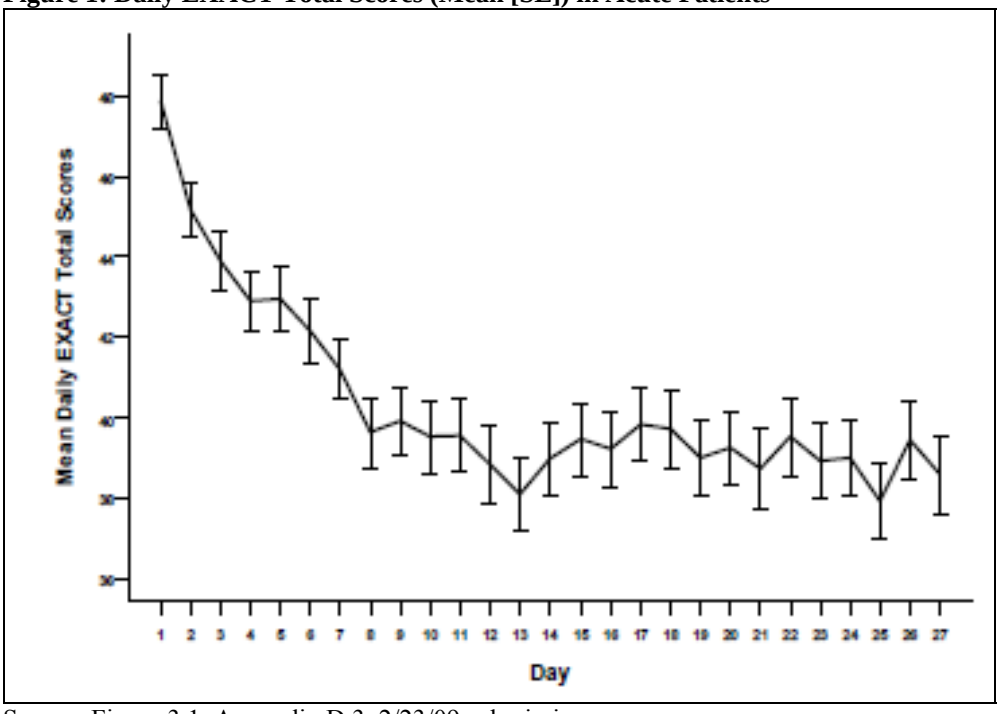
In addition, a repeated measures ANCOVA of the EXACT total score on Days 1-7 was able to discriminate between the Acute and Stable GROUPS ($F=49.9$, $p<0.0001$). The other predictors in the model were TIME, AGE, COMORBIDITIES, and STABLE FEV1. There was a significant decrease over time ($F=94.01$, $p<0.0001$).

4.4 Responsiveness

In the observational study, the Developer attempted to show that the instrument was responsive to change over time, as acute patients improved during the course of an exacerbation (see Figure 1). A repeated measures ANCOVA models for Days 1 to 10 and 1 to 27 showed statistically significant effects for time ($p<0.0001$).

In addition, for acute patients, SGRQ-C change scores from Day 1 to 29 were also used to identify responders and evaluate the EXACT's responsiveness to change. The Developer categorized subjects into three groups: Improved (Responders) (SGRQ-C change score ≤ -4; n=93), No Change (SGRQ-C change score between -3 and 0; n=30), and Deteriorated (SGRQ-C change score ≥ 1; n=48). Significant differences were found in EXACT total score ($p<0.001$) between the three groups. In this analysis, mean changes in EXACT total score were -12.0 (12.1) for Responders, -6.9 (12.3) for No Change, and -3.6 (9.6) for Deteriorated.

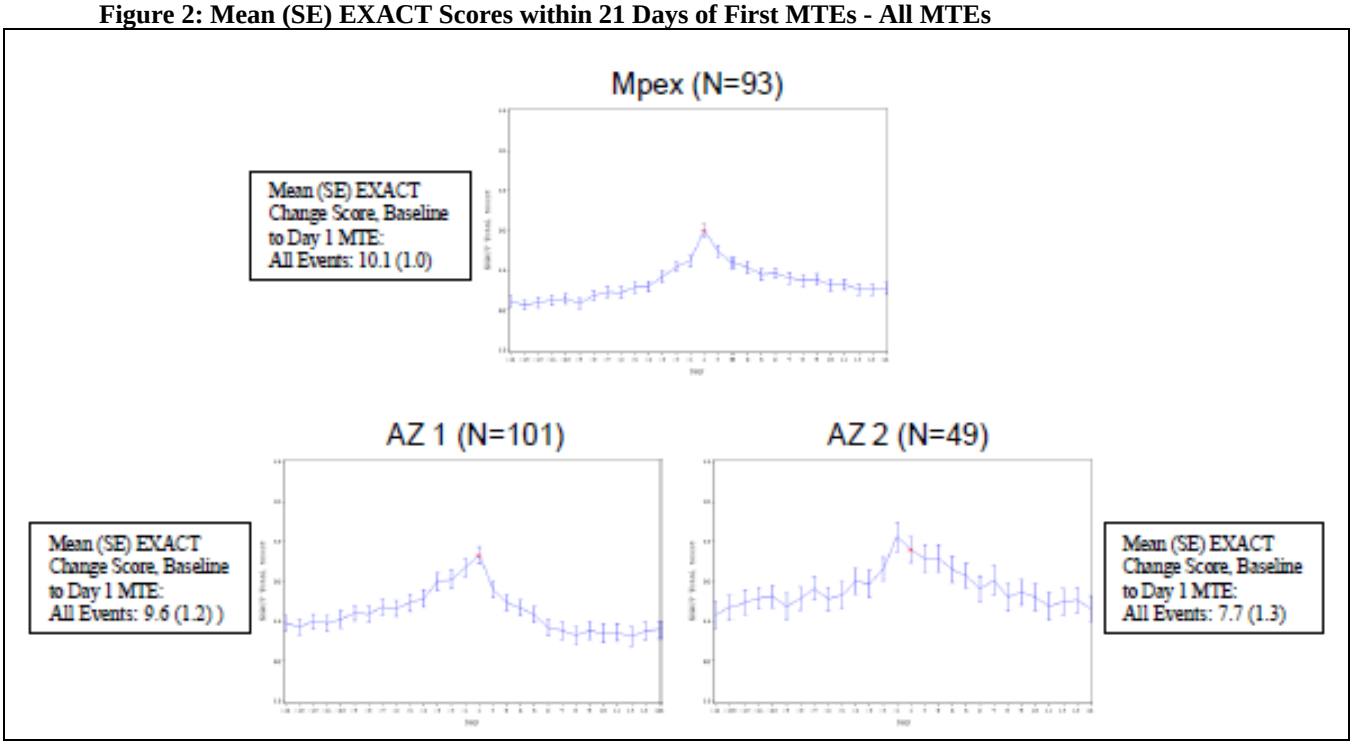
Figure 1: Daily EXACT Total Scores (Mean [SE]) in Acute Patients



Source: Figure 3.1, Appendix D.3, 2/23/09 submission

In the three prevention clinical trials using the 14-item instrument, the mean EXACT total scores increased sharply around the time of the MTE and subsequently decreased (Figure 2). The mean decrease from the peak to improvement in the exacerbation was similar to that seen in the observational study. This provides some assurance that the instrument is responsive.

Figure 2: Mean (SE) EXACT Scores within 21 Days of First MTEs - All MTEs



Source: Figure 1.1, Responses to Agency Queries (4/2013)

However, although changes in the mean EXACT total scores provide some assurance on the responsiveness of the instrument, many of the patients experiencing an exacerbation did not experience a marked EXACT score increase, i.e. beyond the normal day-to-day variability of five points seen in the observational study of stable patients (see Table 2).

In addition, while most patients experienced a modest increase in EXACT score from baseline, a substantial proportion of patients did not see an increase in EXACT score that met the Developer's definition of an EXACT-based exacerbation of ≥9 points.

Table 2: Number (%) of subjects exceeding EXACT Score Change Thresholds with First MTE (±1 day), among subjects with moderate and severe MTEs*

EXACT score change from baseline	MPEX n(%)		AZ 12 n(%)		AZ 20 n(%)	
	Moderate (N=71)	Severe (N=22)	Moderate (N=92)	Severe (N=9)	Moderate (N=47)	Severe (N=2)
≥9 points	30 (42.3%)	13 (59.1%)	38 (41.3%)	7 (77.8%)	19 (40.4%)	0 (0.0%)
≥5 points	47 (66.2%)	16 (72.7%)	54 (58.7%)	8 (88.9%)	28 (59.6%)	2 (100%)
≥2 points	59 (83.1%)	20 (90.9%)	67 (72.8%)	8 (88.9%)	34 (72.3%)	2 (100%)

* A moderate MTE is defined as requiring a clinic visit while a severe MTE is defined as requiring hospitalization
Source: Table 1.1, Response to Agency Queries (4/2013)

4.5 14-item vs. 23-item versions of the EXACT

The cross-sectional measurement properties were assessed primarily in the observational study of acute patients who were given the 23-item version of the EXACT rather than the final 14-item version. There is a question whether patients would provide similar responses for the fourteen items used to calculate the EXACT total score if only these items were asked, rather than all 23 items. However, because the prevention trials provided similar results as the observational study, there is less concern that patients would respond differently if they were only given the 14-item instrument rather than the 23-item, which provides evidence that one can extrapolate cross-sectional measurement properties from the 23-item version to the 14-item version.

4.6 Assay sensitivity

There is no evidence available at this time that the instrument can discriminate between and effective and ineffective therapy for patients who are experiencing an acute bacterial exacerbation.

4.7 Interpretation

There is no evidence available at this time on what constitutes a minimally clinically important difference, i.e. what amount of change would a patient or clinician consider clinically meaningful.

5. Potential deficiencies identified in the prevention trials

Although the prevention clinical trials provided confirmation of the psychometric properties seen in the observational study, there were several potential deficiencies identified in the trials. The first potential deficiency was the poor concordance between the clinical exacerbations as defined by healthcare utilization, specifically receipt of antibiotics and/or systemic corticosteroids and/or hospitalization for the exacerbation, and EXACT defined exacerbations in the three clinical trials. Another potential deficiency identified was the low compliance rate for completing the EXACT during the first several days of hospitalization due to an exacerbation.

The first potential deficiency identified in the prevention trials was that the instrument had low sensitivity to detect medically treated exacerbations (MTEs). The MTEs are categorized as moderate, which are defined as the receipt of antibacterial agents or corticosteroids to treat an acute exacerbation, or severe, which are defined as acute exacerbations that requires

hospitalization. Specifically, the sensitivity rates of the EXACT to detect an MTE are: Mpex: 43.8% (60/137); AZ 12: 70/132 (53.0%); and AZ 20: 26/59 (44.0%) (see Table 3).

The Developer defined EXACT-based exacerbations using a threshold of either

- An increase of greater than or equal to nine points from baseline for at least three days or
- An increase of greater than or equal to twelve points from baseline for at least two days

Table 3: Frequency of MTEs with Corresponding EXACT-based exacerbations within a + 7 Day Window of MTE Onset (Day 1)

	Mpex	AZ 12	AZ 20
Total number of MTEs	137	132	59
Number of MTEs with at least 1 corresponding EXACT-based exacerbation[1]	60 (43.8%)	70 (53.0%)	26 (44.1%)
Number of MTEs without a corresponding EXACT-based exacerbation	77 (56.2%)	62 (47.0%)	33 (55.9%)

[1]EXACT exacerbations corresponding to more than one MTE were counted more than once

Source Appendix E.2, Table 2 (December 2011 submission)

When looking at the stratified analyses, the sensitivity to detect an MTE was low in the Mpex study (moderate: 43.3%; severe: 42.4%). In the AZ trials, the sensitivity was low and similar to the Mpex study with the exception of patients with severe exacerbations in the AZ 12 trial where the sensitivity was 76.9%.

The Developer attempted to identify the characteristics of patients who had poor concordance between EXACT-based exacerbations and MTEs. Unfortunately they were unable to identify these characteristics.

For the proposed context of use in acute bacterial exacerbation treatment trials, the poor concordance between MTEs and EXACT-based exacerbations in the prevention trials, pose a potential problem because the increase in symptoms at the onset of the exacerbation may not be fully captured by the EXACT total score. Thus, because patients will be enrolled in the treatment trials based on clinical symptoms, a subgroup of patients may have an EXACT total score that is only slightly increased. This may limit the sensitivity of the instrument to distinguish an effective from an ineffective treatment.

Another potential deficiency identified in the prevention clinical trials was the low compliance in completing the EXACT observed during the first several days of hospitalization due to an exacerbation. Even though the proposed context of use is in the outpatient setting, some fraction of the patients may worsen after trial initiation and require hospitalization. Thus, there is a concern that a substantial portion of the data may be missing during the peak of the symptoms due to the exacerbation. In the Mpex trial, there was a substantial proportion of the patients who did not complete the instrument during the first few days after they were hospitalized for an exacerbation (Table 4).

Table 4: Compliance rate around time of MTE requiring hospitalization due to exacerbation in the Mpex trial (N=33)

Days from MTE	Compliance Rate
-3	92.9%
-2	88.4%
-1	90.5%
0 **(MTE)**	39.5%
1	34.9%
2	51.2%
3	60.5%

Source: Reviewer's table

6. Future Work

Future work on the EXACT should include the following: determination of a threshold for EXACT to be included as an inclusion criterion, specification of the endpoint for EXACT to be used in analyses, and determination of the clinically meaningful change.

The Agency's ABECB-COPD guidance provides inclusion and exclusion criteria for ABECB-COPD trials based on patient history along with patient signs and symptoms. In addition, the Agency guidance also states, "The same PRO instrument also should be used at baseline to define enrollment criteria." Future work should include a determination of a minimum inclusion threshold for the EXACT in ABECB-COPD trials. Note, the finding that many patients experiencing an exacerbation based on health care utilization did not have an increase in their EXACT score beyond normal day-to-day variability is concerning and could present problems in setting an inclusion threshold.

Future should also include a determination of the endpoint for EXACT to be used in analyses to test the efficacy of a new therapy. This endpoint could be either a comparison of EXACT scores at a fixed time point or time to improvement based on EXACT scores. The future work would then involve either a justification of the fixed time point to conduct the assessment or the specification of a minimum decrease in EXACT scores that represent clinical improvement. Note, the Agency ABECB-COPD guidance states, "A fixed time endpoint may not be as sensitive a measure of treatment effect as a time-to improvement analysis."

In addition, in order to interpret the results of analyses of the EXACT, further work is also needed to determine the magnitude of a clinically meaningful change for the EXACT.

7. Conclusions

For COPD patients who are experiencing an acute exacerbation at baseline, the EXACT has only been studied in a prospective observational study using an earlier version of the instrument, which contained 23 items. The observational study did provide evidence of the cross-sectional measurement properties of reliability and construct validity for the instrument. The similarity of the results from the observational study and the three prevention trials gives some confidence that that patients responded similarly to the 14- and 23-item versions of the instrument. There was some evidence of responsiveness provided in the observational study for acute patients wherein the mean EXACT scores decreased as the exacerbation improved. However, this finding was tempered by the results from the prevention trials where a significant proportion of the patients who were experiencing an

MTE did not have a marked increase in their EXACT scores beyond the normal day-to-day variability of five points seen in the observational study of stable patients.

A major limitation is that the EXACT has not been studied in a treatment trial of patients who are experiencing an acute exacerbation at baseline. As a result, questions remain on the responsiveness of the instrument and on its ability to discriminate between and effective and ineffective treatment. Furthermore, we will have difficulty interpreting the results of the instrument because currently we do not know what constitutes a clinically meaningful change. Also we do not have the information needed to size a treatment trial. In addition, the poor concordance between MTEs and EXACT-based exacerbations in the prevention trials is an additional concern. In the treatment setting, this poses a problem because patients will be enrolled based on clinical criteria not their EXACT scores. As a result, many subjects will have a baseline EXACT score that is similar to that seen when they are not experiencing an exacerbation and it will be more difficult to demonstrate a treatment effect.

Further work is needed to investigate the longitudinal measurement properties of the instrument outlined above. These properties should be studied in a Phase 2 trial prior to using the instrument as a primary or secondary endpoint in a Phase 3 trial.

Scott S. Komo -A

Digitally signed by Scott S. Komo -A
DN: c=US, o=U.S. Government, ou=HHS, ou=FDA, ou=People, cn=Scott S. Komo -A, 0.9.2342.19200300.100.1.1=1300149912
Date: 2014.02.18 11:45:18 -05'00'

Scott Komo, Dr.P.H.
Mathematical Statistician, DB IV

Lisa A. Kammerman -S

Digitally signed by Lisa A. Kammerman -S
DN: c=US, o=U.S. Government, ou=HHS, ou=FDA, ou=People, 0.9.2342.19200300.100.1.1=1300054827, cn=Lisa A. Kammerman -S
Date: 2014.02.18 11:52:15 -05'00'

Concur: Lisa Kammerman, Ph.D.
Mathematical Statistician, DB III