EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

10 April 2015
EMA/CHMP/SAWP/178465/2015
Product Development Scientific Support Department

Draft qualification opinion of qualification of exacerbations of chronic pulmonary disease tool (EXACT), and EXACT-respiratory symptoms measure (E-RS) for evaluating treatment outcomes in clinical trials in COPD

Draft agreed by Scientific Advice Working Party	5 February 2015
Adopted by CHMP for release for consultation	26 February 2015[1]
Start of public consultation	13 April 2015[2]
End of consultation (deadline for comments)	25 May 2015[3]

Comments should be provided using this template. The completed comments form should be sent to Qualification@ema.europa.eu

Keywords	Chronic obstructive pulmonary disease, clinical trial, COPD, endpoint, E-RS, exacerbation, EXACT PRO, patient-reported outcome, PRO, respiratory symptoms

[1] Last day of relevant Committee meeting.
[2] Date of publication on the EMA public website.
[3] Last day of the month concerned.

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

Introduction

The EXACT-PRO Initiative (EXAcerbations of Chronic Pulmonary Disease Tool – Patient-Reported Outcome) brought together clinical, research, methodology, and regulatory experts to develop a new patient-reported outcome (PRO) instrument to standardize the symptomatic assessment of exacerbations of COPD for evaluating frequency, severity, and duration of exacerbations in clinical trials of COPD ("EXACT", 14-items PRO). Furthermore, the EXACT-Respiratory Symptoms ("E-RS", 11-items PRO) was designed to address the need for a standardized PRO measure for evaluating the effect of treatment on the severity of respiratory symptoms in stable COPD. The respiratory symptom items comprising the E-RS were directly (1:1) taken from the EXACT. Hence, the E-RS can be understood as derivative instrument from the EXACT. The E-RS is self-administered by study participants as part of the EXACT daily diary (which is self-administered as well).

The initiative was conducted under the leadership of Evidera scientific staff and supported through funds provided by multiple pharmaceutical companies (www.exactproinitiative.com). The instruments are available for use with permission obtained through Evidera.

Background of development and intended context of use

EXACT (descriptions taken from EXACT User Manual, Vers 6.0, amended/shortened)
Background
Exacerbations are an important feature of chronic obstructive pulmonary disease (COPD), leading to significant morbidity and mortality. Reducing the frequency, severity, and duration of acute exacerbations is of great interest to patients, providers, and payers. These same parameters are often used as primary or key secondary endpoints in clinical trials, including pre- and post-marketing pharmaceutical trials evaluating the efficacy and safety of maintenance and acute therapies for COPD. Despite widespread commitment to understanding exacerbations of COPD and the effects of treatment, there has been no consensus on their empirical definition and no standardized approach to measurement. Historically, exacerbations have been defined in terms of health care utilization, *e.g.,* number of clinic visits, emergency room, or urgent care visits with oral steroid or antibiotic treatment, or hospitalizations for an exacerbation. Health care events have also been used as a proxy for exacerbation severity, with exacerbations requiring an unscheduled clinic or emergency room visit characterized as "moderate," and those requiring hospitalization as "severe." Various approaches have been used to quantify exacerbations that are unreported and self-treated at home, often characterized as "mild".

There are a number of limitations associated with the health care resource utilization (HCRU)-based definition of exacerbation. First, clinic contacts and visits are initiated by patients based on their assessment of the episode, relationship with the provider, cost coverage, and personal or family preferences for care. With as many as 50% to 70% of exacerbations unreported, this definition seriously underestimates exacerbation frequency. Second, HCRU definitions do not take into consideration, standardize, or control for the change or severity of patient symptoms or the physician's assessment of exacerbation. Third, HCRU, particularly hospital admissions, is related to health policy or coverage within a given country or region. Patients undergoing treatment in regions with relatively liberal hospital admission policies will have more frequent and more "serious" exacerbations, while those in regions with conservative admission policies will have less frequent and/or fewer "serious" episodes. These limitations have implications for prevalence estimates in epidemiologic studies, affect

estimates in studies examining the link between exacerbations and disease trajectory, and site selection and treatment outcomes in clinical trials.

A standardized symptom-based method of assessing exacerbations can address many of these limitations. This approach is often traced back to definitions proposed by Anthonison *et al. [1],* who used an empirical definition to identify and classify exacerbations in a clinical trial designed to test the benefits of antibiotic therapy. Seemungal *et al. [2]* extended this definition for the East London (UK) prospective cohort study, to understand causes and mechanisms of exacerbations of COPD. Since that time, diary cards have been used in a significant number of prospective clinical studies and trials to document symptom severity and identify unreported exacerbations. Although most cards include dyspnoea, cough, and sputum, the actual items used to capture these symptoms vary greatly, making comparison across studies virtually impossible and may account for some of the inconsistency in findings across otherwise similar investigations. Further, none of the cards were developed using well-known psychometric procedures with documentation consistent with United States (US) Food and Drug Administration (FDA) and CHMP guidelines. Standardizing the symptom assessment of COPD exacerbations through a common tool and metric is targeted to complement HCRU definitions and improve understanding of these important events, including the prodromal, acute, and recovery phases, and the effects of treatment.

Context of use

The EXACT was developed and validated for use in patients with COPD, including chronic bronchitis. COPD is characterized by persistent airflow limitation with varying degrees of air sac enlargement, airway inflammation, and lung tissue destruction. “The chronic airflow limitation characteristic of COPD is caused by a mixture of small airway disease (obstructive bronchiolitis) and parenchymal destruction (emphysema), the relative contributions of which vary from person to person. Emphysema, or destruction of the gas-exchanging surfaces of the lung (alveoli), is a pathological term that is often (but incorrectly) used clinically and describes only one of several structural abnormalities present in patients with COPD.” Chronic bronchitis, often the target of antimicrobial therapies for acute bacterial exacerbations of COPD (ABECB-COPD), involves persistent or repeated inflammation of the bronchi with excessive bronchial mucus and productive cough with sputum production on most days for 3 consecutive months in at least 2 consecutive years. Cough and sputum production may precede the development of airflow limitation; conversely, some patients develop significant airflow limitation without chronic cough and sputum production.

Exacerbations are events characterized by an acute, sustained worsening in the patient’s COPD beyond normal day-to-day variability, including an increase in respiratory symptoms such as dyspnoea, cough, and sputum production. The EXACT was designed to standardize the assessment of the patient’s condition in order to capture this dynamic process.

Patients with clinically relevant bronchiectasis are often excluded from exacerbation trials and are therefore excluded from the target population for trials using the EXACT. Although asthma is considered a disease of chronic airflow obstruction, the EXACT was not designed for use in this patient population. In addition, although the instrument may prove useful in patients with cystic fibrosis, alpha-1 antitrypsin deficiency, or obliterative bronchiolitis, these COPD phenotypes were not included in the instrument development process and are therefore not part of the target population for the instrument at this time.

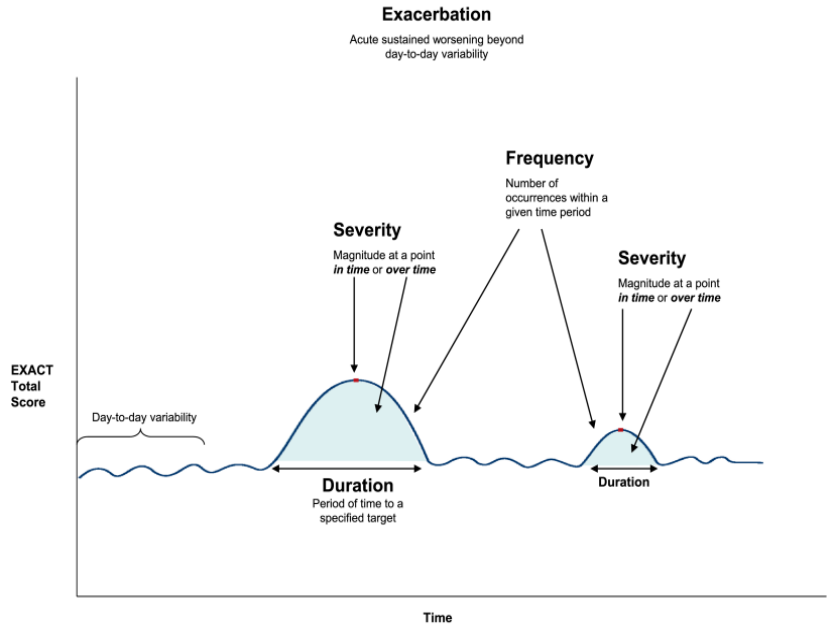
The EXACT was designed for use in 2 types of clinical trials:

1) Maintenance/prevention trials, testing the efficacy of therapies to modify or prevent COPD exacerbations (reduce their frequency, severity and/or duration). Historically, these trials have been 6 to 12 months in duration, enrolling participants during a stable state.

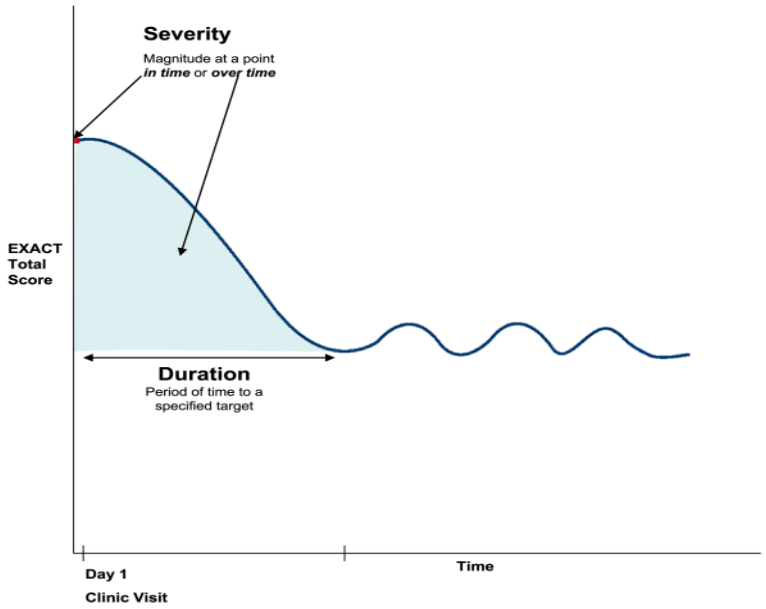
2) Acute treatment trials evaluating therapies to treat exacerbations of COPD (reduce their severity, duration, or recurrence). These trials enrol patients during an acute exacerbation of COPD, *e.g.*, anti-microbial drugs for ABECB-COPD.

Figures 1a and 1b show a schematic representation of exacerbations for these types of trials.

Figures 1a and b. Dimensions of Exacerbation Assessment by Trial Type



1a. Maintenance/prevention trials



1b. Acute-treatment trials

In maintenance/prevention trials exacerbation frequency, severity, and/or duration, may serve as primary, co-primary, secondary, or exploratory endpoints, as appropriate to the study design. In relation to treatment intervention trials, treatment (product-specific) target claims were suggested and discussed at the initiation of the EXACT-PRO Initiative to inform the instrument development process. The following claims were agreed upon and used as a reference point throughout the development and qualification review process, including Expert Panel Meetings (2006–2008), discussions with the FDA and in the EXACT-PRO qualification dossier:

- reduces the frequency of acute exacerbations of COPD
- reduces the duration of acute exacerbations of COPD
- mitigates/attenuates/reduces the severity of acute exacerbations of COPD

In the context of use during an acute exacerbation, the EXACT quantifies patient symptoms during COPD exacerbations treated in an outpatient setting (clinic and urgent care), from the day of diagnosis and enrolment into the trial through the designated follow-up period. The direction and magnitude of symptomatic change, improvement or worsening, can be determined and compared across treatment groups.

The following generic target claims for acute treatment trials were adopted at the initiation of the EXACT-PRO Initiative to inform the instrument development process:

- mitigates/attenuates/reduces the severity of exacerbations treated in clinic or emergency room (outpatient) settings
- reduces/speeds time to symptomatic improvement of exacerbations treated in clinic or emergency room (outpatient) settings

Table 1 in EXACT User Manual 7.0 summarizes the various uses of the EXACT to complement and extend the traditional HCRU definition of exacerbations.

Method of administration

The EXACT is a self-administered daily diary, completed by respondents each evening before bedtime. The instrument was developed as an eDiary (ePRO, PDA), but experience with pen-paper diary booklet administration is available as well.

E-RS (Descriptions taken from E-RS User Manual, Vers 2.0, amended/shortened)

Background

Chronic obstructive pulmonary disease (COPD) is a treatable but progressive disease, characterized by persistent airflow limitation with varying degrees of air sac enlargement, airway inflammation that is not fully reversible, and lung tissue destruction. The disease manifests itself in the cardinal respiratory symptoms of breathlessness, cough, and sputum production. Spirometry is essential for the diagnosis of COPD, provides information related to changes in airflow obstruction over time, and is useful for evaluating the efficacy of treatments intended to effect changes in airflow limitation in this patient population. Spirometry does not measure respiratory symptoms, however. In fact, studies have found that correlations between patient report of respiratory symptoms and forced expiratory volume in 1 second (FEV_1) are weak, and that patient perception of the impact of disease and their health-related quality of life are more closely related to these symptoms than is FEV_1. Clearly, respiratory symptoms are an important component of how patients with COPD feel and function.

Despite consensus on the defining respiratory symptoms characteristic of COPD, there is no validated method for evaluating their severity in clinical trials. Health status questionnaires administered periodically during the course of a trial include an assessment of respiratory symptoms and their impact, but do not capture this information on a daily or weekly basis. Several different daily diaries such as the breathlessness, cough, and sputum scale (BCSS) have been used in clinical trials and tested for reliability and validity. To date, no instrument to assess the respiratory symptoms of COPD has included the patient involvement in concept elicitation and item generation process necessary to provide evidence of content validity.

The E-RS was designed to address the need for a standardized PRO measure for evaluating the effect of treatment on the severity of respiratory symptoms in stable COPD.

Context of use

The E-RS was developed and validated for use in patients with COPD, including chronic bronchitis. COPD is characterized by persistent airflow limitation with varying degrees of air sac enlargement, airway inflammation, and lung tissue destruction. “The chronic airflow limitation characteristic of COPD is caused by a mixture of small airway disease (obstructive bronchiolitis) and parenchymal destruction (emphysema), the relative contributions of which vary from person to person. Emphysema, or destruction of the gas-exchanging surfaces of the lung (alveoli), is a pathological term that is often (but incorrectly) used clinically and describes only 1 of several structural abnormalities present in patients with COPD.” Chronic bronchitis involves persistent or repeated inflammation of the bronchi with excessive bronchial mucus and productive cough for 3 months or more in at least 2 consecutive years. Cough and sputum production may precede the development of airflow limitation; conversely, some patients develop significant airflow limitation without chronic cough and sputum production. The E-RS is intended for use in the following target population:

Clinical diagnosis of COPD or chronic bronchitis: min 40 years of age, current or former smoker with a history of at least 10 pack years, stable COPD, defined by exacerbation-free within 60 days of enrolment.

Although asthma is considered a disease of chronic airflow obstruction, the E-RS was not designed for use in this patient population nor those with clinically relevant bronchiectasis. In addition, although the instrument may prove useful in patients with cystic fibrosis, alpha-1 antitrypsin deficiency, or obliterative bronchiolitis, these COPD phenotypes were not included in the instrument development process and are therefore not part of the target population for the instrument at this time.

The E-RS is intended for use in clinical studies, including Phase II and III randomized, controlled trials testing the efficacy and safety of new treatments for patients with COPD. These trials are generally 12 weeks in duration, with the study length, number and nature of treatment arms, and specific outcome assessments and assessment intervals determined by the sponsor based on the target product profile, target claims, and related data requirements. Trials simultaneously examining exacerbation outcomes may last 6 to 12 months.

E-RS scores may serve as primary, co-primary, secondary, or exploratory endpoints in clinical trials designed to evaluate the effect of treatment on the severity of respiratory symptoms of COPD, as appropriate to the product and trial design.

The following target claims were discussed at the initiation of E-RS development and included in the E-RS evidence dossiers submitted to the FDA and European Medicines Agency (EMA) for instrument qualification:

- Treatment YY reduces the severity of respiratory symptoms of COPD
- Patients treated with YY reported significantly lower respiratory symptom severity scores than patients treated with XX following ZZ weeks of treatment

The 3 subscales embedded in the measure, RS-Breathlessness, RS-Cough & Sputum, and RS-Chest Symptoms, can be used as secondary or supportive endpoints to show the effect of treatment on these respiratory symptoms. In relation to these subscales the following claims were discussed at the initiation of E-RS development and included in the E-RS evidence dossiers submitted to the FDA and European Medicines Agency (EMA) for instrument qualification:

- Patients with COPD treated with YY reported significantly greater reduction in breathlessness severity following ZZ weeks of treatment.
- Patients with COPD treated with YY reported significantly greater reduction in cough and sputum severity following ZZ weeks of treatment.
- Patients with COPD treated with YY reported significantly greater reduction in chest symptom severity following ZZ weeks of treatment.

Method of administration

The E-RS is usually/always administered as part of the 14-item EXACT, which is a daily diary completed by respondents each evening before bedtime. The EXACT was developed following e-Diary administration technology, but experience with pen-paper diary booklet administration is available as well.

Methodological assessment of the EXACT and E-RS and scientific discussion

Qualitative development

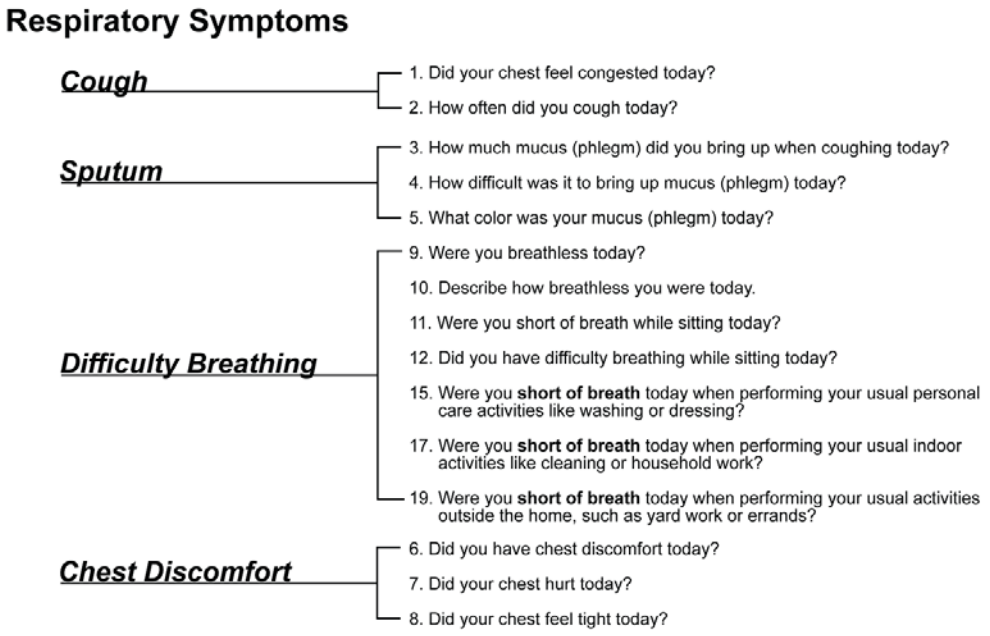
The qualitative development work for the EXACT was done in light of the goal to standardize the symptomatic assessment of exacerbations of COPD for evaluating frequency, severity, and duration of exacerbations in clinical trials of COPD [4]. Qualitative development work for the E-RS included data gathered during EXACT development and additional data on respiratory symptoms in stable COPD from a new set of subjects without recent exacerbation experience [5].

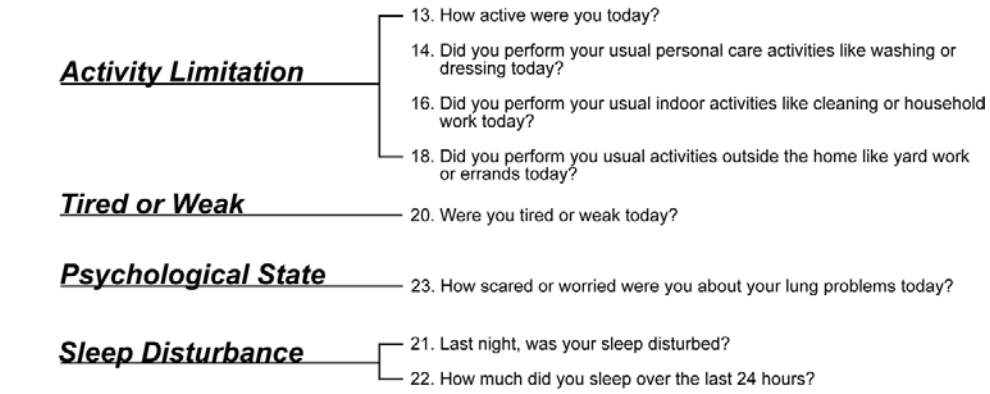
In the very first part of the project a comprehensive review of the existing literature on exacerbations in COPD was carried out, confirming the lack of a standardized symptom-based tool to assess duration, frequency and severity of exacerbations. The review was also important to identify and evaluate existing PRO instruments used in clinical trials of exacerbations of COPD. This informed the development of protocols and interview guides used in the qualitative research that formed the foundation of the tool. The first goal in development was then to determine the features and essential attributes of an exacerbation as perceived by patients to inform the instruments' content and structure. This was primarily done by targeted patient interviews and focus group sessions. Based on the outcome and the information retrieved, draft items were developed and further discussed within an expert panel. After further cognitive debriefing interviews with patients, an item pool of 23 questions emerged, which was taken as the basis for further quantitative development with item reduction.

From the methodological perspective, CHMP considers the measures and procedures taken in this early phase of development as adequate. CHMP also confirms that the resulting set of 23 items covers all topics/domains which are judged relevant by the EMA qualification team (QT) experts. See figure 3 for the initial conceptual framework to cover the relevant aspects concerning exacerbation in COPD.

As regards the particular wording of the item-questions, cognitive debriefings with patients were only conducted item-wise, and not in context of a (final) PRO questionnaire, which would have potentially also taken into account the patients' understanding of single items in relation to answers (already given) to other item-questions (in the same domain). This was identified as a deficiency by the EMA QT during the assessment of the qualification dossier. This was criticised in particular in relation to the fact that, in the final PRO tools, patients are not 'guided' through the questionnaire dependent on their answers given so far, but have to answer all items (no 'item skipping'), despite the fact that some might no longer seem applicable under certain circumstances. Consequently, this may lead to seemingly illogical answer profiles under certain conditions. The nature as well as the potential consequences of this methodological issue are further described and discussed in the next section.

Figure 1: Initial Conceptual Framework: 23-item instrument





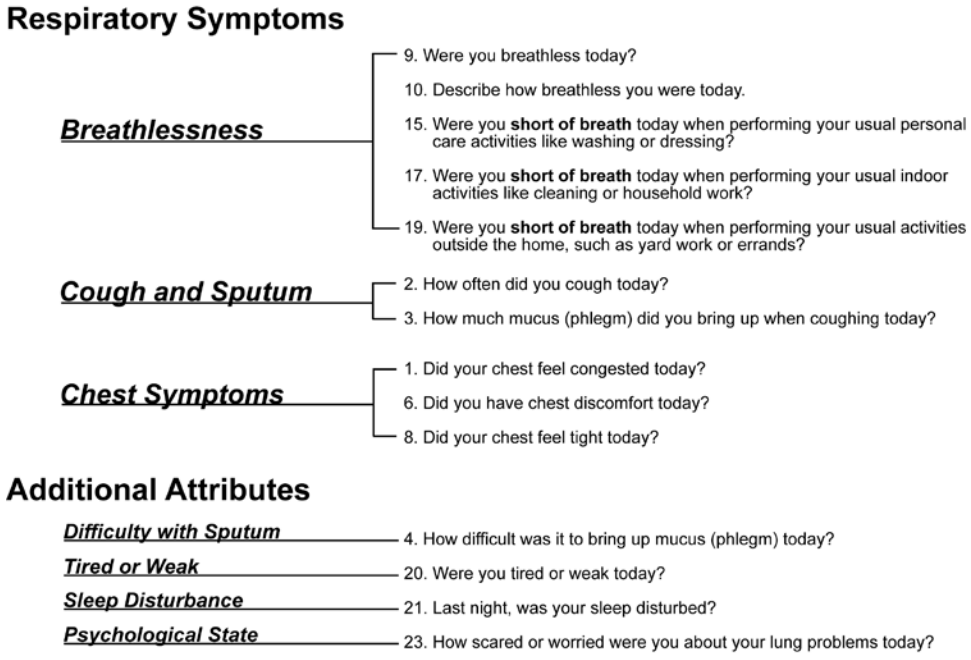
Quantitative development/validation

The next step of PRO development was item reduction and identification of domains in order to efficiently and exhaustively describe the concept of interest. For that purpose, in-depth quantitative analyses were carried out based on data coming from a two-group, prospective, observational study of 410 patients with COPD [6]. The patient population comprised 222 acute patients with a clinician-confirmed exacerbation and 188 clinically stable (non-exacerbating) patients, who all repeatedly completed the draft EXACT item pool (23 items) via personal digital assistant (PDA). In addition, patients and clinicians provided further relevant data, including clinical history, pulmonary function, St. George's Respiratory Questionnaire-COPD (SGRQ-C), Modified Medical Research Council (MMRC) assessment, physician assessment of patient's exacerbation manifestations (Acute Group); and patient and clinician global assessments of exacerbation severity (Acute Group).

State-of-the-art statistical/psychometric methodology was applied in the analyses of the resulting data set [7]. Rasch models (item response theory analyses) were used for item reduction and to identify distinct response categories per item. Subsequently, factor analyses were applied for item-structuring and domain definition. This resulted in a 14-item PRO tool (the EXACT), having a total score ranges from 0 to 100, where higher scores indicate a more severe condition. Factor analysis identified three factors (domains) embedded in the instrument: breathlessness, cough and sputum, and chest

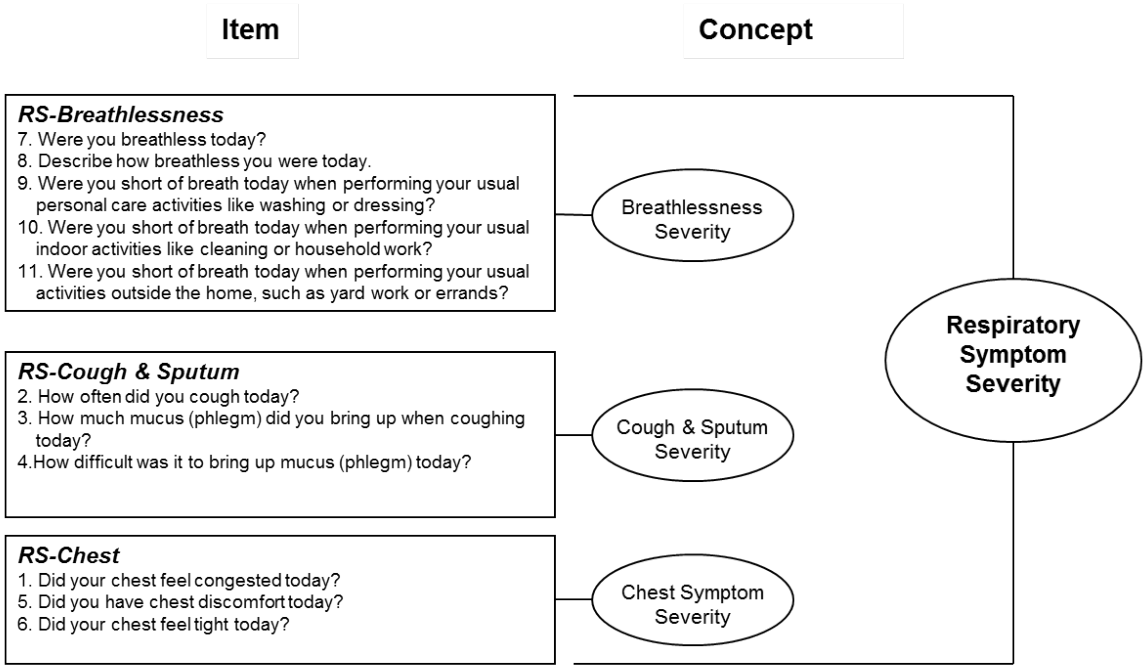
278 symptoms. Scores on these domains also range from 0 to 100 and provide information on these
279 specific attributes of exacerbation.

280 Resulting conceptual frameworks for the EXACT and the E-RS (which includes all items of the EXACT
281 related to respiratory symptoms.) are displayed in figures 4 and 5.Figure 4: Final EXACT conceptual
282 framework (showing all items with numbering according to draft item-pool)



283

284 Figure 5: Final E-RS conceptual framework



285

286 As regards the item selection process, the resulting domains and the structuring of items, the CHMP
287 has the following comments:

According to figure 1, the draft 23-items pool contained items related to patients' daily activity (limitations). In the final PROs, the "daily activity" domain was dropped. In the discussion with the analysts, they confirmed that this decision reflects the technical process of item selection together with discussion among the developers that the instrument should assess the symptoms associated with exacerbation events. In the reduced item sets (figures 4 and 5) activities of daily living are now only covered indirectly in 3 items of the breathlessness domain. The issue was discussed during the assessment of the qualification dossier, as (amount of) physical activity per se needs to be considered as one important domain with clear association to and influence on symptoms and other aspects covered with the reduced item-set. The EMA QT concluded that this issue needs to be seen in context of the future role of the EXACT/E-RS as an endpoint in clinical trials. As the EXACT and E-RS do not directly cover patients' physical activity, it might be necessary to cover this aspect by separate adequate tools in clinical trial setting to put (change of) EXACT/E-RS data in appropriate context, in order to better understand (the change of) a patient's disease condition (depending on the trials objectives).

One further issue identified in relation to item-categorisation was the fact that the symptom domain for cough and sputum in the EXACT and the E-RS do not comprise the same set of items, as the item: "How difficult was it to bring up mucus (phlegm) today?" is in this domain in the E-RS, but is a separated item in the EXACT. From the discussion with the developers of the PROs it was understood that this again was the result of the technical item analyses, and the resulting categorisations can be considered most efficient and optimal to describe the concepts of interests per PRO-tool. However, CHMP considers this divergence not optimal from a practical user's perspective, requiring additional explanation and description for user's who might be interested to make use of both PRO tools (including separated subdomain analyses) in parallel in one trial.

As already mentioned in relation to the assessment of qualitative development, the issue of an 'obvious' dependency between items did, according to the opinion of EMA QT experts, not receive sufficient attention in the development and validation of the PRO tools. Given the wording of the items and the corresponding response categories, a naïve approach of viewing the domain-specific subsets of items can in principle lead to the perception that two 'nested' item structures exist (see below), and that this dependency between items would actually call for a 'respondents-guiding' to (next) applicable items, dependant on answers given to an obvious superordinate item.

Nested item structures identified:

- 'How often did you cough today?' → 'How much mucus did you bring up when coughing?' → 'How difficult was it to bring up mucus today?'
- 'Were you breathless today?' → four items to specify breathlessness further.

As an example, it might not be considered logically consistent and straight forward to ask a patient the question of how much mucus he/she was able to bring up when coughing, if the superordinate item answer revealed that there was no coughing at all that day.

It is understood that neither the PDA device, nor the instructions in the pen & paper version would allow 'skipping' of items based on answers given to previous items. This issue was discussed with the developers in more detail and additional descriptive data analyses from the first validation trial (cross-tabulation of corresponding item responses) revealed and confirmed that seemingly inconsistent response profiles do/can result when administering the PROs to patients. However, logically inconsistent response profiles were seen in a relatively small number of observations. Furthermore, from the developers' perspective, the advantages of a 'multiple items' approach (over single item) in terms of better estimation of an underlying construct was illustrated in the framework of the

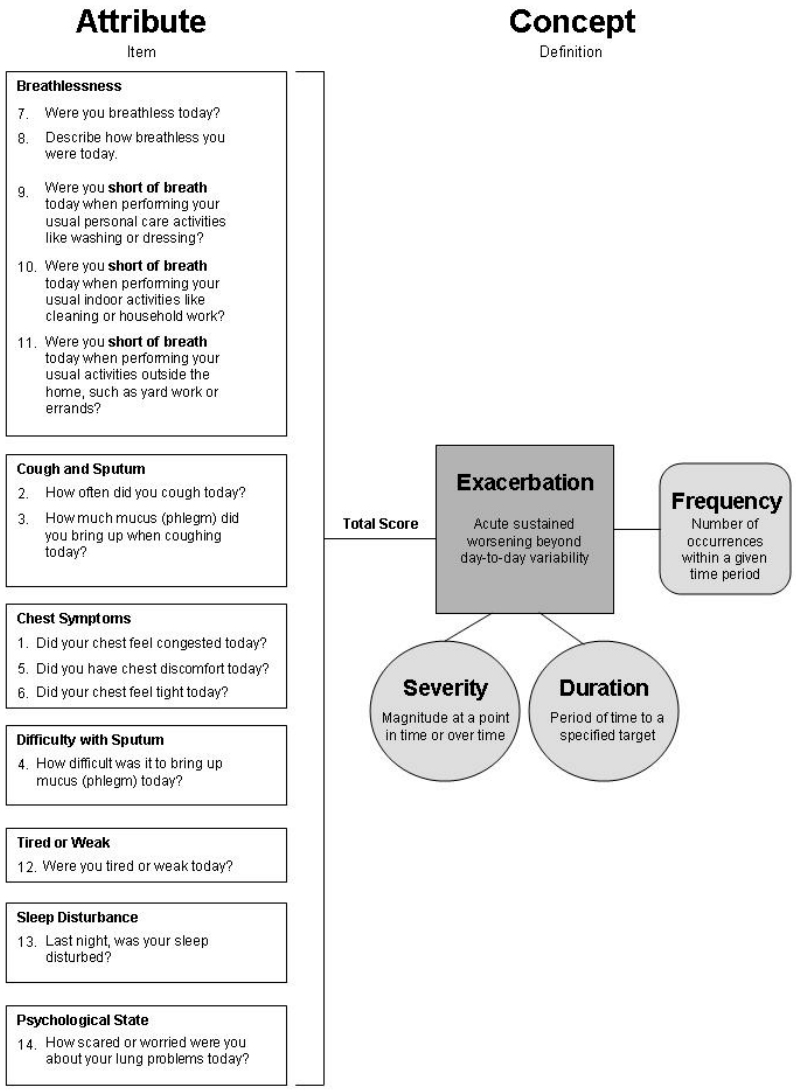
qualification procedure. In addition, developers reported that patients cognitively debriefed on the 23-items did not raise this as a concern and no signals of respondents' frustration or non-compliance (attributable to that issue) have been reported so far when using these tools in patient trials. Also, the final 14-item tool has been subjected to cognitive interviews during the translation process (over 20 languages (54 to date with at least 5 interviews per language), and no corresponding criticism was brought up from the patient side. This additional information was acknowledged by CHMP, alleviating the concern in relation to patient perception and face-validity of the PROs.

However, from a theoretical/methodological perspective, there remains a slight concern regarding interpretability of individual patient's EXACT total score changes, especially in cases where increases or decreases over time would be primarily driven by changes in answers to the mentioned items which would need to be interpreted as logically inconsistent (as explained, e.g. patient answers that even more mucus could be brought up when coughing as compared to earlier days, but still answers 'no coughing at all' to previous item). Change of that kind would also have a knock-on effect on the metrics used to describe intensity, frequency and duration of exacerbations events. Hence, in rare cases, the interpretation of (such) individual patients' development of the disease status will most likely be hampered. For statistical analyses of scores and exacerbation metrics on the group level (*e.g.* when comparing mean outcome between treatment arms), this methodological peculiarity of the PRO tools can indeed be expected as negligible, as it is considered very unlikely that systematic bias could be introduced which would favour one treatment condition (arm) in a clinical trial setting. This point of criticism is rather related to the content validity of the tool, as it might finally remain unclear in individual cases of inconsistent replies, what real facts regarding the disease condition would be underlying such response behaviour.

The evaluation of psychometric properties of the EXACT/E-RS included evaluation of internal consistency, test-re-test reliability, construct and discriminant validity, and responsiveness. CHMP considers this evaluation complete in the sense that all important properties of a newly developed PRO have been investigated. The advantage of having data from stable as well as from acute patients was utilised in these analyses. Detailed results of these evaluations are available in dedicated reports, and these are not subject to detailed assessment in this document. The consortium reports excellent internal consistency as well as excellent overall reproducibility, leading to the conclusion that the EXACT (E-RS) was found sufficiently reliable for the targeted context of use. In terms of validity, CHMP agrees that adequate content validity is given (see also assessment of qualitative development above). In terms of construct (external) validity, the EXACT showed pronounced correlation with SGRQ-C, MMRC and the amount of rescue medication, but weak or no correlation to FEV1% predicted. Analyses to investigate discriminant validity showed that using EXACT total score allowed to discriminate patients according to clinician rating of exacerbation severity (and hence also according the separation at inclusion: stable vs acute).

Investigations regarding responsiveness and magnitude of change (over time) are closely related to the project's primary goal to develop metrics for intensity, frequency and duration of exacerbations events based on observed patient trajectories of EXACT total scores over time. Taking this aspect into consideration, Figure 6 displays the full concept of the EXACT.

Figure 6: Final EXACT conceptual framework including the higher-level concept to derive algorithms and metrics for intensity, frequency and duration of exacerbations events (showing all items with numbering according to current version of EXACT)



A separate part of the analytical work was dedicated to the development of rules and algorithms to finally derive metrics for intensity, frequency and duration of exacerbations events. In this context, definitions have been set for: baseline (stable disease condition), onset of an event (start of acute worsening of condition), event duration, recovery and event severity, all based on sudden changes/stable phases in individual patients' EXACT total scores trajectories. In the framework of the qualification procedure, several methodological issues have been discussed in relation to these definitions. Among others, the question of whether onset or recovery of an exacerbation event can be triggered by worsening or improvement in one symptom domain only was addressed. Here, separate additional analyses revealed that majority of EXACT event onsets and recoveries would be triggered by pronounced changes in at least 2 symptom domains, according to the current metric definitions. For the sake of better understanding, the suggested/used rules/definitions for the EXACT are given below:

- baseline: within-patient mean over 7 days (4 minimum)
  - reset: every 4 exacerbation-free weeks to allow for improvement or deterioration
- onset: first day of worsening
  - $\geq$ 9 points for 3 days or $\geq$ 12 points for 2 days from baseline

- recovery: First day of persistent, sustained improvement
  - improvement: ≥ 9 point s from the maximum observed value Day 1-14
  - sustained: 7 consecutive days using a 3-day rolling average
- duration: days from Onset to Recovery
- severity: worst day of the event
  - In the time span between 'Onset' and 'Recovery', as defined above
- frequency: number of EXACT-defined events

In the related discussions with the Consortium it became clear that, as a matter of principle, the choice and settings of these definitions and algorithms determine the correspondence between the EXACT-data based exacerbation events and traditional HCRU-based definitions (*e.g.* medically treated exacerbations/events, 'MTE') of exacerbation. Further evaluation in this regard have been carried out based on data coming from further validation work/trials mentioned in the next paragraph. From a methodological perspective, the algorithms and settings chosen to define an EXACT-data based exacerbation event can be considered meaningful and acceptable on its own. Other choices and definitions might have been acceptable as well, leading to different correspondence (and hence comparative interpretation) to traditionally used HCRU-based definitions of exacerbation (*e.g.* MTE).

Further validation work for the EXACT and the E-RS was carried out based on data from three clinical trials where the 14-item EXACT was administered throughout the conduct of the individual trials and from which study raw data was fully accessible. In all these trials, the experimental drugs were found to be ineffective, allowing for an assessment of the performance of the EXACT and E-RS in moderate to severe COPD settings, involving patients on maintenance therapy. The outcome of additional performance evaluation based on these three trials is reported in detail in Leidy *et al.[8 and 9]* Primary focus is given to further evaluation of the correspondence between HCRU-based definitions of exacerbations (MTE) and the EXACT-defined events. In this context, the specific ability of the EXACT to record (otherwise) unreported events of worsening of the disease condition needs to be mentioned. Results discussed in this context reveal that, in general, EXACT-defined events are more frequent than MTEs, and that around 70-90% of EXACT events remain unreported. One further important finding is that – overall - only about half of the MTEs seen in the trials reached the threshold for an exact event, leading to an estimated sensitivity of around 50% for the EXACT event definition to 'detect' a MTE. It becomes evident from these figures that the different strategies evaluated to capture time phases of sustained worsening of disease condition measure rather different underlying concepts. Potential explanations of the differences observed are provided in the mentioned publication, and the authors' views and reasoning in this context are shared in principle by CHMP. Based on these findings and the limited extent of correspondence observed, the qualification of EXACT derived clinical endpoints is aggravated, as insufficient additional evidence currently exists for how differences in EXACT derived metrics for severity, duration and frequency of exacerbation events should be interpreted. Hence, the current lack of a common understanding of (minimum) clinical important differences in the evaluation of the EXACT derived metrics for severity, duration and frequency makes a qualification of the suggested endpoints as key efficacy measures (primary or secondary in late phase clinical trials) impossible at this point in time.

As additional validation evidence, and here in particular in relation to the PRO's ability to detect change, the EXACT User Manual (Versions 6 and 7) mentions the ATTAIN study, a 6-month phase III randomised, controlled trial which investigated the efficacy of aclidinium for the maintenance

treatment of COPD. This trial showed significant differences in (HRCU-defined) exacerbation rates between active and placebo group, which could also be reproduced by making use of the symptom-driven EXACT-based event definitions as described by Jones *et al. [10]*. However, only summarised results of this study were available to the EMA QT at the time of the review which limited the ability to explore the utility of the PRO in this setting.

So far, the EXACT has not been used in clinical trials evaluating the potential effect of experimental drugs on acute exacerbations.

An additional separate issue discussed in the framework of the qualification procedure was related to the notion that the patients' compliance to complete the EXACT in the hospital setting was rather low (~62%-72%) in the three validation trials. In that matter, it can be agreed to the Consortium that the EXACT was primarily developed to ask patients to rate their symptoms within the context of their home, rather than in a hospital setting, and as discussed above the strengths of the EXACT might indeed be to record episodes of symptom/condition worsening, which would otherwise not be reported following the usual trials standards without EXACT administration. However, it seems important to disentangle two issues in this context: the first being the applicability of the EXACT tool during hospitalisation in general, and the second being the reasons for/ consequences of reduced compliance during hospitalisation. At the moment, it appears that EXACT data coming from the home- and the hospital setting have different underlying quality. This will most likely further aggravate the interpretation of EXACT derived exacerbation metrics in clinical trials, where a noteworthy proportion of patients would be hospitalised.

Scientific questions discussed during the qualification procedure

First set of questions posed and discussed

Question 1

Does the EMA agree that the EXACT is acceptable as a method for measuring frequency, severity, and duration of exacerbations as efficacy endpoints in medical product development trials of chronic obstructive pulmonary disease (COPD)?

Question 2

Does the EMA agree that the EXACT-RS is acceptable as a method for measuring the severity of respiratory symptoms as an efficacy endpoint in medical product development trials of COPD?

SAWP response

Ad 1) The rather general wording of the questions leads to difficulties in decision making in relation to the sought qualification. The reason being that, as of today, frequency, severity and duration of exacerbations in COPD cannot readily be assessed in clinical trials in a standardised/validated manner, as methodological difficulties in that regard already arise in context of a universally accepted definition of an 'exacerbation' per se. The consortium themselves describe the whole spectrum of approaches to understand and detect phases of acute worsening in COPD disease conditions, ranging from HCRU-based to purely symptom-based strategies. Based on that, the question of whether a PRO has the ability to metrically characterise the medical condition of 'an exacerbation' is difficult to answer, as long as the nature of the targeted concept of an 'exacerbation' remains unspecific (as in the wording of the question originally posed). Hence, it was suggested to have a set of more specific questions as the basis for the qualification of the EXACT.

Ad 2) From CHMP perspective, E-RS (as compared to the EXACT) has only limited innovative elements to it as it can finally be used as COPD symptom score. As mentioned by the Consortium during the qualification procedure, the E-RS should be analysed and interpreted in a manner similar to other stable-state clinical measures like spirometry, SGRQ and TDI.

As a derivative of the EXACT - which had a different and innovative development objective behind it - the development of the E-RS appears more as a by-product of EXACT development rather than a 'stand-alone' development of a COPD symptoms PRO. E-RS can be interpreted as the symptom domain of the EXACT tool. Against this background it remains open whether the E-RS in its current form (11 items) would have resulted from qualitative and quantitative development as the optimal (=most valid, reliable and efficient) tool, if only the description of respiratory symptoms via a score would have been the primary focus of development. Despite this criticism, and the expected limited additional value of the E-RS in the presence of an available armamentarium of established tools to describe respiratory symptoms in COPD, the E-RS may finally qualify as an endpoint as proposed by the applicant. Some of the issues of lacking evidence concerning validation described for the EXACT also apply for this derivative tool at this point in time. So far, some important performance aspects could not be sufficiently explored. In particular, these are the PROs' ability to detect (treatment induced) change in stable as well as in acute disease conditions, and secondly the interpretability of observed differences in E-RS scores in the context of other accepted and frequently used relevant endpoints (definition/understanding of minimum relevant change, predictive validity). In parallel to the updating of the EXACT qualification questions (as mentioned above) the Consortium also decided to update the set of questions for the E-RS, see further below.

Second set of questions posed and discussed
For the EXACT
Question 1
Does the Agency agree that the EXACT measures symptoms of acute exacerbations of chronic obstructive pulmonary disease (AECOPD)?

SAWP response
In principle, the Agency agrees that the EXACT measures symptoms of acute exacerbations of chronic obstructive pulmonary disease. In close relation to the intended context of use, it is important to state that exacerbations need to be understood as events characterised by an acute, sustained worsening in the patients COPD disease condition, going beyond normal day-to-day variability. The conceptual framework of the EXACT comprises symptom domains which in total appear to cover all specific symptoms which are commonly judged relevant from a patient's and clinician's perspective. Hence, adequate content validity has been demonstrated, and also other performance measures indicate that the EXACT is a suitable PRO to measure symptoms as intended. One methodological issue has however been identified in this context, and this is related to the two item blocks for the domains of cough and breathlessness. As described in more detail in the scientific discussion above, the PRO does not foresee respondent's routing which would allow skipping of items which would seem not applicable given answers to superordinate item-questions. This may, in rare cases, result in logically inconsistent response profiles for individual patients, making single case interpretation of such profiles difficult in terms of understanding of the true symptom status. This issue is however considered of less relevance for any kind of data analyses on a group level.

Question 2
Does the evidence to date support its use as an exploratory endpoint in drug development trials for the prevention of exacerbations of COPD?

SAWP response

The Consortium applied state-of-the-art methodology during development and validation of the EXACT PRO tool. Some methodological issues have been identified in the course of the qualification assessment (see details in the scientific discussion above) which need to be taken into consideration when administering the EXACT in its current form. However, the totality of the evidence generated in the development and validation package supports the use of the EXACT PRO (including the related methodology to define metrics for severity, duration and frequency of exacerbation events) as an exploratory endpoint in drug development trials for the prevention of exacerbations in COPD. Not only the EXACT total score, but also the derived metrics for severity, duration and frequency of exacerbation events appear to be sufficiently sensitive to changes in an individual patient's disease condition. However, when administering/using the EXACT in the targeted context, the rather low extent of correspondence between the EXACT-based definition of exacerbations and other commonly used HCRU-based definitions (as discussed in the scientific discussion) has to be kept in mind and adequately reflected in the interpretation of study outcome.

Question 3

Does the evidence to date support its use as an exploratory endpoint in drug development trials of antimicrobial therapies for acute bacterial exacerbations of chronic bronchitis in patients with COPD (ABECB-COPD)?

SAWP response

The Consortium themselves indicate in the current version of the EXACT User's Manual that the performance of the tool has not been adequately investigated in the setting of acute exacerbations. CHMP has no objection to further exploration of the performance characteristics of the EXACT in this setting. The research field of anti-microbial therapies might be one option to further test the PRO tool, but CHMP sees no limitations for evaluating the tool also in other settings of acute COPD exacerbation.

Question 4

With further evidence, might the instrument be used as a primary or secondary endpoint to demonstrate effectiveness in drug development clinical trials of AECOPD?

SAWP response

In principle, CHMP confirms that the suggested attempt to characterise COPD exacerbation events in terms of severity, duration and frequency in a highly-standardised and more symptom-driven manner can be considered a valuable contribution to search for suitable efficacy endpoints in COPD trials. The primary open issue in relation to the question posed is whether the scientific community will be ready to move away from commonly used HCRU-based definitions due to the limitations described, and to accept symptom-driven definition (*e.g.* the EXACT methodology) to describe exacerbation events. The willingness to do so will depend on the degree of understanding which can be achieved in terms of putting outcome data of (changes in) the EXACT in good relation to other relevant (changes in) outcome measures commonly used in the past. One important aspect will be the judgement of the importance of unreported worsening events, which can be expected to be the majority of events detected by the EXACT in many instances (future clinical trials). However, sensitivity alone cannot be expected to be persuasive on its own. A clear context to clinical relevance would need to be established with this tool, and this is currently identified as the last important (and per se difficult) step for any future validation work.

As mentioned in answer to question 2, some methodological issues have been identified in relation to the technical makeup of the PRO, e.g. the peculiarity to theoretically reveal logically inconsistent response profiles in the domains of cough and breathlessness items. At this stage of the validation, it

remains difficult to judge in how far this property could aggravate the acceptability of the EXACT as a key endpoint in clinical trials in the future.

One final important aspect to mention in the context of whether the EXACT methodology would qualify for primary or secondary efficacy evaluation is the fact that patients' physical activity is not directly covered in the suggested PRO tool. However, amount of physical activity per se needs to be considered as one important domain with clear association to and influence on symptoms and other aspects covered with the EXACT. Therefore, for a more complete description of potential treatment success in clinical trials, it seem advisable to discuss the future role of EXACT for primary/secondary efficacy evaluation always in context of separate/parallel concepts to measure (amount of) physical activity.

For the E-RS

Question 5

Does the Agency agree that the E-RS measures respiratory symptoms of chronic obstructive pulmonary disease (COPD)?

SAWP response

CHMP agrees that the E-RS measures symptoms of chronic obstructive pulmonary disease. The development concept of the E-RS was to cover and exclusively contain the respiratory symptom domains which have been identified by the joint development work for EXACT and E-RS. According to this plan, 'item-wise' the E-RS is a direct derivative of the EXACT. Against this background, many of the comments made in answer to Question 1 in relation to the performance characteristics of the EXACT-PRO apply also to the E-RS. The presented conceptual framework of the E-RS comprises three symptom domains. Of note (as also mentioned in the scientific discussion above) the symptom domain for cough and sputum in the EXACT and the E-RS do not comprise the same set of items, as the item: "How difficult was it to bring up mucus (phlegm) today?" is in this domain in the E-RS, but is a separated item in the EXACT. CHMP considers this divergence not optimal from a practical user's perspective, requiring additional explanation and description for user's who might be interested to make use of both PRO tools (including separated subdomain analyses) in parallel in one trial.

The methodological issue related to the potential to trigger logically inconsistent response profiles is also of relevance for the use of the E-RS (see limitations and related concerns as described above).

Question 6

Does the evidence to date support its use as an exploratory endpoint in drug development trials evaluating the effect of treatment on respiratory symptoms of COPD?

SAWP response

The Consortium applied state-of-the-art methodology during development and validation of the E-RS PRO tool. Some methodological issues have been identified in the course of the qualification assessment (see details in the scientific discussion above) which need to be taken into consideration when administering the E-RS in its current form. However, the totality of the evidence generated in the development and validation package supports the use of the E-RS as an exploratory endpoint in drug development trials evaluating the effect of treatment on respiratory symptoms of COPD.

Question 7

With further evidence, might the instrument be used as a primary or secondary endpoint to demonstrate effectiveness in drug development clinical trials of COPD?

SAWP response

In this answer CHMP refers to demonstration of 'efficacy' rather than 'effectiveness', a term that is usually used differently in context of health technology assessments.

Despite the expected limited additional value of the E-RS in the presence of the available armamentarium of established tools to describe respiratory symptoms in COPD, the E-RS may finally qualify as an endpoint as proposed by the Applicant. Some of the issues of lacking evidence concerning validation described for the EXACT at the time of the review also apply for this direct derivative of the EXACT at this point in time. So far, some important performance aspects could not be sufficiently explored. In particular, these are the PROs' ability to detect (treatment induced) change in stable as well as in acute disease conditions, and secondly the interpretability of observed differences in E-RS scores in context of other accepted and frequently used relevant endpoints (definition/understanding of minimum relevant change, predictive validity).

As mentioned in answers to Questions 2 and 6, some methodological issues have been identified in relation to the technical makeup of the PRO, e.g. the peculiarity to theoretically reveal logically inconsistent response profiles in the domains of cough and breathlessness items. At this stage of the validation, it remains difficult to judge how far this property could impact on the acceptability of the E-RS as a key endpoint in clinical trials in the future.

CHMP qualification opinion

The EXACT PRO is a self-administered daily diary developed and validated for use in patients with COPD. It was designed to standardize the symptomatic assessment of exacerbations of COPD for evaluating frequency, severity, and duration of exacerbations in clinical trials. The EXACT PRO is intended for use in two types of trials; (i) trials testing the efficacy of therapies to modify or prevent COPD exacerbations, and (ii) trials evaluating therapies to treat acute exacerbations of COPD.

The CHMP concludes that the EXACT PRO currently can be used as an exploratory endpoint in drug development trials for the prevention of exacerbations in COPD. Not only the EXACT total score, but also the derived metrics for severity, duration and frequency of exacerbation events appear to be sufficiently sensitive to changes in an individual patient's disease condition.

In order to be used as a primary or secondary endpoint to demonstrate efficacy in drug development clinical trials of exacerbations in COPD, a clear context to clinical relevance would need to be established with EXACT PRO. There is a rather low extent of correspondence between the EXACT-based definition of exacerbations and other commonly used HCRU-based definitions. Furthermore, the clinical relevance of unreported worsening events, the expected majority of events detected by EXACT, needs to be established. Finally, as physical activity is not directly covered by EXACT, it seems advisable in future trials to use EXACT in parallel with measures of physical activity.

Further exploration of the performance characteristics of the EXACT in drug development trials of antimicrobial therapies for acute bacterial exacerbations of chronic bronchitis in patients with COPD (ABECB-COPD) would be of interest.

The E-RS is a derivative instrument from the EXACT designed to address the need for a standardized PRO measure for evaluating the effect of treatment on the severity of respiratory symptoms in stable COPD.

The CHMP concludes that the E-RS can be used as an exploratory endpoint in drug development trials evaluating the effect of treatment on respiratory symptoms of COPD. E-RS is expected to provide only limited additional value in the presence of available established tools to describe respiratory symptoms in COPD.

In order be used as a primary or secondary efficacy endpoint in drug development clinical trials of COPD, E-RS's ability to detect treatment induced change in stable as well as in acute disease conditions needs to be demonstrated. Furthermore, the interpretability of observed differences in E-RS scores in context of other accepted and frequently used relevant endpoints should be established (definition/understanding of minimum relevant change, predictive validity).

References

[1] Anthonisen NR, Manfreda J, Warren CPW, Hershfield ES, Harding GKM, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. Ann Int Med. 1987;106:196-204.

[2] Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. May 1998;157(5 Pt 1):1418-1422.

[3] European Medicines Agency, Respiratory Drafting Group. Guideline on clinical investigation of medicinal products in the treatment of chronic obstructive pulmonary disease (COPD). EMA/CHMP/483572/2012. London: European Medicines Agency. 2012; http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/08/WC500130880.pdf. Accessed January, 2015.

[4] Leidy NK, Wilcox TK, Jones PW, Murray L, Winnette R, Howard K, Petrillo J, Powers J, Sethi S; EXACT-PRO Study Group. Development of the EXAcerbations of Chronic Obstructive Pulmonary Disease Tool (EXACT): a patient-reported outcome (PRO) measure. Value Health. Dec 2010;13(8):965-975.

[5] Leidy NK, Sexton CC, Jones P, Notte SM, Monz BU, Nelsen L, Goldman M, Murray LT, Sethi S. Measuring respiratory symptoms in clinical trials of COPD: reliability and validity of a daily diary. Thorax. May 2014;69(5):424-430.

[6] Leidy NK, Wilcox TK, Jones PW, Roberts L, Powers JH, Sethi S; EXACT-PRO Study Group. Standardizing measurement of chronic obstructive pulmonary disease exacerbations. Reliability and validity of a patient-reported diary. Am J Respir Crit Care Med. Feb 2011;183(3):323-329.

[7] Jones PW, Chen WH, Wilcox TK, Sethi S, Leidy NK. Characterizing and quantifying the symptomatic features of COPD exacerbations. Chest. Jun 2011;139(6):1388-1394.

[8] Leidy NK, Murray LT, Jones P, Sethi S. Performance of the EXAcerbations of Chronic Pulmonary Disease Tool Patient-reported Outcome Measure in Three Clinical Trials of Chronic Obstructive Pulmonary Disease. Ann Am Thorac Soc. 2014 Mar;11(3):316-25.

[9] Leidy NK, Murray LT, Monz BU, Nelsen L, Goldman M, Jones PW, Dansie EJ, Sethi S. Measuring respiratory symptoms of COPD: performance of the EXACT-Respiratory Symptoms Tool (E-RS) in three clinical trials. Respir Res. Oct 2014;15(1):124.

[10] Jones PW, Lamarca R, Chuecos F, Singh D, Agustí A, Bateman ED, de Miquel G, Caracta C, Garcia Gil E. Characterisation and impact of reported and unreported exacerbations: results from ATTAIN. Eur Respir J. Nov 2014;44(5):1156-1165.

Table 1.0: Standardizing Exacerbation Outcomes in Clinical Studies of COPD (EXACT User Manual 7.0)

Endpoint[a]	**Definition**	**Measurement Approach**	
		Medically-Treated Events (MTEs)	**Symptom-Defined Events:**
Frequency **Event rate**	– Event rate: per person per year – Event definition: acute sustained symptomatic worsening of COPD; treated with antibiotics,	Number of health care resource utilization (HCRU) events: – Clinic or urgent care visit for an acute sustained symptomatic worsening of COPD, treated with antibiotics and/or steroids	Number of symptom-defined events: – Acute, sustained symptomatic worsening of COPD, defined as an increase in EXACT score ≥9 points for 3 days or ≥12 points for 2 days, above

	steroids, in hospital, or self-treated at home	– Hospitalization for an acute sustained symptomatic worsening of COPD EXACT score changes may be used to document change in symptoms associated with HCRU events.	Baseline Reported: accompanied by clinic or urgent care visit with antibiotic and/or steroid treatment or hospitalization Unreported[b]: no associated visit or hospitalization; self-treated at home
Time to first event **Time to subsequent (next) event**	– Days from initiation of treatment/placebo to first event – Days from recovery to subsequent (next) event	First HCRU Event: – Days to Day 1, clinic or urgent care visit – Days to Day 1, hospitalization Subsequent HCRU event: – Days from end of treatment for first HCRU event to Day 1 of next HCRU event	First symptom-defined event: – Days to Day 1 of sustained increase in EXACT score exceeding event threshold Subsequent symptom-defined event: – Days from Recovery from first symptom-defined event to Day 1 of next symptom-defined event
Proportion of patients with ≥ 1 event	– % patients with ≥1 event	% with ≥1 HCRU event: – % with ≥1 clinic or urgent care visit – % with ≥1 hospitalization	– % with ≥1 symptom-defined event: – % with ≥1 unreported symptom-defined event
Severity	– Degree or magnitude of the event(s)	Type of treatment: – Moderate: antibiotics or steroids – Severe: hospitalization Symptom severity: – Maximum EXACT score during the HCRU event – Change in EXACT score, baseline to HCRU Day 1 – Mean EXACT score during treatment; area under the curve (AUC)	Unreported, symptom-defined events: – Mild: self-treated at home[b] Symptom severity: – Maximum EXACT score during the event – Change in EXACT score, baseline to event Day 1 – Mean EXACT score during the event; AUC
Duration	– Length of the event(s)	Duration of treatment: – Days of treatment with antibiotics or steroids	Duration of symptoms: – Days from symptom onset to symptom recovery

		– Days of hospitalization	– Recovery: improvement in EXACT score ≥9 points from the maximum value, sustained for ≥7 days

[a]If 1 of these endpoints is chosen as the primary efficacy endpoint, the others also should be assessed to ensure that another exacerbation outcome has not worsened.

[b]Characterized as "mild" in EMA COPD Guideline.[EMA [3]]

Annexes

- Applicant submission – EXACT and E-RS – Updated User Manuals from cy version_2_0_3
- Applicant submission – EXACT_User_Manual_Version_6_0_20