Garcia-Aymerich J, Puhan MA, Corriol-Rohou S PROactive consortium, *et al*. Validity and responsiveness of the Daily- and Clinical visit-PROactive Physical Activity in COPD (D-PPAC and C-PPAC) instruments. *Thorax* 2021;76:228-238.

On today's episode, I am excited to discuss an article published in Thorax in 2021 – entitled: Validity and responsiveness of the Daily- and Clinical visit-PROactive Physical Activity in COPD (D-PPAC and C-PPAC) instruments. The corresponding author is Dr. Judith Garcia-Aymerich from ISGlobal in Barcelona, Spain. I was fortunate to visit ISGlobal in 2019 and spent some time with Judith, so it's especially nice to be able to discuss one of her papers. This is an open access article, so freely available – I will put the link in the show notes.

In this study, the overarching objective was to better understand the PROactive Physical Activity COPD instruments: D-PPAC and C-PPAC. Before I talk about the paper, I'll provide a brief background about PROactive. The PROactive consortium is a 5-year European Innovative Medicines Initiative and is made up of 19 partners, which includes 8 pharmaceutical companies, several research institutions, and patient organizations. The PRO part of the name is in reference to Patient Reported Outcomes, which you might recall was the topic of a previous episode. The mandate of the consortium is to develop tools that capture daily physical activity, based on the patient's needs, that are inclusive of the whole experience of physical activity.

So, over the last several years, the PROactive consortium has been conducting different types of research to create these tools. They have completed qualitative research projects where patients spoke of their experience of physical activity. They have looked at different wearable devices to measure physical activity. They have provided guidance on how to use those wearable devices. And, they have developed these two instruments, the Daily PROactive Physical Activity in COPD instrument, which has the acronym D-PPAC, which I'll call D-PAC in this podcast, and the Clinical PROactive Physical Activity in COPD instrument, which has the acronym C-PPAC, which I'll call C-PAC.

The Consortium has tested these tools in validation studies, but they wanted to assess validity and reliability across more countries and across a wider range of disease severity. They also wanted to examine if the measure was responsive, and to estimate the minimal important difference. I'll speak a bit more about what this means in a bit.

So, as I said, these instruments were developed to capture the experience of physical activity. Through their research, the Consortium found that neither an activity monitor alone nor a questionnaire alone could discriminate the 'full experience of physical activity'. Patients reporting their activity doesn't provide an accurate quantity, and it can be difficult to estimate the change in activity due to an intervention, so the self-report has some problems. But step counts on their own don't tell the whole story either – they don't properly describe how a person's physical activity can be affected by a disease, or by an intervention. But when self-report and an activity monitor are combined, then they provide a more complete picture. Therefore, the D-PPAC and the C-PPAC combine the questionnaire and activity monitor to estimate both the 'amount of physical activity' and the 'difficulty with physical activity' in order to better capture the patient's activity experience with COPD.

So, what do these instruments look like? D-PPAC questionnaire includes 7 items which are answered every day, in the evening, for one week, via an app. The questions include asking about the amount of physical activity, for example: How much walking did you do outside today? as well as the difficulty of physical activity, for example: How much difficulty did you have getting dressed today? Then, the instrument has a section where the step counts of the data are categorized and scored, for example: if you walked less than 1000 steps, you would get a score of 0, whereas 1001-3000 steps would get a score of 1, etc. The vector magnitude units, or VMU, are also included in the scoring. This reflects the velocity of movement. It should be noted that for the whole instrument to be accurate, the step counts and VMU need to be accurate. So, a simple wearable, such as a FitBit, wouldn't be appropriate, as it doesn't measure steps accurately enough. To use this tool, either an Actigraph or a Dynaport accelerometer must be used. So with both these measures: the self-report and the activity monitor, scores are the summed, converted and then you have one total score, with possible ranges between 0 and 100, with higher values indicating a better score.

The C-PPAC questionnaire is similar but includes 12-items and has a 1-week recall that is completed the day of each study visit. The activity monitors are worn during the waking time for 1 week at each study visit; therefore, they had to have more than 8 hours of data for at least 3 days during that week for the data to be considered valid.

So, you can see that these instruments could be an improvement on measures that include just patient report, or just activity monitor data. By combining them, weighting each component and scoring them within the same instrument, you can have an instrument that both includes how active people are with how difficult activity is. But in order for it to be useful in research studies and clinical settings, it needs to be responsive. What does that mean? Well, responsiveness reflects the ability of an instrument to actually measure change, when change occurs. You can have an instrument that measures something, but if its designed in a way that doesn't capture change, it doesn't have much use as an outcome measure. So the investigators wanted to see if the instruments were able to measure change when change does occur. The other thing they wanted to do was to estimate the minimal important difference, or MID. The MID is the smallest amount of change in an instrument that the patient perceives as being related to actual change. This is important to calculate because you can often calculate a difference that is statistically different, but if the difference is too small for the patient to perceive, or to be clinically important, then what does it matter? So the MID allows you as the reader and clinician to have a good idea if the change that is measured is likely to be clinically important to you, as the clinician, and your patient.

So, to calculate these things, the investigators included data from 7 prospective randomized controlled trials from 17 countries in Europe and North America. These trials were testing the effect of pharmacological and non-pharmacological interventions of patients with COPD. Each of these trials contributed slightly differently to the study for the evaluation of different measurement properties; however, all studies contributed baseline data for reliability and validity.

In addition to the D-PAC and C-PAC scores, the investigators also had data on demographics, lung function, 6MWD, the modified MRC Dyspnoea scale, the Chronic Respiratory Disease

Questionnaire, the Clinical COPD Questionnaire, and/or the COPD Assessment Test (CAT). They had a lot of different measures on different attributes of the disease. Finally, the patients were also asked to rate their global change in physical activity with regards to amount, difficulty, and overall since the start of the study they were in, and they used a 7-point Likert-type scale to quantify this. The Likert-type scale ranged from 'much worse' to 'much better'.

How are validity, reliability, and responsiveness calculated? For validity, you look at relationships between the instruments, and other measures. So, if the measure is valid, you'd expect it to have a strong association with another measure, so an example, if you are measuring physical activity you would probably expect to see an association with the 6MWD. So across the whole sample, if those that had lower 6MWD also tended to have lower scores on their C-PACC or D-PACC, that would be an indication of validity. Reliability reflects the ability of the measure to provide similar values over time, or between one assessor and another, if no actual changes happened in the patient. So you would want to see a strong association between the measurement at one time and the next – and in this case, they had data from a 2 week period with no intervention. As well, for the reliability and validity analyses, it was investigated by sex, age groups, COPD severity, country, and language.

Responsiveness was examined by determining if change occurred in other outcomes known to improve after an intervention, such as pulmonary rehabilitation, for example. As they also had patient report of change using that Likert scale, they could determine if when patient's reported change as part of the study they were in, did the D-PAC and the C-PAC also change? They can also then calculate how much change in these instruments was necessary for it to 'register' with patients as actually being a change.

So, they had data from over 1300 patients, and they created one group of 950 patients to assess the D-PAC and 651 patients to assess the C-PAC. On average, these patients were about 66 years old, but were approximately 70% male. Most were ex-smokers, with an FEV1 of about 55% of predicted. They came from many countries but predominantly in North America and Europe. The instrument scores showed the whole range of 0 to 100.

Their analysis showed the instruments were highly reliable. In patients with no clinical change in a two-week period, the instruments also showed little change in values. When looking at validity, the instruments showed little association with measures of quality of life. This is as expected, as although there is some relationship between how active we are and the quality of our lives, these are not the same things, so you would not expect to see a strong relationship there. The amount of physical activity as scored by the instruments had a moderate association with 6MWD values, and strong correlations with objective physical activity measures. The instruments were able to differentiate between those low physical function measures and high. They were also able to differentiate between those with worse and better lung function, and those with worse and better dyspnea levels.

When looking at responsiveness, the investigators also found that the instruments were responsive to change in both pharmacological studies and pulmonary rehabilitation studies. They calculated the MID to be a score of 6 for the amount and difficulty scores, and 4 for the total scores.

So, this is good news, the results from this study show that these instruments are valid, reliable, and responsive. They would be useful in both clinical trials of pulmonary rehabilitation, as well as outcome measures for pulmonary rehabilitation programs. They work well across a broad spectrum of COPD severity, patient ability, and in different countries. This study also showed that patients scored differently in the amount scores versus the difficulty scores. This reinforces the fact that amount of physical activity and difficulty of physical activity are two different things and need to both be considered when assessing physical activity.

The investigators found that the C-PACC values were higher than the D-PACC values in a given patient, which they attributed to the possibility that patients, when recalling physical activity over a 7 day period, tend to overestimate the amount of activity they are doing compared to when they report on their activity on a given day. It is not just in patients with COPD where this occurs this is quite well known, the longer thee recall period the more we tend to estimate the answers usually to our favor in many cases. So, they reinforce that the C-PACC and the D-PACC cannot be used interchangeably, so in a study they could not use the clinical measure in one day and then the daily measure in another day and expect those measures are going to be comparable.

Some of the limitations of this study were noted. The trials were not trials that recruited individuals with an acute exacerbation of COPD. So, although they had patients with very low scores, they were all still considered clinically stable. Another limitation is that the sample were people who had entered clinical trials. These people are often motivated to be a part of research, and stable enough to withstand the demands of participating in research. It is difficult to know if the same results would occur outside of clinical trial recruits. They didn't mention it, but it continues to be a problem to investigate measures interventions and measures with a markedly smaller proportion of women in the sample. 25% women or 30% of women does not reflect the prevalence of COPD in women anymore, and study investigators need to work harder to remove barriers for women to participate. Some of the strengths of the study were the large sample size, across many countries with different languages, and a wide range of physical abilities, and with different interventions.

So clinically, can you use these instruments in your clinical practice? If you have the activity monitors that they used in this study, namely the Actigraph or the Dynaport, then you might find it interesting to use these instruments with your patients, to learn more about their baseline physical activity and changes to it. However, it is important that it is these activity monitors that are used. The results would not be valid if you swapped out a different activity monitor, or just used the questionnaire part of the instrument without the activity monitor part. Unfortunately, those activity monitors are not cheap, so it is unlikely that many pulmonary rehabilitation programs will invest in these.

So how do programs benefit? Well, informally I would say that this study, and these instruments give you better insight as to how to discuss physical activity with your patients. If your questions related to physical activity are only focused on the amount they do, then you're missing an important dimension of physical activity, that of difficulty. So both areas should be part of the discussion. But I think also it is important for clinicians to understand how instruments are developed, and their validity, reliability, responsiveness, and MID, because these will be the instruments that will be used to report physical activity in pharmacological and pulmonary

rehabilitation studies in the future. So, you'll see these instruments reporting values, and it will help your interpretation of the results if you have a background understanding as to how these measures were developed and validated.

In conclusion, as the investigators state, this study supports that the D-PPAC and C-PPAC instruments are valid and reliable across sexes, age groups, COPD severities, languages, and countries and are responsive to drug and non-drug treatments and changes in clinically relevant variables, especially many of those that are quite relevant in pulmonary rehab. But, there does need to be future work to confirm this in people with an AECOPD, across further countries and languages in Africa, Asia, and the Middle East, and of course we need to recruit more women.

Now this paper was long, with a lot of technical language that is important to report but can be a bit challenging to interpret. But I hope that this episode introduced you to a measure that you'll likely see in research papers in the years to come and gave you an understanding of some of the concepts they addressed.

Thank you for listening to this episode, and until next time, keep well and keep moving.