5   Draft qualification opinion on Proactive in COPD

6
7

Draft agreed by Scientific Advice Working Party	26 October 2017
Adopted by CHMP for release for consultation	09 November 2017[1]
Start of public consultation	20 December 2017[2]
End of consultation (deadline for comments)	29 January 2018[3]

8
9

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

12
13

Keywords	Activity monitor, chronic obstructive pulmonary disease, clinical trial, COPD, endpoint, patient reported outcome, physical activity, PRO.

14

---

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

15 Background information based on the Applicant's submission

16 Under the Innovative Medicines Initiative Joint-Undertaking (IMI-JU) framework, the public-private
17 PROactive Consortium developed two Patient Reported Outcome (PRO) instruments to capture physical
18 activity (PA) data in patients with Chronic Obstructive Pulmonary Disease (COPD) in clinical trial
19 settings. One of those tools is the D-PPAC which is supposed to enable daily data collection (recall
20 period of 1 day). The other developed PRO tool is the C-PPAC with a recall period of 7 days, intended
21 to collect PA data during specified clinical study visits. The two PRO instruments have been developed
22 as 'hybrid' tools, i.e. classical questionnaire items are combined with activity monitor readouts
23 collected separately. The Consortium has produced electronic and paper-pencil versions of both the D-
24 PPAC and C-PPAC instruments. Also, translations to several languages have been done for both tools.
25 The English versions of the D-PPAC and the C-PPAC can be found in [Annex Link 1 and Annex Link 2].

26 During the development/validation phase, the Consortium sought advice from EMA in 2011 and in
27 2013 via the qualification advice procedure. These advice requests introduced the project, described
28 the proposed conceptual framework (CFW) and sought advice on elements of the Consortium's
29 approach to develop and validate the PRO instruments. In the framework of this (now third) interaction
30 with EMA, the Consortium presented validation work carried out in their project's last phase (work
31 package 6, WP6), which was based on 'final' versions of the two PRO instruments. Figure 1 below
32 illustrates the project's work flow and its structure consisting of three important work-packages (WP2,
33 WP4 and WP6). Details on these work-packages as well as corresponding assessment comments are
34 found in a later section of this document.

35 Figure 1: Overview of PROactive development stages
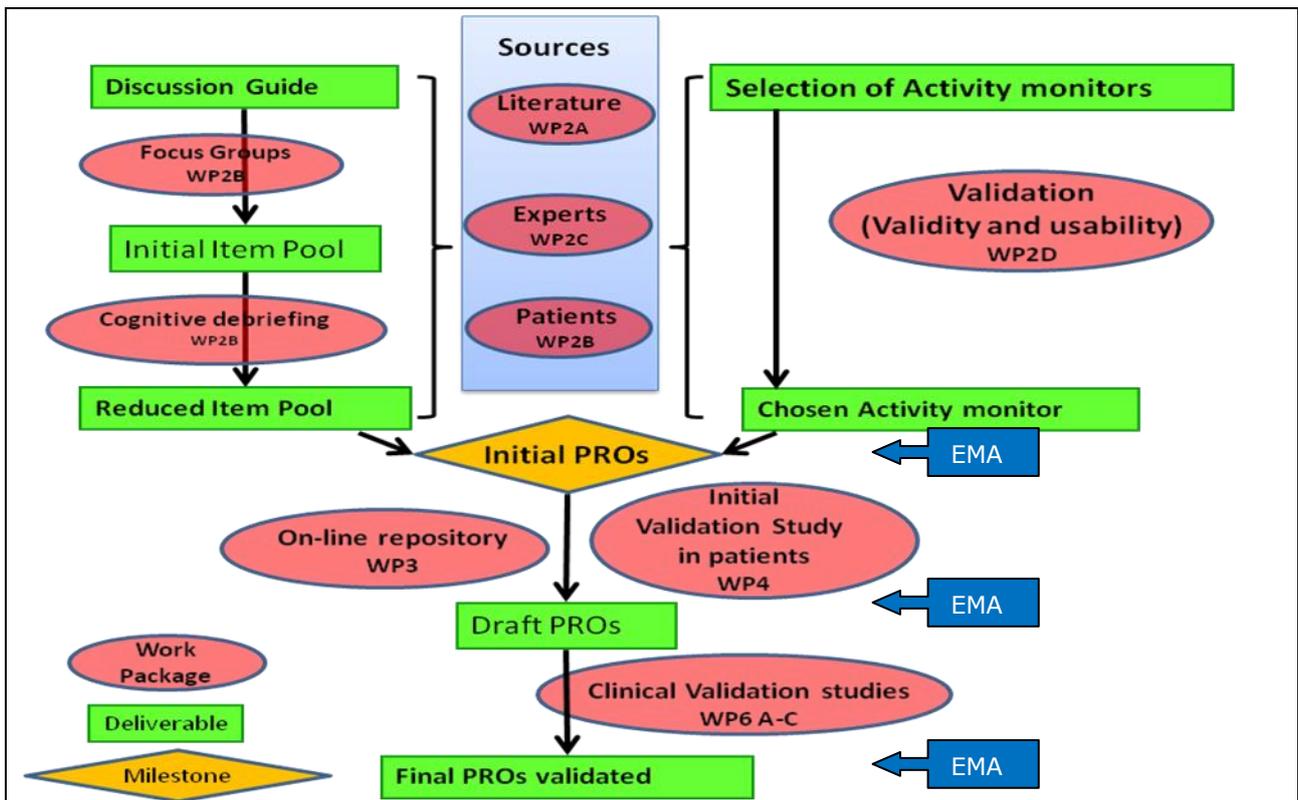
36

37 Based on the totality of validation work as presented, the Consortium suggests that the PRO tools are
38 ready for use in clinical trial settings having similar COPD patient populations as chosen in the
39 WP4/WP6 trials.

<u>The disease/condition in which the PPAC instruments are intended to be applied</u>

41 COPD, the 3rd leading cause of death worldwide, represents an important public health challenge that
42 is both preventable and treatable. COPD is a major cause of chronic morbidity and mortality
43 throughout the world; many people suffer from this disease for years, and die prematurely from it or
44 its complications. Globally, the COPD burden is projected to increase in coming decades because of
45 continued exposure to COPD risk factors and aging of the population [Annex Link 3 GOLD 2015].

46 Physical inactivity and its associated symptoms as a consequence of COPD are a hallmark of the
47 disease potentially contributing to the disease progression [Annex Link 4 Hopkinson and Polkey, 2010].
48 Patients are discouraged from being physically active due to the complex interplay of impaired exercise
49 tolerance, symptoms, exacerbations and co-morbidities (e.g. heart disease, osteoporosis,
50 musculoskeletal disorders, and malignancies) which may also contribute to restrictions of activity.
51 Impaired activity leads directly and indirectly to increased morbidity and even increases mortality in
52 COPD. The PA in which patients engage is the net result of the capacity patients have available to
53 engage in and their active choice to use the available capacity.

54 As a consequence, both disease impact, mainly determined by symptom burden and activity
55 limitations, and future risk of disease progression (e.g. exacerbations) should be considered when
56 managing patients with COPD [Annex Link 3 GOLD 2015].

57 Drug developers have traditionally used spirometry, laboratory parameters, exercise capacity, clinical
58 events (e.g. exacerbations) and/or health related quality of life as clinical trial outcome measures,
59 which do not fully cover the patients' experience of the consequences of the disease.

60 While it is important to measure changes in respiratory function and symptom endpoints when
61 evaluating new treatments in COPD, measuring their impact on aspects of daily life such as PA may be
62 more meaningful to patients and physicians/healthcare providers. There is now considerable evidence
63 that the level of FEV_1 is a poor descriptor of disease status [Annex Link 3 GOLD 2015].

64 Physical activity as defined by Caspersen (any bodily movement that results in an increase in energy
65 expenditure) can be measured with activity monitors [Annex Link 5 Caspersen et al. 1985]. However
66 these devices were, at the outset of the present project not well validated in COPD. More importantly
67 they provide only quantitative indices of PA and do not capture the patient's experience with PA. A
68 number of exercise capacity measures exist, e.g. Field Walking Tests [Annex Link 6 Holland et al.
69 2014] or Ergometry, which can inform researchers and developers about the patients' capacity for
70 exercise. However, engagement in PA is a different concept, as not only it calls on the patient's
71 physiological capacity, but also refers to a patient's self-efficacy and willingness to engage in activities.
72 The latter two are potentially influenced by a complex and individual interplay of exercise related
73 symptom perception, past behavior, health beliefs and motivation. Capturing all the dimensions of
74 daily PA that are relevant to patients should provide a unique perspective of treatment effectiveness.
75 However, despite its importance, no (other) existing PRO captures PA in a way that it maximally
76 reflects the experience of patients with COPD. Also, there is no PRO that is sensitive enough to
77 measure small but important changes in PA in clinical trials.

78 <u>Presentation of development, validation and regulatory assessment of the PROs</u>

79 Early development work forming the basis of both PRO instruments was carried out in the framework
80 of Work Package 2 (WP2). There were 4 sub-work-packages that contributed to the development:
81 systematic reviews of the literature (WP2A), patient input (WP2B), input from experts (WP2C) and the
82 validation and selection of activity monitors (WP2D).

83 Under WP2A five systematic reviews of the literature have been done. Figure 2 illustrates the different
84 objectives of these reviews.

85 Figure 2: Objectives of Systematic Scientific Reviews (SR) conducted as part of WP2A
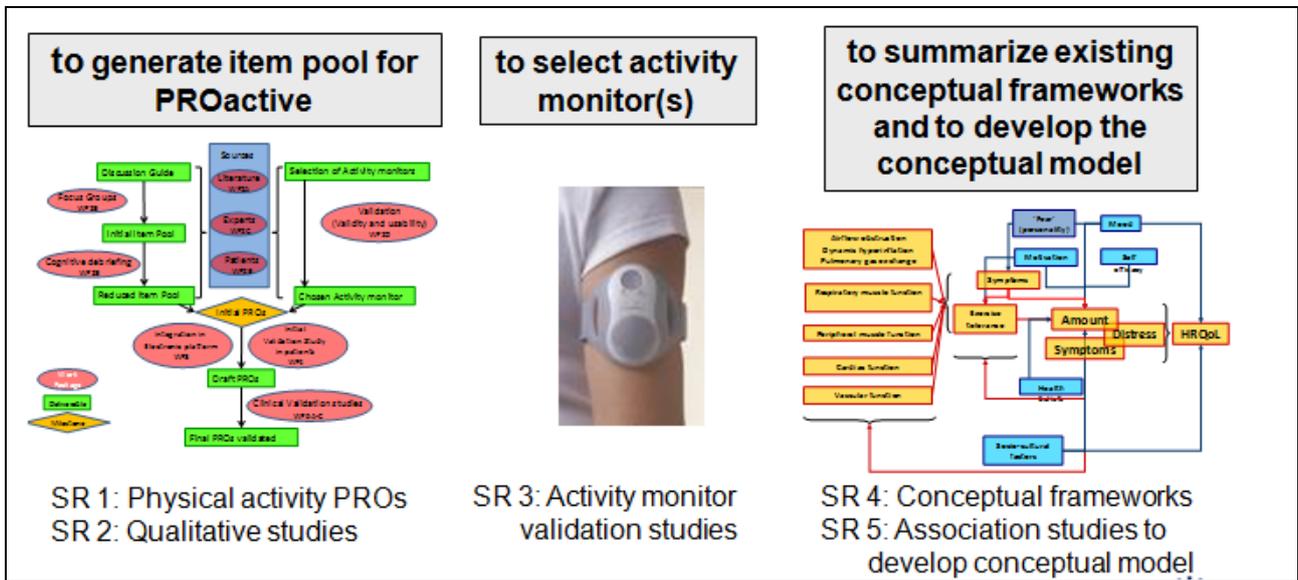

86

87 In summary, the literature reviews have helped to support the construct of the initial PROactive
88 conceptual framework and the drafting of the endpoint model, developed specifically for patients with
89 COPD, which is the intended population in which the PRO tools are supposed to be used. Reviews also
90 revealed that no valid instruments or scales existed at the time of development start which would
91 comprehensively capture PA from a COPD patient perspective. For more detailed descriptions of the
92 outcome of the WP2A-reviews the reader is referred to [Annex Link 7, Link 8, Link 9].

93 In parallel to WP2A, another work package WP2B covered qualitative research involving COPD patients.
94 This work package comprised one-to-one interviews, focus groups and cognitive debriefings which
95 were conducted in four European countries: the UK, the Netherlands, Belgium and Greece. Involved
96 COPD patients had different disease severity level. 116 patients participated in this qualitative
97 research. WP2B activities allowed identification of the draft concept of experience of PA (Figure 3).

98 Figure 3: Initial Draft of the Concept of experience of Physical Activity

99
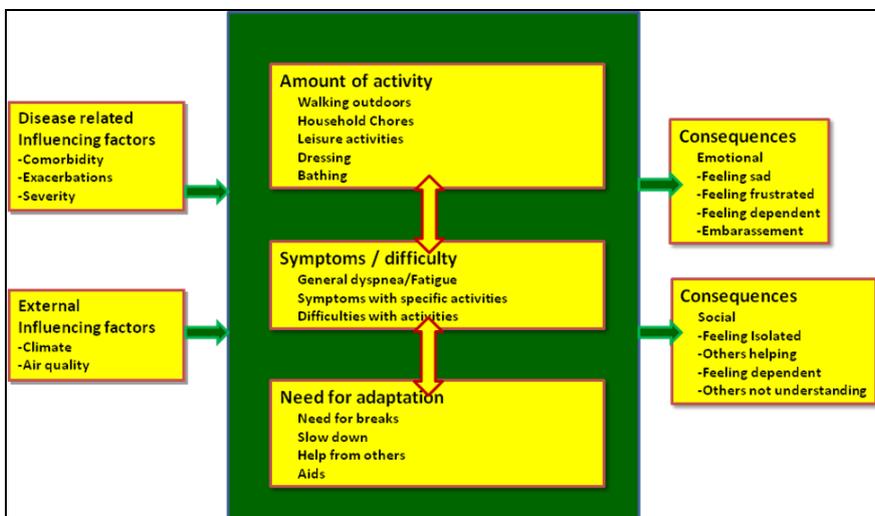
100

101 The qualitative studies also generated sufficient potential items shown to be of 'universal' importance
102 to patients. An initial item pool was derived and items thereof were tested in WP4 in conjunction with
103 the two selected activity monitors (see subsequent sections).

104 The work package WP2C assigned to expert input in the early stages of the instrument development
105 was primarily implemented to determine the criteria to characterize the general patient population or
106 give advice on the item pool. Here PRO developers used the complementarities of highly specialized
107 experts in their respective fields from 18 different organizations actively involved in the PROactive
108 consortium. In addition, as part of the advisory board, the PROactive consortium has met bi-annually
109 with a further set of 12 clinical and PROs experts as well as members from regulatory agencies that
110 provide guidance on the PRO development and validation. Furthermore, through the European
111 Respiratory Society, who was partner within the project, the consortium was also able to consult with
112 multidisciplinary experts at key stages of the PROs development to ensure that the construct meets
113 the clinicians' expectations.

114 Based on literature review, patient- and expert input, the initial conceptual framework (as shown in
115 Figure 4) was developed. Of note, items for the clinic visit PRO were similar to the items of the PRO to
116 be completed on a daily basis, with the exception of the items in bold, which only appeared in the clinic
117 visit PRO. This preliminary conceptual framework comprised 3 domains: 'Amount of PA', 'Symptoms
118 experienced during PA' and 'Need for adaptations'.

119 Figure 4: Initial Conceptual Framework
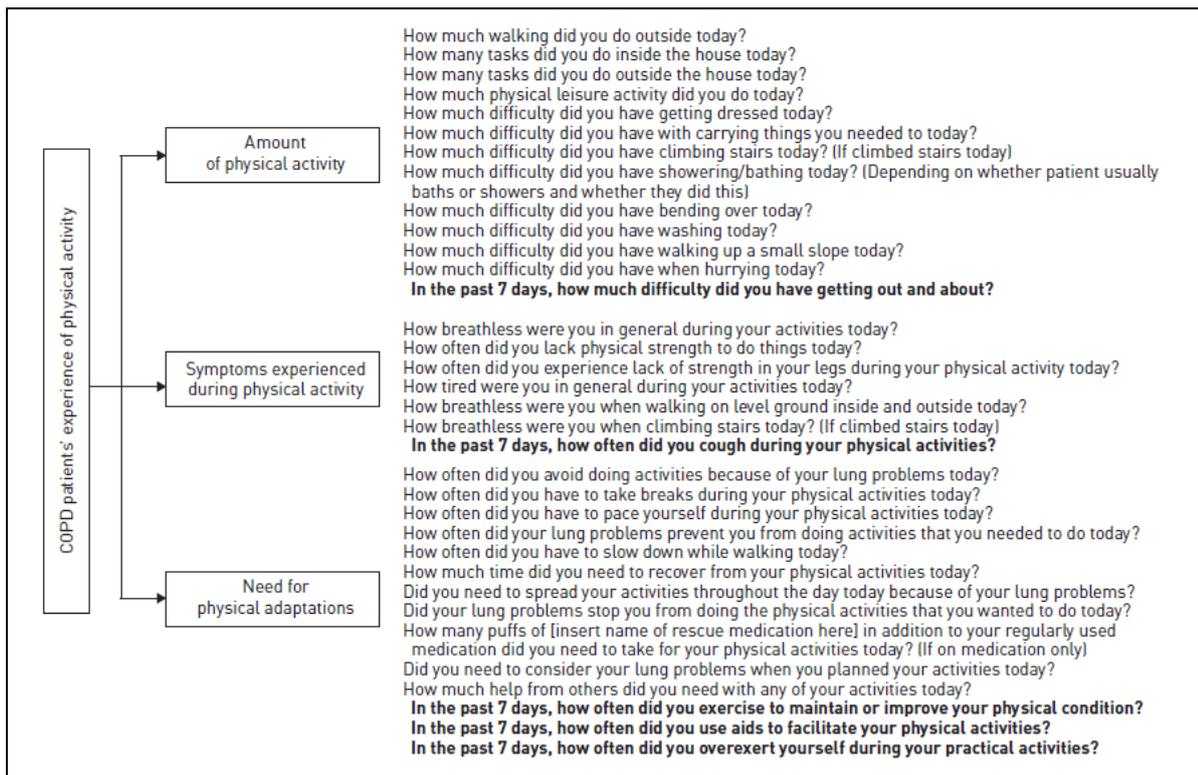
120

121 This initial conceptual framework was subject to discussion during the first interaction with the SAWP
122 qualification team (QT). In the course of assessing the first qualification advice request, the QT
123 challenged the assumption that a PRO tool based on the domains 'symptoms during PA', 'amount of PA'
124 and 'need for physical adaptations' will indeed be optimal to meet the Consortium's goal to have a
125 reliable and valid measure for PA in COPD patients. Especially the 'symptoms'-domain was felt to
126 contribute only little direct information about actual PA. At that time the Consortium explained the

127 findings from qualitative interviews with a variety of patients with COPD, namely that symptoms
128 patients experience during PA as well as adaptation required relate to the amount of PA they actually
129 do. Although the QT agreed that all these themes related to PA are closely interlinked, and that the
130 three proposed domains may be exhaustive to cover all relevant aspects to derive a PA score, it was
131 considered important that the wording of the items (especially for the symptom-domain) reflects the
132 link to (limited) PA. A pure domain on COPD symptoms without such a link was doubted to be
133 supportive for a new concept. Concern was expressed that the new PRO tools would conceptually be
134 very similar to already existing COPD questionnaires. In subsequent development and validation steps,
135 the Consortium considered that point of criticism. The result was eventually an altered conceptual
136 framework, not comprising a 'symptom'-domain anymore (see later sections).

137 One further aspect discussed with the Consortium at that stage of development was that improved PA
138 should generally not be at the expense of other aspects of QOL in COPD patients. It was recommended
139 by the QT that this issue required dedicated investigations during PRO validation. The Consortium
140 agreed and referred to their plans to also include measures of health status or health-related quality of
141 life in the PROactive studies planned to investigate this issue. Furthermore, it was mentioned that most
142 clinical studies in COPD include measures of health status or health-related quality of life, which would
143 allow for investigating such a potential impact in specific drug developments later on.

144 In relation to the Consortium's goal to adequately cover the theme 'amount of PA' with their PROs, the
145 idea of implementing read-outs from PA monitoring devices was introduced early during development.
146 Early plans to possibly develop the PROs as hybrid tools merging monitor readout data with item
147 response data were supported by the QT. PA monitors are frequently used to estimate levels of daily
148 PA. A variety of PA monitors are available to measure bodily movement. These devices use
149 piezoelectric accelerometers, which measure the body's acceleration, in one, two or three axes
150 (uniaxial, biaxial or triaxial activity monitors). Signals are transformed into various measures of energy
151 expenditure using specific algorithms, or are summarized as activity counts or vector magnitude units
152 (reflecting acceleration). With the information obtained in the vertical plane or through pattern
153 recognition, steps or walking time can also be derived by some monitors.

154 Reduced PA is an important feature of COPD. However, most of the monitors that were available at
155 project start had been validated in healthy subjects, but not necessarily in patients with chronic
156 diseases. As patients are less physically active and move slower than healthy subjects, the validity of
157 these monitors to pick up movement needed to be evaluated further.

158 With work-package WP2D, two studies were conducted to identify suitable activity monitors to be used
159 in validation studies as part of the PROactive instruments.

160 The first study, carried out in laboratory environment, followed the aim to evaluate the validity of six
161 monitors in COPD patients (ranging in severity from mild to very severe according to GOLD stages)
162 against a gold standard of indirect calorimetry in the form of VO2 data from a portable metabolic
163 system. It was hypothesized that triaxial activity monitors (transducing body's acceleration in three
164 axes) would be more valid tools when compared to uniaxial activity monitors. Indeed, the study found
165 that three triaxial activity monitors (Dynaport Move Monitor, Actigraph GT3X and SenseWear Armband)
166 were the best monitors to assess standardized and common physical activities in the range of intensity
167 relevant to patients with COPD. Changes in walking speed were most accurately registered by the
168 Dynaport Move Monitor and Actigraph, which are both devices that are worn on the hip. For further
169 details on the study see [Annex Link 10 Van Remoortel et al. 2012].

170 The second study in WP2D was carried out as a follow up to the previous study. It was supposed to
171 further assess the utility of activity monitors for use in clinical trials via a multicentre evaluation of the

172 six commercially available monitors ('field study'). All tested monitors showed good correlations with
173 'active energy expenditure'. The best correlations were obtained with two of the triaxial monitors
174 tested: the DynaPort MoveMonitor and the Actigraph GT3X. Another monitor, the 'Sensewear',
175 (BodyMedia Inc) also passed all preset validation criteria. However this monitor is branded as a
176 consumer device, rather than a medical device, and therefore was not further tested in subsequent
177 PROactive-related studies. The DynaPort MoveMonitor and Actigraph GT3X monitors were also the best
178 able to explain variability in total energy expenditure associated with PA, and were therefore most
179 representative of what patients were actually doing. For further details on the study see [Annex Link
180 11 Rabinovitch et al. 2013].

181 In summary, the data generated with these 2 studies, the laboratory validation study and the field
182 study, have supported the use of the DynaPort MoveMonitor and the Actigraph GT3X in subsequent
183 PROactive work packages WP4 and WP6 to further develop and validate the PROactive instruments.

184 Work package 4 (WP4) comprised an item reduction- and initial validation study with the primary
185 objectives to

186 - derive the set of items that measure PA in both the daily and clinic visit versions of the
187   PROactive instruments,

188 - confirm the draft PROactive conceptual framework of PA in patients with COPD for both the
189   daily and clinic visit versions of the PROactive instruments,

190 - perform an initial validation of the two PROs instruments

191 The design of this multicentre study was randomised 6-weeks observation 2-way cross-over (Fig. 5).

192 Figure 5: WP4 Study design
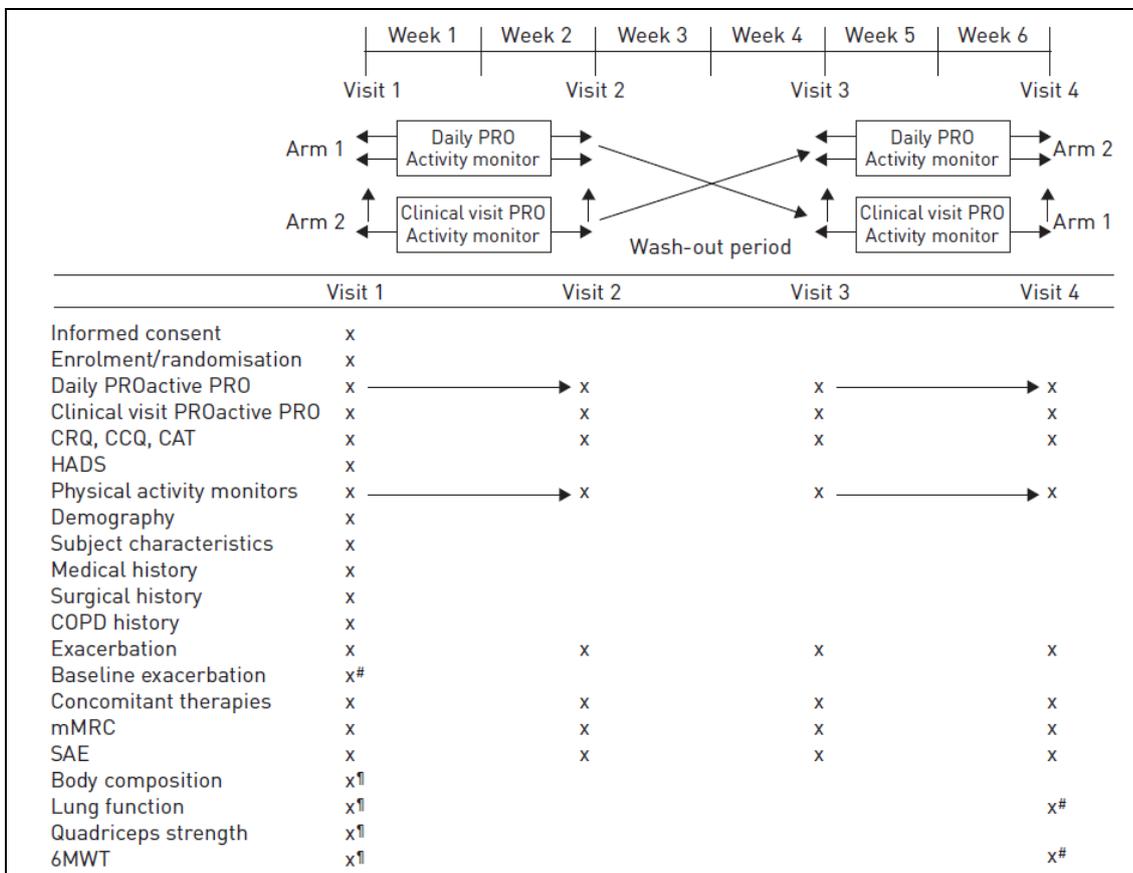
193

194 Both stable and exacerbated COPD patients were recruited, to cover the whole range of PA. In the first
195 2 week study period patients were randomised to complete either the daily PROactive item pool
196 consisting of 30 questions asking patients to report their PA experience on a daily basis, or the clinical
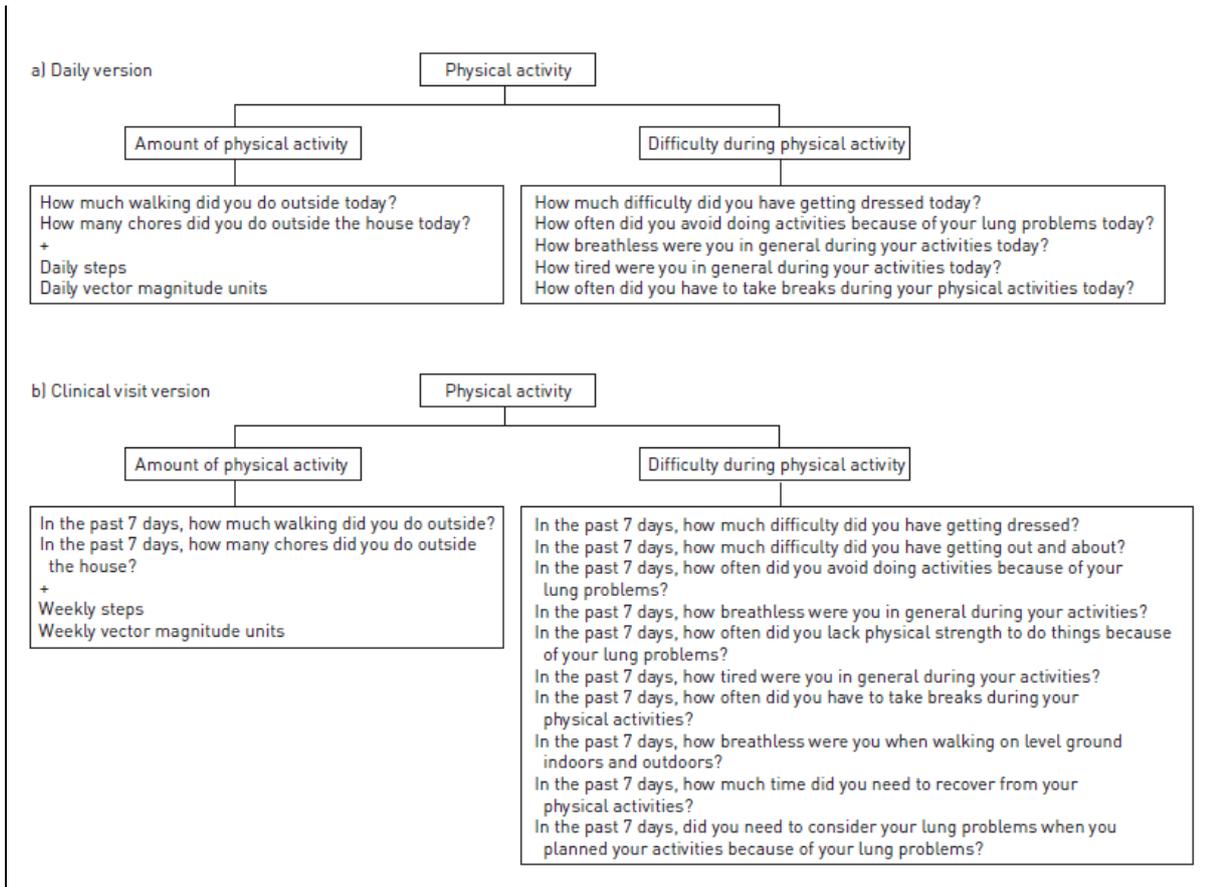197 visit PROactive item pool of 35 questions using a 7 day recall. Following a 2 week wash-out patients
198 completed the other questionnaire during the second study period. During the study periods, patients
199 had to wear two accelerometers: the Actigraph G3TX and the Dynaport MoveMonitor.

200 The design of the WP4 study was finalized following discussion with the QT which had some
201 reservations regarding the adequacy of a cross-over design. However the Consortium's view was that
202 the cross-over design allowed the use of a single large cohort, hence a broader range of COPD
203 phenotypes to be included when compared with a two armed study using matched groups. The full
204 cohort allowed for the evaluation of relationships between the two PROactive instruments of PA using
205 paired data. Thirdly, the design lead to a substantial reduction in the burden of phenotyping these
206 patients. The QT eventually agreed to the suggested design, also based on the review of draft versions
207 of study protocol and statistical analysis plan [Annex Link 12, Link 13].

208 Two hundred and thirty six (n=236) patients with COPD were included in the WP4 study. Patients were
209 mostly male (68%), with mean ±SD age of 67±8 years, FEV1 of 57±21% and body mass index of
210 27±5 kg·m^{-2}. Most of them were GOLD II or III, 9% were GOLD IV, 46% had co-morbidity, and 60%
211 had already been hospitalised for an exacerbation. A total of 228 patients (97%) had valid (≥3 days
212 with ≥10 h wearing time) data from activity monitors, showing good compliance and moderate levels
213 of PA.

214 For each of the two PROs two major methodological steps were carried out: domain identification was
215 done first by exploratory factor analysis methods, which was then followed by domain-wise item
216 reduction analyses (Rasch analyses). This sequential methodological approach actually carried out was
217 sufficiently described and CHMP could finally support the Consortium's interpretation of the WP4
218 analyses' results. The analyses carried out suggested that both the daily and clinical visit versions of
219 PPAC had a bi-dimensional structure, with a clear distribution of items in two factors. The two resulting
220 domains 'amount of PA' and 'difficulties during PA' had been reported to be quite robust. As compared
221 to the preliminarily conceptual framework (Figure 4), the revised conceptual framework (Figure 6) no
222 longer contains a symptom-specific domain, which indicates that the newly developed PROs have the
223 potential to cover specifically the (isolated) concept of PA as targeted.

224 Figure 6: Conceptual frameworks of a) the daily version of PROactive Physical Activity in COPD (D-
225 PPAC) and b) the clinical visit version of PROactive Physical Activity in COPD (C-PPAC) instruments:
226 final domains and items

a) Daily version

Physical activity

Amount of physical activity

How much walking did you do outside today?
How many chores did you do outside the house today?
+
Daily steps
Daily vector magnitude units

Difficulty during physical activity

How much difficulty did you have getting dressed today?
How often did you avoid doing activities because of your lung problems today?
How breathless were you in general during your activities today?
How tired were you in general during your activities today?
How often did you have to take breaks during your physical activities today?

b) Clinical visit version

Physical activity

Amount of physical activity

In the past 7 days, how much walking did you do outside?
In the past 7 days, how many chores did you do outside the house?
+
Weekly steps
Weekly vector magnitude units

Difficulty during physical activity

In the past 7 days, how much difficulty did you have getting dressed?
In the past 7 days, how much difficulty did you have getting out and about?
In the past 7 days, how often did you avoid doing activities because of your lung problems?
In the past 7 days, how breathless were you in general during your activities?
In the past 7 days, how often did you lack physical strength to do things because of your lung problems?
In the past 7 days, how tired were you in general during your activities?
In the past 7 days, how often did you have to take breaks during your physical activities?
In the past 7 days, how breathless were you when walking on level ground indoors and outdoors?
In the past 7 days, how much time did you need to recover from your physical activities?
In the past 7 days, did you need to consider your lung problems when you planned your activities because of your lung problems?

The resulting item sets as shown in the figure above were presented as 'draft PROs' after conduct of WP4. At the same time, the Consortium stated that no further changes in the PROs were foreseen at that point in time, and that all trials in WP6 were supposed to validate these very PRO versions. At that point in development the QT advised to maintain a certain amount of flexibility to amend/optimise the PROs (e.g. minor changes to response categories might turn out to be beneficial after broader use and testing). However, the Consortium stated that the items have been selected based on patient research and best statistical practice so should be robust going into WP6, where validation studies were planned to be running simultaneously, so timing of reporting would not permit adjustments of the PROs as part of WP6. For the QT, this fact constituted a minor deficiency in the PROs development and validation process. It was however understood that at least parts of the late phase validation trials would need to test and validate the final version of the PROs. As regards the intended implementation of monitor device data, the consortium considered different combinations of PRO question-items plus read-out variables from the activity monitors in the item reduction process. The two read-out variables 'daily steps' and 'mean Vector magnitude units per minute (VMU/min)' were found to be most informative in combination with the questionnaire items identified. Daily steps is understood to serve as a proxy for quantity of movement, whereas VMU serves as a proxy for overall intensity of effort. Cut-offs within the observed data ranges were chosen that maximised person separation index values in Rasch analyses. Interestingly, cut-offs differ between the two monitor devices investigated (Actigraph G3TX and the Dynaport MoveMonitor), which corresponds to a differential mapping from steps/day and VMU/min observed to PROs' response scores (0-4 or 0-5) finally assigned per monitor item included. Given that observation, it remained unclear for the QT in how far other monitoring devices than the two used in the validation trials could replace those monitors in the PROs without (repeated) thorough item-combination analyses including data cut-off investigations. The consequence is that the Opinion given with this document is formally restricted to the PRO use involving either Actigraph G3TX and the

252 Dynaport MoveMonitor. No recommendation is currently possible in relation to the use/implementation
253 of other monitor devices in the data capturing of the D-PPAC and the C-PPAC.

254 Overall, CHMP agreed that the information presented indicate that a combination of monitor device
255 read-outs and PRO items gives advantages in capturing amount of PA. Potential bias of wearing the
256 monitor device on the actual amount of PA was discussed with the Consortium, and evidence exists
257 that such bias might be negligible. The expectation that any potential bias of that kind would affect all
258 parallel intervention (treatment) groups in a clinical trial in the same manner was acknowledged.
259 Nonetheless, this general issue of biased estimation of PA might require dedicated consideration in the
260 interpretation of future trial results.

261 Based on WP4 study data, some psychometric properties of the two PRO tools had been investigated.

262 According to the reports provided, both instruments showed strong internal consistency and test–
263 retest reliability. Construct validity was explored via convergent-, known groups- and discriminant
264 validity investigations. In both PROs instruments, the domain 'amount of PA' exhibited weak
265 correlations with health-related quality of life and moderate correlations with dyspnoea and exercise
266 capacity. The domain 'difficulty with PA', however, showed moderate to strong correlation with health-
267 related quality of life, dyspnoea and exercise capacity. Known-groups validity was good in both
268 instruments, with scores differentiating across grades of dyspnoea, stable from exacerbated patients at
269 baseline and tertiles of PA levels (using variables not included in the PPAC scoring, such as intensity).
270 Analyses for discriminant validity revealed low correlations with unrelated constructs.

271 For further details of analyses results see [Annex Link 14].

272 Throughout the qualification advice procedures, the question of whether the PROs should reveal one
273 single total score each or, alternatively, separate scores for each of the two domains was repeatedly
274 discussed. Based on the (early) descriptions of the Consortium's motivation to develop PROs to
275 measure PA in COPD, the QT had a clear preference and advised to come up with one metric (per PRO)
276 to describe PA as one entity. For the Consortium it was important to note that, according to their
277 understanding, improving PA in COPD would either mean to improve the amount without negative
278 impact on difficulty, or to improve difficulty without negative impact on amount, or to improve both
279 amount and difficulty. With the advice provided, the QT saw no necessity to implement this 'restricted'
280 definition of improved PA already into the scoring system of the PRO tools. It was felt that observed
281 effects on a total score resulting from a mix of a slight negative change in one domain and substantial
282 improvement in the other might still be relevant from a clinical perspective.

283 Such an understanding would be in line with the interpretation of the outcome of many other
284 questionnaires (used in different disease areas) which feature more than one domain and one overall
285 sum score. It is quite common that domain sub-scores are planned to be reported and interpreted in
286 addition to allow for further exploration of the origin of observed effects. In the last round of discussion
287 between the Consortium and the QT, the Consortium confirmed their concept to suggest the use of a
288 total score (per PRO instrument), with the need to keep track of the two sub-domain scores. Both sub-
289 domain scores are mapped to a range from 0 to 100 points, and the total score is derived by taking the
290 arithmetic mean of the two domain scores (amount & difficulty), giving the two domains equal weights
291 in computation. According to the Consortium, additional ICC analyses revealed that alternative
292 weighting (60/40 or 70/30) would not improve psychometric properties, and hence equal weights were
293 considered suitable. For each of the two PROs, the total score is also defined on the range from 0-100
294 points. Finally, agreement was reached that an overall effect in (perception of) PA may be driven by
295 either or both domains, also reflecting the outcome of qualitative research with COPD patients.

296 With work package 6 (WP6) the PROs where further tested in clinical studies investigating the effect of
297 different pharmacological and non-pharmacological interventions in patients with stable moderate to
298 severe COPD, reflective of contemporary COPD management strategies [Annex Link 3 GOLD 2015].

299 With WP6, the Consortium was planning to address the following comments received in the final CHMP
300 advice from the two qualification advice procedures:

301 • Interpretation of PRO results on PA has to be seen in the context of the pharmaceutical class of
302   the drug used and the expected mechanism of action,
303 • Improved PA should not be at the expense of other aspects of Quality of Life (QOL) in COPD
304   patients,
305 • The instrument may not be optimal for patients with milder COPD;

306 WP6 was therefore designed to:

307 • Confirm the internal consistency of the two PRO instruments
308 • Confirm test-retest reliability
309 • Evaluate and confirm construct validity
310 • Evaluate and confirm known groups validity
311 • Investigate the ability to detect change over time, i.e. the PROs' responsiveness
312 • Investigate these changes in relevant subgroups of patients, e.g. age, gender, COPD severity
313 • Determine the definition of response and investigate the minimal clinically important difference
314   (MID)
315 • Verify the variables to use from the monitors and cut-offs from the activity monitors, and
316   confirm the monitor outcomes as part of the PRO instrument scores.
317 • reconfirm the conceptual framework established after WP4

318 In line with WP6 objectives, the consortium has longitudinally validated the PROs in six clinical studies
319 performed by EFPIA- and Academia partners. These studies are summarized below:

320 1. *PHYSACTO study*: An exploratory, 12 week, randomised, partially double-blinded, placebo-
321   controlled, parallel group trial to explore the effects of once daily treatments of orally inhaled
322   tiotropium + olodaterol fixed dose combination or tiotropium (both delivered by the Respimat®
323   inhaler), supervised exercise training and behaviour modification on exercise capacity and PA
324   in patients with COPD. The primary objective was to confirm that bronchodilator monotherapy
325   (tiotropium) plus behavioural modification, bronchodilator combination therapy (tiotropium +
326   olodaterol FDC) plus behavioural modification, and bronchodilator combination therapy
327   (tiotropium + olodaterol FDC) plus exercise training plus behavioural modification improve
328   exercise capacity as compared to placebo plus behavioural modification. The study population
329   consisted of outpatients with COPD of either sex, aged 40 - 75 years with a smoking history >
330   10 pack years, post-bronchodilator FEV1 $\geq$ 30% and < 80% predicted, and post-bronchodilator
331   FEV1/FVC < 70%.

332 2. *TRIGON - T9 study*: A Phase IIb, double blind, randomised, multinational, multi-centre, 2-way
333   crossover, placebo controlled study designed to demonstrate the superiority of CHF 5259 (i.e.
334   glycopyrronium bromide) vs. placebo, administered by pMDI over a 4-week treatment period in
335   patients with moderate to very severe COPD (GOLD stage III and IV). Primary Outcome
336   Measure was the change from baseline in pre-dose morning FEV1 on Day 28. Male and female
337   adults (40 $\leq$ age $\leq$ 80 years) with a diagnosis of COPD being current or ex-smokers with a post-
338   bronchodilator FEV1 < 60% of the predicted normal and a post-bronchodilator FEV1/FVC < 0.7
339   were included.

3. *URBAN TRAINING (CREAL) study*: This cross sectional and longitudinal RCT has – on top of validating the PROactive instrument - also provided opportunity to test an innovative intervention in patients with COPD. This study involved a training intervention adapted to each patient needs and capabilities and using public spaces and urban walkable trails. Primary objective was to assess 12 months effectiveness of the intervention with respect to PA level (primary outcome), and COPD admissions, exercise capacity, body composition, quality of life, and mental health (secondary outcomes) compared to "usual care". COPD patients aged >45 years with a ratio of forced expiratory volume in one second (FEV1) to forced vital capacity (FVC) $\leq$ 0.70 and clinically stable (i.e. least 4 weeks without antibiotics or oral corticosteroids) were included.

4. *ExOS study*: A cross-sectional and longitudinal open labeled 3 arm study was performed to primarily assess the functional capacity in patients with COPD and secure a wider understanding of the stability and sensitivity of commonly employed exercise tests so as to guide clinical trial outcome selection. This 7-9 week study compared the outcomes of the exercise tests following an (known) effective intervention, of either pulmonary rehabilitation or an inhaled bronchodilator (LAMA) therapy for 6 weeks. There was also a control arm with no intervention. Secondary objectives were to explore the relationship between PA and exercise testing and their responses to pulmonary rehabilitation and LAMA, and to report the MID of studied tests in response to pulmonary rehabilitation and LAMA. COPD patients with a GOLD stage 2-4 and MRC grade dyspnea 2 or greater, aged 40-85 years were included.

5. *MrPAPP study*: A cross sectional and longitudinal randomised clinical trial assessing the impact of a telecoaching program (COACH) on PA in patients with COPD on top of usual care, compared to usual care alone for 3 consecutive months. The COACH program included a step counter, an exercise booklet, an application installed on a Smartphone, the use of text messages and occasional telephone contacts with the investigator. PA was measured using the PROactive monitors (ActiGraph® and DynaPort®) and the PROactive questionnaire. A daily goal (number of steps) was sent to the patient, and revised every week. Patients were 66 years old on average, with an FEV1=56±21% predicted, and 1/3 were female.

6. *ATHENS study*: Longitudinal randomised 4-arm study intended to compare paper-pencil versus the electronic scoring version of the PROactive instruments. All the patients who participated in the rehabilitation program were randomised in four groups: Group A included patients who only used the paper-pencil version of the clinical visit version of the PROactive instrument; in Group B patients used the electronic version of the clinical visit version of PROactive instrument; Groups C and D were used as control groups including patients who did not participate in a rehabilitation program while receiving the usual standard of care. Groups C and D were also randomized to those patients using the paper-pencil version (Group C) or the electronic version (Group D) of the PROactive instrument. The rehabilitation programme was multidisciplinary including mandatory supervised aerobic training 3 days a week, at appropriate training intensity, which was to be increased on a weekly basis. Resistance training was performed with fitness equipment also for 3 days/week. Other components of the program were breathing control and relaxation techniques, methods of clearance of pulmonary secretions, disease education, dietary advice, and psychological support on issues relating to chronic disability. Clinically stable patients with COPD were to be recruited from the academic centers' Outpatient Clinic on the following entry criteria if they had a post-bronchodilator FEV1 lower or equal to 70% predicted without significant reversibility (<12% change of the initial FEV1 value or <200 ml) and optimal medical therapy according to GOLD stage 2.

386   In the trials of WP6 D-PPAC and C-PPAC were implemented for use according to the
387   descriptions as presented in Table 1.

388   Table 1: PPAC capture in WP6 individual trials

	PHYSACTO (BI)	URBAN TRAINING (CREAL)	T9 TRIGON (Chiesi)	ExOS (UK NHS Trust)	Pulmonary Rehabilitation (ATHENS)	MrPAPP (Academic-TT)
CT number	NCT02085161	NCT01897298	NCT02189577	-	NCT02437994	NCT02158065
N (included in analysis)	283	308	161	33 (Pilot)	59	361
Activity Monitor(s)	Dynaport	Dynaport	Dynaport	SenseWear & ActiGraph	Actigraph	Dynaport & Actigraph
Overall duration of study	19 weeks	12 months	12 weeks	7-9 weeks	8 weeks	3 months
PROactive	Key 2nd endpoint	Exploratory endpoint	Exploratory endpoint	Co-Primary endpoint	Primary endpoint	Key 2nd endpoint
D-PPAC	X	-	X	X	-	X
C-PPAC	-	X	-	-	X *	X
PPAC administration	•At Baseline, for 1 week (between V1 & 2) prior to randomisation at V4 •1st follow-up assessment: for one week between V5&6 in week 9 •2nd follow-up assessment: for one week between V7& 8 in week 12 PHT LogPad	At Baseline and At Month 12 Internet interface	Daily during 14 days during the run-in period for test-re-test purpose PHT LogPad	At Baseline and at the end of the study PHT Log¨Pad	At Baseline and at the end of the study Paper and computer version	1 week before randomization (V2) and at the end of study during week 12 (V3) PHT LogPad and Internet Interface

389

390 It should be noted that high-level data from two additional trials were expected to become available
391 during the qualification procedure but have not been reflected on during preparation of this opinion
392 document. (ACTIVATE Phase IV study evaluating a LABA/LAMA FDC (DUAKLIR®, GENUAIR®) in GOLD
393 II-III COPD patients; AZ Phase IIa study in GOLD III-IV COPD patients with a history of frequent acute
394 exacerbations with AZD7624, a new compound).

395 PHSYACTO, T9 TRIGON, EXOS and MRPAPP used/incorporated the D-PPAC. URBAN TRAINING, ATHENS
396 and MRPAPP used/incorporated the C-PPAC. Both tools have accordingly been validated independently.

397 As has to be expected, adherence to protocol differed between trials and this resulted in only a part of
398 patients contributing data to the final PROactive analyses for each trial (varying from 55% in study T9
399 TRIGON to 93% in PHYSACTO). Adherence criteria determining sufficient compliance for inclusion were
400 set arbitrarily. For validation purposes it is endorsed to focus on a sample indeed contributing data
401 points. No comparison of baseline characteristics between adherers and non-adherers were performed
402 and the possibility of systematic exclusion of certain patient groups (e.g. based on severity of
403 impairment) from the validation exercise cannot be fully ruled out. At the same time, it is understood
404 that the baseline and EOT data reported only reflect those patients eventually included in the analyses
405 which mitigates respective concerns.

406 Key demographics were largely comparable across trials and agreeably representative of a COPD
407 population. Overall, about half of patients were younger than 65 years, about two thirds were male.
408 Participants were predominantly non-smoking, retired and not living alone.

409 Table 2: Baseline demographics and comorbidities in WP6 trials

	Total	Physacto	T9	Exos	MrPapp	Urban Training	Athens
	m (SD) / n (%)	m (SD) / n (%)	m (SD) / n (%)	m (SD) / n (%)	m (SD) / n (%)	m (SD) / n (%)	m (SD) / n (%)
n	995 (100)	195 (20)	88 (9)	22 (2)	330 (33)	308 (31)	52 (5)
Age (years)	66 (8)	65 (6)	62 (8)	64 (7)	66 (8)	69 (9)	67 (9)
Age groups:							
<65	384 (39)	78 (40)	55 (63)	10 (45)	131 (40)	89 (29)	21 (40)
≥65 & <75	445 (45)	114 (58)	27 (31)	11 (50)	145 (44)	131 (43)	17 (33)
≥75 & <85	159 (16)	3 (2)	6 (7)	1 (5)	53 (16)	82 (27)	14 (27)
≥85	7 (1)	0	0	0	1 (0.3)	6 (2)	0
Male	701 (70)	124 (64)	52 (59)	17 (77)	209 (63)	257 (83)	42 (81)
Socioeconomic status: low	256 (67)			11 (50)		208 (68)	37 (71)
Living alone	167 (23)			5 (23)	83 (25)	40 (13)	39 (75)
Active worker	90 (13)			3 (14)	45 (14)	36 (12)	6 (12)
Current smoker	302 (30)	76 (39)	57 (65)	3 (14)	85 (26)	71 (23)	10 (29)
Weigth (kg)	77 (17)	79.9 (17.7)	78.6 (18.5)	75.2 (20.4)	75 (16.7)	77 (14.9)	77.6 (17.5)
Heigth (cm)	168 (9)	170 (10)	170 (9)	169 (10)	168 (9)	164 (7)	168 (9)
BMI (kg/m^2)	27.3 (5.1)	27.4 (4.8)	27 (6)	26 (5.7)	26.4 (5)	28.4 (5)	27.3 (5.1)
Heart Rate (bpm)	76 (13)	73 (11)	72 (11)	79 (11)	77 (13)	77 (14)	78 (11)
Systolic BP (mmHg)	136 (18)	133 (18)	132 (10)	125 (20)		140 (18)	
Diastolic BP (mmHg)	79 (11)	78 (10)	80 (8)	80 (14)		79 (11)	
Doctor diagnosed co-morbidities							
Anxiety	10 (2)	7 (4)	0	0	3 (1)		
Depression	44 (7)	27 (14)	0	3 (14)	14 (4)		
Cancer	26 (4)	14 (7)	0	2 (9)	10 (3)		
Any cardiovascular	149 (23)	66 (34)	18 (20)	3 (14)	62 (19)		
Diabetes	45 (7)	13 (7)	0	5 (23)	27 (8)		
Musculoskeletal	206 (20)	89 (46)	3 (3)	3 (14)	65 (20)		
Asthma	3 (1)	0	0	1 (5)	2 (1)		
Hypertension	183 (29)	92 (47)	0	12 (55)	79 (24)		

Some variables have missing data: 614 in Socioeconomic status, 283 in Living alone, 292 in Active worker, 464 in co-morbidities, 4 in HR, 385 in BP.

Relevant co-morbidities are listed in Table 2 as well. Importantly, keeping in mind the patient demographics, concomitant musculoskeletal disorders seem underrepresented in some of the trials, or respective data are missing (UT, Athens trials). Drawing on the inclusion/exclusion criteria of the concerned trials, all but one trial (i.e. T9 Trigon) explicitly exclude concomitant conditions that could interfere with PA, including orthopaedic, neurological but also, more generally, "other" respective complaints unrelated to COPD. Whereas it is evident that concomitant diagnoses interfering with a patients activity level would hamper demonstrating PPAC performance related to pulmonary activity limitations or improvement thereof, this might have created a somewhat artificial setting. As seen in the table above, the exclusion criteria did not prevent all patients suffering from potentially relevant conditions from entering the trials. Still, whether the PPAC tools would perform similarly (well) in a broad COPD population without abovementioned restrictions as regards co-morbidities in terms of staging COPD-related PA and being responsive to pulmonary improvement cannot conclusively be answered.

Table 3: Baseline COPD/physical activity in WP6 trials

|	Total	Physacto	T9	Exos	MrPapp	Urban Training	Athens
	m (SD)/n (%)	m (SD)/n (%)	m (SD)/n (%)	m (SD)/n (%)	m (SD)/n (%)	m (SD)/n (%)	m (SD)/n (%)
n	1083 (100)	283 (26)	88 (8)	22 (2)	330 (30)	308 (28)	52 (5)
Lung function							
FEV_1 (% predicted)	53.1 (18.1)	48.3 (12.7)	48.3 (12.1)	46.2 (19.6)	56.6 (21.7)	55.8 (17.79	51.3 (19.8)
ATS/ERS stages:							
Mild: $FEV_1 \geq 80\%$	86 (8)	1 (0.4)	1 (1)	0	53 (16)	27 (9)	4 (8)
Moderate: $FEV_1 < 80\%$ & $FEV_1 \geq 50\%$	493 (46)	124 (44)	41 (47)	8 (36)	136 (41)	161 (52)	23 44)
Severe: $FEV_1 < 50$ & $FEV_1 \geq 30$	409 (38)	139 (49)	41 (47)	9 (41)	104 (32)	95 (31)	21 (40)
Very Severe: $FEV_1 < 30$	95 (9)	19 (7)	5 (6)	5 (23)	37 (11)	25 (8)	4 (8)
FVC (% predicted)	88.7 (22.7)	104.1 (20.4)	79.7 (14.5)	82.6 (13.6)	92 (23.2)	75.6 (17)	79.4 (18.9)
FEV_1/FVC (%)	49.7 (13.4)	47 (10)	48 (13)	43 (14)	49 (15)	54 (12)	48 (14)
IC (% predicted)	76 (29)	72 (17)	68 (15)		86 (40)		73 (33)
SaO_2 (%)	95 (2)	96 (2)		96 (2)	95 (2)	95 (2)	95 (2)
Exercise capacity & muscle strength							
6MWD (m)	456 (103)	452 (100)		424 (81)	443 (105)	485 (94)	400 (113)
ESWT time (s)	303 (207)	295 (199)		401 (287)			
Patient Global Rating PA							
Not limited	169 (26)	50 (18)			87 (27)	32 (62)	
A little bit limited	324 (49)	162 (58)			146 (45)	16 (31)	
Limited	142 (22)	59 (21)			80 (25)	3 (6)	
Severely limited	21 (3)	9 (3)			11 (3)	1 (2)	

With regards to baseline lung function and exercise capacity the large majority of patients can be classified as GOLD 2/3, showing some degree of limitation regarding PA. Whereas this is expected to represent the COPD population at large, it is noted in the context of validating an outcome tool that for lung function patients at both ends of the scale are not well represented and for PA this particularly applies for those being severely limited. The 6MWD averages also indicate a reduced, yet considerable residual performance level. Accordingly, the Applicant stated that at the current stage, very severe COPD and/or patients currently suffering from an exacerbation (implying a rather dynamic disease state) are not considered a target population for applying the PPAC outcome tools.

Two (likely interdependent) observations can be made regarding the distribution of baseline D-PPAC and C-PPAC scores in the aggregated study sample:

438   Figure 7: Distribution of D-PPAC scores (left panel) and C-PPAC scores (right panel) at
439   baseline
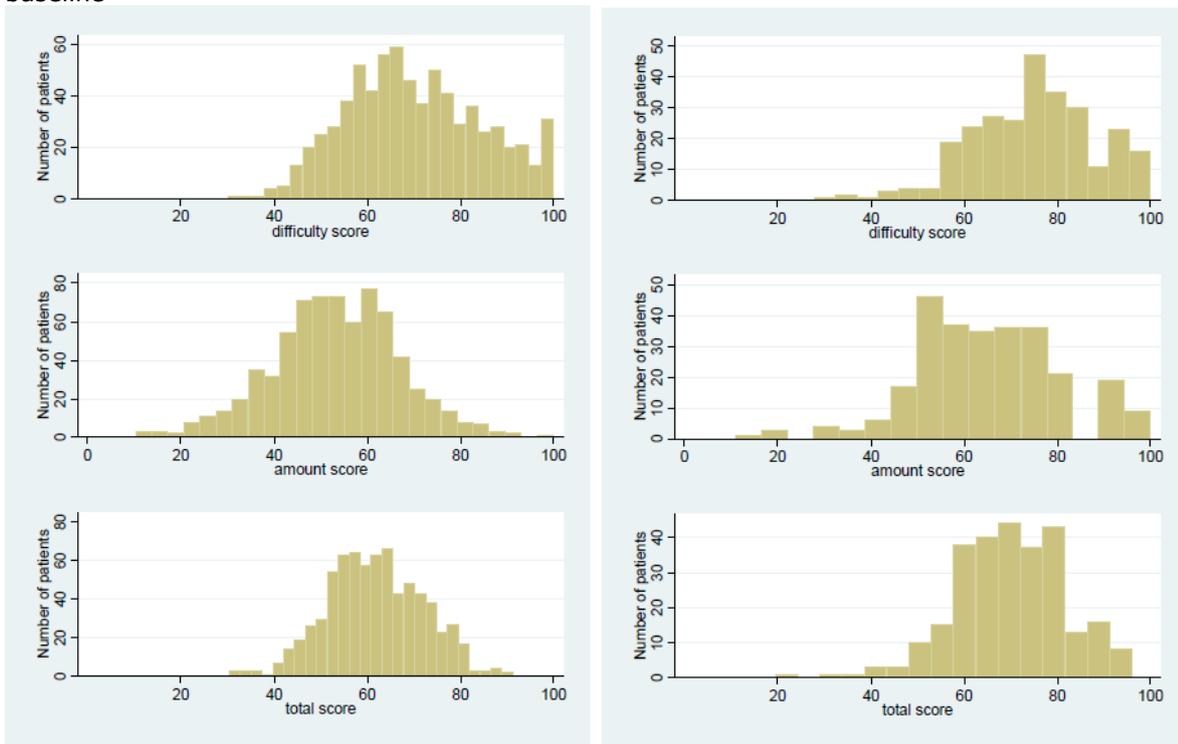
440

441   Firstly, it appears that, in line with statements made above on the disease severity of included
442   subjects, no patients scored at the lower end of either D-PPAC or C-PPAC in any of the clinical studies.
443   This applies to all three scales ('difficulty', 'amount' and 'total'), but is most pronounced for the
444   'difficulty' and 'total' scales where apparently no subjects scored below 40 (out of 100) and the
445   majority substantially higher. This means that the psychometric properties of the tools at the lower
446   end of possible scores were essentially left unaddressed during the WP6 validation exercise. Secondly,
447   when looking at known-groups validity, i.e. comparing PPAC scores with GOLD stage at baseline, it
448   appears that while showing variably pronounced separation in PPAC scores depending on GOLD stage,
449   even those patients with substantially impaired lung function (i.e. GOLD 4) scored relatively well on D-
450   PPAC and C-PPAC. The same holds true for dyspnoea (mMRC) and 6MWD results if employed as well-
451   known group denominators. Whereas these observations might be explained by patient selection,
452   populations appear rather comparable between the WP4 study conducted for initial validation and item
453   reduction and the WP6 trials, and the existence of a floor effect cannot be ruled out.

454   Given the differences in trial designs and PPAC data capture schedules, the different trials were not
455   equally able to contribute information for all the validation sub-tasks as listed in Table 5 below. From
456   the different trials, PRO response data of similar structure were pooled to obtain new datasets, each
457   one eventually foreseen for a specific part of the validation analyses.

458   For the D-PPAC three different datasets were derived for different validation analysis tasks:

459   'PDDR'-dataset: Pooled Daily PPAC day-by-day retest, to analyse Test-retest reliability;

460   'PDRB'-dataset: Pooled Daily PPAC Random baseline, to test Construct validity and confirm the

461   conceptual framework;

462   'PDRR'-dataset: Pooled daily PPAC Random repeated, to analyse responsiveness;

463   For the C-PPAC two different datasets were derived for different validation analysis tasks:

464   'PCB'-dataset: Pooled Clinical visit PPAC Baseline, to analyse Internal Consistency, Construct Validity,

465   and to confirm the conceptual framework;

466   'PCR'-dataset: Pooled Clinical Visit PPAC Repeated, to analyse responsiveness;

467   Details on the data-pooling/data-merging approaches are provided in the Statistical Analysis Plan of
468   WP6 [Annex Link 15]. The data management in this context was adequately described and
469   documented, and the data-sets used as basis for different validation tasks were considered suitable by
470   the QT.

471   One further important aspect in relation to the handling of data captured by the D-PPAC and C-PPAC is
472   the standardised approach of actual data aggregation. It was agreed with the Consortium that
473   qualification can only be considered for the format of data aggregation used in development and
474   validation of the tools.

475   For the D-PPAC the intention is to derive weekly averages, based on daily recordings and the need to
476   merge on a daily basis:

477   -   Response to valid daily questionnaire (no missing answers)

478   -   Values of steps and VMU/min if valid activity monitor data (valid means at least 8h of

479   monitoring)

480   -   calculate daily amount, difficulty and total score

481   -   calculate weekly mean if at least for 3 days in the week the questionnaire and monitor data

482   are available; data from days were only questionnaire data or only monitor data are available

483   are not taken into consideration for calculation of scores;

484   For the C-PPAC the intention is to use one weekly single measure, based on

485   -   Response to valid clinical visit questionnaire (no missing answers)

486   -   Median values of steps and VMU/min of three to seven valid days prior to clinical visit
487   questionnaire (at least 8h per monitoring day irrespective of weekdays/weekends)

488   -   Calculation of amount, difficulty and total score

489   One finding in the review of WP6 data was the rather divergent estimation of 'baseline' data in the
490   MrPAPP trial, dependent on which PPAC tool was used for data capture. The MrPAPP trial was the only
491   WP6 study in which both PROs were scored at baseline. According to the study results provided, the
492   PROs score 8-10 points differently on average in the same study population. Although the actual
493   patient set used was not identical for the two PROs to derive total scores (different 'n' obviously due to
494   differences in missing data structure), the differences seen in average scores are quite extensive, so
495   that interchangeable use of the two PROs within one trial setting cannot be supported based on these
496   findings.

497   Patient compliance to the PROs was another topic discussed in the framework of the qualification
498   procedure. Given the hybrid nature of the two tools (monitor + questionnaire data required from the
499   same data capture period/days), there is in principal an elevated risk for lower patient compliance if
500   data capturing is relying on more than one source. However, the Consortium concluded from the
501   different WP6 trials that in general compliance increased with 'importance' of measuring PA in the

502 specific trial setting. In this context, it is important to note that, whenever one of the two PRO tools is
503 intended to be applied, investigators and study personnel need to be adequately trained to
504 use/introduce the PPAC in a specific trial. This is expected to positively impact patient compliance. Of
505 course, also on the patient side, there is a need to provide appropriate information on how PPAC –
506 related activities are supposed to be handled during the conduct of the trial. For all of these purposes,
507 the adequacy of the User's guide [Annex Link 16] is of importance.

508 Data capture for the D-PPAC is supposed to be done with an electronic hand-held device. Relevant
509 experience was gathered in the clinical WP4 and WP6 trials. Questions regarding device selections as
510 well as questions relating to technical validity/performance where not directly addressed in the
511 framework of the qualification procedure. For the C-PPAC, a paper and pencil version as well as a web-
512 based interface was developed and tested by the Consortium. As for the D-PPAC, technical details to
513 support the electronic version of the C-PPAC have not been subject to assessment in this qualification
514 procedure.

515 So far, the D-PPAC in available in 62 languages whereas the C-PPAC is available in 14 languages.
516 Translation programmes included cognitive interviews performed with patients having the
517 corresponding language as mother tongue. Assessment of the translation work was not subject to this
518 qualification procedure.

519 Reliability, construct validity and responsiveness of both D-PPAC and C-PPAC were investigated in WP6
520 as outlined below:

521 Table 4: Psychometric properties tested per study

		EFPIA			Academic		External
		Physacto	T9	EXOS	MrPaPP	Athens	UT
PROactive instrument	Daily	X	X	X	X		
	Clinical visit				X	X	X
	Dynaport	X	X		X	X	X
	Actigraph			X	X		
Reliability Internal Consistency Test-Retest Reliability		X	X X	X	X	X	X
Construct validity Convergent validity Discriminant validity Known groups validity		X X X	X X X	X X X	X X X	X X X	X X X
Ability to detect change (responsiveness)		X		X	X	X	
Confirmation of Conceptual Framework		X	X	X	X	X	X

523 Psychometric properties D-PPAC:

524 As regards reliability measures, internal consistency and test-retest reliability were addressed.
525 Crohnbach's alpha was consistently >0.7 for both 'difficulty' and 'amount' domains in the total dataset
526 and in each of the 4 included studies. Test-retest reliability was tested using Intraclass Correlation
527 Coefficient values and Bland Altman plots. Only data from the T9 TRIGON study were used since it was
528 the only study that had repeated measures within a range of 7 (+/-1) days. Analysis was done by
529 comparing average measures of Week 1 with those of Week 2 but it should be noted that patients were
530 subjected to a change in medication at the beginning of week 1 compared to baseline. Results

531   (suggesting a high correlation) therefore have to be interpreted with caution, also because this
532   strategy was apparently chosen over a comparison of day 6 vs. day 13 scores in a data-driven manner.

533   Construct validity was addressed via correlation with related and unrelated constructs and with known-
534   groups expected to have differences in PA. Convergent validity was tested against different known
535   measures of dyspnoea, health status, exercise capacity and PA. Correlations were modest and varied
536   widely depending on domain and related construct applied and the Applicant attributes this to the fact
537   that PROactive instruments measure different concepts than already existing instruments, which is
538   difficult to ascertain. It is noted that for 'global rating of PA', a presumably simple construct, good
539   correlation with the PROactive instruments across domains would have been expected which was
540   apparently not the case. Expectedly unrelated constructs (i.e. height, heart rate, BP) were found to not
541   correlate with PPAC scores. As already stated above, known groups comparisons support the
542   differentiation of impairment severity via D-PPAC but only so over a limited range of the scale.

543   Caution is warranted regarding interpretation of responsiveness because clinical trials included in this
544   analysis did not include interventions of known efficacy. Thus, the PRO may falsely seem not
545   responsive, when the interventions are not effective. According to the Applicant, EXOS study results
546   were removed from responsiveness analysis because only 22 patients (distributed in 3 different groups)
547   participated. In PHYSACTO the response was more pronounced across all three domains in all
548   interventional arms tested, compared to the placebo arm. In MrPapp no change from baseline was
549   observed for either arm with the 'amount' domain being the sole exception where minor improvement
550   was observable for the telecoaching intervention and minor worsening for the usual care arm.

551   For the investigation of longitudinal validity, MrPaPP and PHYSACTO data were pooled and three
552   variables of self-reported global rating of change were categorised and possible responses to each
553   were grouped as follows:

554   • Global rating of change 'difficulty'

555   o much more difficult, more difficult, a little more difficult

556   o no change, a little easier

557   o more easy, much more easy

558   • Global rating of change 'amount'

559   o much less active, less active, a little less active

560   o no change, slightly better

561   o more active, much more active

562   • Global rating of change 'overall'

563   o much worse, worse, slightly worse

564   o no change, slightly better

565   o better, much better

566   Whereas the grouping of response possibilities into -/=/+ can be criticised as it limits a further
567   differentiation for quantity of change, the direction of effect as evident from all three D-PPAC domains
568   was concordant for each category of global rating.

569   Furthermore, differences between final and baseline PA levels were calculated using variables from the
570   activity monitors not included in the calculation of PPAC scores, including time in light, moderate and

571 vigorous PA, intensity, and lying, sitting, standing and walking time. According to the distribution of
572 the differences and their values, the following variables were used for longitudinal validity: changes in
573 time in moderate-to-vigorous activity, changes in time lying or sitting and changes in intensity and
574 each categorised in quintiles. Only results on change in time in PA are provided which support the
575 assumption of the scales being responsive, at least for the 1st and 5th quintiles, i.e., in those patients
576 with most increase or reduction in time in PA. Finally, 6MWD changes were compared to PPAC changes
577 and results indicate that concordance in response was only there for those patients increasing their
578 walking distance but not for those showing a reduction as in these patients PPAC scores stayed stable
579 over time.

580 For determining a potential MID of the D-PPAC, anchor-based as well as distribution-based methods
581 were used relying on PHYSACTO and MrPapp data. 6MWD, CCQ and SGRQ as well as change in global
582 rating ('total', 'difficulty' and 'amount') were considered as established outcomes that could serve as
583 candidate anchors. Correlations between these candidate anchors and the three PPAC domains were
584 however rather low, somewhat surprisingly also so for change in global rating. Since there are three
585 categories of GRCs (worse, no change or little easier, better), the mean change in the amount score in
586 patients which reported an improvement in the global ratings of change was chosen to represent the
587 MID. In order to be consistent with the estimation of MIDs based on GRC the mean change in the
588 difficulty score in patients who had improvement in the CCQ of at least -0.4 (MID of CCQ (Kocks et al.
589 2006) or of at least -4 (MID of SGRQ - Schünemann et al. 2003) were selected as MIDs. For the GRC
590 the mean change in the difficulty score in patients who reported an improvement in the GRC difficulty
591 was considered to represent the MID. 6MWD was disregarded for the low correlation with PPAC scores.
592 The obtained MID estimates were between 5.2 and 7.8 for the difficulty score and 4.7 and 6.7 for the
593 amount score. The anchor- and distribution based methods yielded similar results but it is noted that
594 SDs were quite large. Based on that, a MID of 6 for the amount score and a MID of 6 or 7 for the
595 difficulty score was deemed optimal. In order to simplify the interpretation it was suggested to use a
596 MID of 6 for both scores of the D-PPAC. For the total score the MID estimates were between 2.0 and
597 5.7. For this score it was suggested to use a MID of 4.

598 The anchors and their respective MIDs used seem reasonable based on the cited literature but the low
599 correlations with PPAC and the assumed independency of concepts clearly renders "global rating"
600 anchors more meaningful than others. Derived estimates for MIDs for 'amount' and 'difficulty' derived
601 showed some differences and where pragmatically and uniformly set across tools and scores for
602 reasons of simplification. In this context it is noted that the Company states: "PA can be considered
603 relevant (i) when a given improvement in amount is achieved without more difficulty, (ii) when less
604 difficulty with PA occurs without deterioration in the amount, or (iii) both less difficulty with activity
605 and a greater amount of activity are demonstrated." This simple approach can be followed to jointly
606 consider the 'amount' and 'difficulty' domains in specific scenarios but does not consider situations
607 where certain deteriorations in either domain might be accompanied by substantial gains in the other
608 (which could result in a net benefit). The 'total' domain combining amount and difficulty can be a
609 remedy but the lower proposed MID is clearly questioned as less than meaningful improvement on
610 either amount or difficulty paired with no change in the respective other domain, could be considered
611 meaningful in the total scale which is counterintuitive. Overall, how certain changes in the three
612 domains would be perceived by the patient, likely also depending on baseline values, seems not
613 conclusively answered. MID determination usually focuses on the immediate benefit associated with
614 certain quantitative changes in the concerned score rather than the predictive value of such changes
615 for other (preferably long-term) outcomes with established or intrinsic clinical relevance such as
616 survival. The latter however also constitutes a viable strategy for making PRO outcomes interpretable
617 and informative for benefit assessment of experimental interventions. So far, the predictive properties

618 of certain baseline and/or changes in PPAC or subdomains for e.g. survival, dependency, or lung
619 outcomes such as exacerbations, etc. were not investigated during validation. Feasibility constraints for
620 such analyses are acknowledged however, at least for survival, looking at the duration/size of studies
621 included in WP6.

622 <u>Psychometric properties C-PPAC:</u>

623 As regards reliability measures, only internal consistency was addressed on MrPapp and UT data. Test-
624 retest reliability was not studied because of the design of the included studies. None of the studies
625 included a repeated questionnaire one week apart. Crohnbach's alpha was consistently >0.7 for both
626 'difficulty' and 'amount' domains in the total dataset and in each of the 2 included studies.

627 Construct validity was addressed via correlation with related and unrelated constructs and with known-
628 groups expected to have differences in PA. Convergent validity was tested against different known
629 measures of dyspnoea, health status, exercise capacity and PA. As seen for the D-PPAC, correlations
630 were modest and varied widely depending on domain and related construct applied and the Applicant
631 attributes this to the fact that PROactive instruments measure different concepts than already existing
632 instruments, which is difficult to ascertain. It is noted that for 'global rating of PA', a presumably
633 simple construct, good correlation with the PROactive instruments across domains would have been
634 expected which was apparently not the case. Expectedly unrelated constructs (i.e. height, heart rate,
635 BP) were found to not correlate with PPAC scores. As already stated above, known groups comparisons
636 support the differentiation of impairment severity via C-PPAC but only so over a limited range of the
637 scale.

638 MrPapp and ATHENS data were used to analyse responsiveness of the C-PPAC. The Athens study was a
639 4-arm study designed to compare the paper-pencil with the electronic version of the PROactive
640 instrument. With this study, the patients who participated in the rehabilitation program were
641 randomized in four groups: Group A included patients who only used the paper-pencil version of the
642 clinical visit PPAC; in Group B patients used the electronic version of the clinical visit PPAC; Groups C
643 and D were used as control groups with patients only receiving the usual standard of care. In both
644 trials, the intervention arms displayed higher response across C-PPAC domains compared to control.
645 The control arms, particularly those in the ATHENS trial, also reflected varying degrees of worsening
646 across domains. Overall, and as seen for the D-PPAC, C-PPAC seems capable of reflecting changes to
647 PA. The subjects dealing with the paper version showed more marked response (in both directions)
648 than those dealing with the electronic version, but no formal comparison of the two modalities was
649 made.

650 For the investigation of longitudinal validity, only MrPaPP data are referred to and three variables of
651 self-reported global rating of change were categorised and possible responses to each were grouped in
652 the same manner as described above for the D-PPAC. The direction of effect in all three C-PPAC
653 domains was concordant with each category of global rating, thus supporting the notion of longitudinal
654 validity. Furthermore, as for the D-PPAC, differences between final and baseline PA level and 6MWD
655 were calculated and grouped in quintiles. Concordance with C-PPAC changes can be observed with
656 exception of the 'difficulty' domain not reflecting changes in PA which can however potentially be
657 explained by patients adapting 'amount' while maintaining stable levels of 'difficulty'.

658 For MID determination, same methods as for the D-PPAC were used but only MrPapp data were
659 considered. As seen for the D-PPAC, correlations between candidate anchors and the three PPAC
660 domains were rather low. The MID estimates ranged between 2.8 and 6.8 for the difficulty score and
661 4.5 and 7.9 for the amount score across anchors, all estimates with little precision. A MID of 5-6 was
662 considered appropriate for the amount and difficulty scores of C-PPAC by the Applicant, but 6 was kept

663 for reasons of consistency with the D-PPAC. For the total score the MID estimates ranged between 3.4
664 and 5.9. For this score it was also suggested to use the same MID of 4 as for the D-PPAC. The critical
665 discussion provided above on MID derivation applies similarly for the C-PPAC.

666 For further details of analyses results of WP6 see [Annex Link 17].

667 During the Qualification procedure the topic of the 'Context of Use' for the two different PRO tools
668 (separately) was further discussed with the Consortium. The idea was that the choice of Daily or
669 Clinical Visit tool is driven by the clinical hypothesis being tested and therefore the study design. The
670 suggestion for the C-PPAC was that it would more likely be used where patients' experience of PA is a
671 supportive outcome and/or where patient burden of completing PROs is high. In relation to the
672 intended use of the C-PPAC, also 'pragmatic studies' to gather real-world evidence (e.g. 'minimal'
673 intervention studies) where suggested. The D-PPAC was suggested to be used in the context of study
674 settings where measurement of patient experience of PA is an outcome of primary interest. The
675 Consortium's idea was that whenever "label-claims" could result from PA data analyses, the basis for
676 calculation should be the D-PPAC.

677 Finally, the presented results of WP6 could be continuously updated/confirmed with data from trials
678 still ongoing during consultation or planned for the future. It is further stated that stratified results for
679 all validity analyses are available (i.e. based on gender and COPD staging) and respective high-level
680 data might also be useful for public domain to support applicability of PPAC across relevant substrata.

681 **CHMP opinion**

682 The Consortium developed two PRO tools, the D-PPAC and the C-PPAC to capture physical activity (PA)
683 data in patients with COPD in clinical trial settings. Both tools are hybrid tools, combining information
684 from questionnaire items with PA monitors read-out data. State-of-the-art qualitative methodology has
685 been applied in the development phase to build a conceptual framework that eventually combines two
686 domains: 'amount of PA' and 'difficulty with PA' into one concept for each of the two PRO tools. This
687 conceptual framework is considered appropriate to describe PA in COPD. In general, adequate
688 quantitative methods have been used to identify the optimal sets of items, monitor read-outs and
689 response categories which finally comprise the D-PPAC and the C-PPAC. In the framework of the
690 qualification advice/opinion procedures, there was no dedicated assessment of technical details of
691 electronic formats for the D-PPAC (hand-held) and the C-PPAC (web-based solution). It is also
692 important to note that translation work carried out for the two PRO tools was also not subject to this
693 qualification procedure.

694 With a recall period of 24 hours the D-PPAC allows to collect data on a daily basis. Derived data is
695 converted to two domain scores and one total score, based on weekly averages. The recall period as
696 well as the actual data aggregation approach is endorsed for the use of the D-PPAC. It is agreed to the
697 consortium that due to the higher amount of information collected on a daily basis, the D-PPAC
698 qualifies for a context of use where a clear (primary) focus is on measuring PA. In the decision to apply
699 the D-PPAC in a specific study, expected patient burden should however be considered and weighed
700 against the importance of PA as study objective.

701 The C-PPAC has a recall period of 7 days, which is indeed considered an adequate period to capture PA
702 data reflecting weekly (repeated) routines of COPD patients' daily life. As for the D-PPAC, data is
703 converted to two domain scores and one total score for the C-PPAC. The suggested context of use for
704 trial settings where patients' experience of PA is a supportive outcome and/or where patient burden of
705 completing PROs can be expected to be high can be endorsed in principle.

706 It is important to note that for both tools, the D-PPAC and the C-PPAC, item selection/optimization was
707 done separately for the two domains, respectively. There was no overall (items) evaluation of
708 optimality regarding the PROs' single components (i.e. items and monitor read-out data). The derived
709 *total* score is the arithmetic mean of the two domain scores (amount & difficulty), giving the two
710 domains equal weights in computation. According to the Consortium, also different weighting was
711 explored, but equal weighting was considered most adequate. Currently, reporting domain scores
712 separately is considered to improve the information as one domain may be more (or exclusively)
713 affected by a specific intervention. Therefore, it seems advisable to focus eventual interpretation of
714 PRO results on the two resulting domain scores for 'amount' and 'difficulty' next to each other, rather
715 than on the *total* score. Further development work seems indicated to pursue the goal of having a total
716 score being most informative for PA in the trial settings targeted.

717 The Consortium's validation work contains an attempt to determine minimal important differences
718 (MIDs) on the PROs' domain- and total scales. Whilst anchor-variables and their respective MIDs seem
719 reasonably selected, low correlations between some of the anchors and PPAC scores were observed.
720 These findings might just reflect the fact that PA - as the new entity of interest - is indeed rather
721 independent from other established measures commonly used in COPD. Uncertainty remains how
722 certain changes in the PRO domains would be perceived by the patient (likely also depending on
723 baseline values). MID determination focused on the immediate benefit associated with certain
724 quantitative changes in the concerned score rather than the predictive value of such changes for other
725 (preferably long-term) outcomes with established or intrinsic clinical relevance (e.g. survival).
726 Although it has been demonstrated that activity limitation is associated with poorer prognosis and
727 reduced survival in COPD, the predictive properties of certain baseline and/or changes in total PPAC or
728 domains for e.g. survival or lung outcomes such as exacerbations etc. have not been investigated at
729 this time.

730 In the validation work for the two tools, psychometric properties were evaluated on basis of patient
731 sets which excluded individuals who might have scored at the lower end of the domain/total scales.
732 This means that interpretation of derived psychometric properties for the two tools is limited to data
733 ranges corresponding to central and upper parts of the underlying score-data distributions. In how far
734 this corresponds to restrictions in targeted COPD patient population or is related to a potential floor
735 effect of the tool (i.e. being insensitive to differentiate among worse PA scores) remains currently
736 unclear (e.g. GOLD 4-categorised patients were found to score relatively high on the D-PPAC and C-
737 PPAC). Patients with relevant comorbidities potentially interfering with PA have also been
738 systematically excluded from validation trials which might either require further restrictions or careful
739 interpretation of PA data collected in such patients.

740 The D-PPAC and the C-PPAC are not designed to be used and should therefore not be used
741 interchangeably in a single study.

742 From a technical perspective, the Opinion provided here is formally restricted to the PROs' use
743 involving either Actigraph G3TX and the Dynaport MoveMonitor worn at the waist. No recommendation
744 is currently possible in relation to the use/implementation of other monitor devices in the data
745 capturing of the D-PPAC and the C-PPAC.

746 The original Consortium's request for Qualification opinion contained a suggestion for two Clinical trial
747 endpoint models where the new C-PPAC and D-PPAC were proposed to be used as secondary, or even
748 as primary efficacy endpoints in COPD trials. The current EMA *Guideline on clinical investigation of*
749 *medicinal products in the treatment of COPD* (EMA/CHMP/483572/2012-corr) mentions PA as a
750 potential secondary endpoint, and contains clear recommendations regarding primary endpoints to be
751 envisioned in various study/patient population settings. During the qualification review, and as

discussed with the consortium at the discussion meeting, it became clear that discussion around clinical endpoint models and potential positioning of PA in the hierarchy of important endpoints in COPD trials should be kept separate from the actual qualification aim, which is to declare the two new PRO tools suitable to capture PA in COPD patients as intended. It was therefore decided to strive for qualification without touching the issue of whether the PROs are suitable to inform (co)primary/secondary (etc.) endpoints in the various suggested contexts of use. Against this background, positioning of endpoints and targeted claims have not been discussed/agreed in the margins of this qualification procedure.

Incorporating findings based on the PRO tools in 5.1 of the SPC of a compound targeting COPD seems possible but specific content or wording cannot be pre-empted at this point in time and will largely depend on the effects shown in a specific development programme and the perceived relevance of such information to the patient/prescriber, accounting for overall results. As discussed above, the interpretation of certain changes observable on PPAC and its subdomains in terms of magnitude and associated patient-perceived benefit is considered difficult and might require further context, i.e. embedding in other (secondary) outcomes.

Annexes

- Applicant submission – Link 1. The English versions of the Daily PROactive Physical Activity in COPD (D-PPAC).

- Applicant submission – Link 2. The English versions of the Clinical Visit PROactive Physical Activity in COPD (C-PPAC).

- Applicant submission – Link 3. Reference GOLD 2015. See PROactive references in final request.

- Applicant submission – Link 4. Reference Hopkinson and Polkey, 2010. See PROactive references in final request.

- Applicant submission – Link 5. Reference Caspersen et al. 1985. See PROactive references in final request.

- Applicant submission – Link 6. Reference to Holland et al. 2014. See PROactive references in final request.

- Applicant submission – Link 7. PROactive Work Package 2A: Input from the literature - Report on Systematic Review 1

- Applicant submission – Link 8. Report on work Package 2A review 3 - Activity monitors for potential use in COPD

- Applicant submission – Link 9. PROactive Work Package 2A: Input from the literature - Report on Systematic Review 5

- Applicant submission – Link 10. Reference Van Remoortel et al. 2012. See PROactive references in final request.

- Applicant submission – Link 11. Reference Rabinovitch et al. 2013. See PROactive references in final request.

- Applicant submission – Link 12. WP4 study protocol. See Appendix 10 WP4 Clinical Study Protocol

791   -   Applicant submission – Link 13. WP4 statistical analysis plan. See Appendix 7 - WP4 Statistical
792   Analysis Plan.

793   -   Applicant submission – Link 14. WP4 "details of analyses results". See Appendix 11 WP4 Clinical
794   Study Report

795   -   Applicant submission – Link 15. "Details on the data-pooling/data-merging approaches are provided
796   in the Statistical Analysis Plan of WP6". See Appendix 12 - WP6 Clinical study synopses.

797   -   Applicant submission - Link 16. User Guide. See Appendix 16 in final request

798   -   Applicant submission –Link 17. "analyses results of WP6". See Appendix 14 - WP6 clinical validation
799   study report.

800   [1] All annexes mentioned under the Applicant's position refer to the documentation submitted with the request.