Health-related quality of life is a clinical outcome that is unique to each individual. It can be assessed subjectively by a clinician, but cannot be measured in a standardised manner. In contrast, health status which is a marker of impaired health-related quality of life, is measurable and such measurements are now performed routinely in clinical trials to estimate the size of treatment benefit. Outside the setting of clinical trials, health status questionnaires have shed important insights into a number of areas of chronic obstructive pulmonary disease, particularly in terms of symptomatic progression over time, which appears to be confined largely to effects on physical activity and psycho-social areas of health-related quality of life, rather than specific respiratory symptoms. Development of short simple instruments has brought health status measurement into routine practice where it may be particularly useful for baseline assessment, monitoring trends, and assessment of recovery after acute events such as exacerbations or hospitalisation.

Key words: COPD. Health status. Questionnaire. St George’s Respiratory Questionnaire.
INTRODUCTION: TERMINOLOGY

The title of this review refers to health-related quality of life (HRQoL), which is appropriate since it refers to the impact of the disease on a patient’s daily life and physical and social well being. Each of us is an individual, so our lives will be affected by illness in ways that are unique to each of us. The concept of quality of life is entirely valid in that context. However, there are many factors that affect a patient’s quality of life and health may not be the most important one, hence the more specific term “health-related quality of life”. Improvement in HRQoL is an important treatment outcome in chronic obstructive pulmonary disease (COPD) and can be assessed at an individual patient level through the clinical history, since it is “narrative” assessment that can take into account the individual’s personal circumstances. However, even the assessment of an individual patient may differ between clinicians since it is a subjective and largely non-standardised evaluation. Measurement is at the heart of clinical medicine: blood pressure, body mass index (height and weight), heart rate, degree of airflow limitation, arterial oxygen saturation; all of these are assessed routinely in a standardised way using measurement instruments. The case for standardised assessment of symptoms in routine practice is overwhelming, however measurement of symptoms is not straightforward. For example, the UK Medical Research Council Breathlessness Questionnaire measures disability due to dyspnoea or the effect of exercise on breathlessness (it is not clear which) rather than breathlessness itself. More recent instruments have recognised this problem and are more specific in terms of what they are trying to measure; for example, the Shortness of Breath with Daily Activities questionnaire (SoBDA), whose name is self-explanatory.

It can be argued that it is impossible to measure HRQoL because measurement implies standardisation and, since all patients are different, standardisation is not possible. There have been attempts to develop individual quality of life questionnaires, but these are rare due to the conceptual difficulty of designing and validating such instruments. Standardisation of symptom questionnaires can only be achieved by selecting items that should be applicable to all patients who have the condition being assessed. This may require difficult decisions; for example, for many COPD patients the inability to play with grandchildren significantly impairs their HRQoL, but not all patients have grandchildren so an item like that would be inappropriate. It is for these reasons that I prefer the term “health status measurement” to draw attention to that fact that what is being measured is health impairment, not HRQoL. The latter is an individual patient-level clinical outcome, whereas health status measurement is a marker of that outcome, measured using standardised instruments.

WHY AND WHAT OF HEALTH STATUS MEASUREMENT

Chronic obstructive pulmonary disease is a very complex condition, even without considering its comorbidities. There are multiple pathways from disease in the lungs to impaired health status, and whilst breathlessness and its impact on exercise capacity may be the single most important factor, it is not the only one (Fig. 1). Whilst assessment of a specific problem such as breathlessness is useful,
what is required more often is an overall measure of the disease's impact. Measurement of health status is performed using questionnaires completed by patients and in this respect they are subjective, not objective. Impaired HRQoL can only be experienced by the patient and only be captured through their account of their symptoms and the resulting impact on their lives and wellbeing. It is in this respect that they are subjective, but if we accept that a clinical history is valid and reliable, then a properly designed health status measure will also be valid and reliable (when used for the purposes for which it was designed).

Measurement implies a scale, like a ruler or thermometer, and one of the characteristics of these instruments is that the distance between any two points implies the same size of difference, regardless of where the two points lie along the scale. A 2°C difference in temperature is the same whether between –1 and +1 or 99 and 101; this example illustrates why it is important to distinguish between the size of change and the implications of that change. At one end of the scale, ice becomes water and at the other, water becomes steam – same temperature change, but the consequences of that change depend on the baseline state. This distinction is important because a commonly asked question is, “Does a change in health status score mean the same thing in mild as opposed to severe patients?” The answer is “yes”, but only if the questionnaire has invariant measurement...
properties, i.e. it behaves like a ruler. That cannot be assumed, however.

Let us take a simple example of a questionnaire item that has five response options. It is usually assumed that the severity “distance” between each category is exactly the same (Fig. 2 A); however, that may not be so. Modern methodology using Rasch modelling has shown that the picture is much more likely to be similar that shown in figure 2 B. Indeed, the categories may even be disordered, such an apparent logical severity sequence is not followed, when data from patients are collected and analysed appropriately. This is a problem since most questionnaires have been developed using the assumption of equal-interval response scales. A related question concerns the severity associated with each item. All too often it is assumed that each item has an equivalent level of severity. This is important, since calculation of an overall or total score requires the responses to the different items to be summed, but simple sums may not be appropriate if each item has a different degree of importance. This problem was recognised during the development of the St George’s Respiratory Questionnaire (SGRQ) and for that reason each response to each of the items has an empirically derived weight4,5 that plays a key part in calculating the total score.

Rasch methodology resolves the issue of creating a valid total score by using an entirely different technique. It contains a suite of tests, one of which includes testing the probability of a patient of a given level of severity responding to one specific item in the same way as they did to all of the other items in the set; i.e. it tests how well each item fits with all the others as a measure of overall severity. This is illustrated in figure 3, which shows two very good items, but the one that assessed sleep disturbance was a better fit to the model of overall severity than the one about wheeze. Only by applying tests of this kind can the developer be sure that they have selected the very best and most appropriate items. A related methodology can be used to test whether men and women answer items differently or whether there are differences between countries/languages.

My focus on total scores should be set into context. Whilst most often the estimate that is required is the overall health status of a patient, it can be useful to look at domains or profile scores to understand in which way a treatment has most befitted the patient. A warning should be applied here, though. The domains of an instrument will contain fewer items than those that make up the total score (in some cases perhaps only three), which may make them generally less reliable/repeatable than the total scores. Also, much more effort is usually put
into validating overall scores than domains. Finally, there is a risk of “cherry picking” a favoured domain to illustrate a particular treatment effect. Generally speaking, total scores can be used for hypothesis testing, whereas domain scores are best used for exploring reasons for overall improvement and for hypothesis generation.

USE OF GENERIC HEALTH STATUS INSTRUMENTS IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Generic questionnaires were the first to be developed, originally as health surveys in general populations, but later they were applied to groups of patients with specific diseases. That then illustrated their weaknesses since they generally contain too few items for any specific questionnaire, which makes them relatively insensitive to differences between patients and to changes within patients. They now have rather limited application in COPD, with the exception of the EQ-5D, which is a generic measure that has a utility scale (http://www.euroqol.org/about-eq-5d.html). Such scales are weighted to estimate the patient’s preference for a given state and incorporate death as a health state. For that reason, health economists favour them for cost-utility analyses and the calculation of quality adjusted life-years (QALY), but they are otherwise too insensitive for most other purposes in COPD trials.

Figure 3. Item response curves obtained from Rasch analysis during development of the CAT. The plots show the relationship between overall severity (measured in logits) on the X axis and responses to a single item on the Y axis (as expressed in response to a 5-point category of response). The solid line indicates the theoretical relationship for item that has a perfect fit to the Rasch model, i.e. if it behaves exactly like all the other items combined. The dots indicate average measured responses from 1,503 patients. It can be seen that the measured responses to the item about sleep fits more closely to the line than the item about wheeze.
DISEASE-SPECIFIC MEASURES

Four disease-specific measures are in wide use for COPD. These will be described only briefly here since they are very well documented elsewhere.

Chronic Respiratory Questionnaire

The Chronic Respiratory Questionnaire (CRQ) was the first to be developed. It was originally developed as an interviewer-led instrument\(^6\), but later a self-administered version was created\(^7\). Its items were identified from patient interviews. It contains 20 items grouped into four domains: dyspnea, fatigue, emotional function, and mastery. It uses a seven-point equal interval category scale for responses. Its developer never created a total score, although CRQ users often report total scores by summing or averaging responses across domains. A systematic review of its performance characteristics has been published. It is widely used, particularly in the field of rehabilitation, and is available in several languages. The CRQ scores are used most often to describe changes rather than to characterise the severity of a population of patients. It is used largely for research and audit, particularly of rehabilitation programmes. No website appears to be available.

St George’s Respiratory Questionnaire

This self-administered instrument was developed and validated in asthma and COPD\(^5\), although until recently its use has largely been confined to COPD. It has 50 items derived from literature reviews and clinician input and grouped into three domains: symptoms, activity, and (psycho-social) impacts. Each item has an empirically derived weight. It was extensively validated when first developed and there is a very large number (> 700) of publications to be found in PubMed by typing “SGRQ”. After a number of years of use it was subjected to detailed statistical re-analysis to identify and remove less well-performing items. Since this was carried out using COPD data, this shorter 40-item set was termed the SGRQ for COPD (SGRQ-C)\(^9\). It is important to understand that there are no SGRQ-C scores; the scoring algorithm of the SGRQ-C creates SGRQ scores. The potential advantage of the shorter instrument is that the signal-to-noise ratio should be better. The other change with the SGRQ was to remove the recall period for the symptoms component.

The SGRQ is a research tool that is too complex for routine use. It is available in 63 language versions across 42 countries. Manuals and language versions for SGRQ and SGRQ-C are available on the website (www.healthstatus.sgul.ac.uk).

Clinical COPD Questionnaire

The Clinical COPD Questionnaire (CCQ) began life as the COPD Control Questionnaire and was constructed using items identified by patients, but item selection was based upon physician judgement about COPD control\(^9\). It has 10 items, each with a seven-point equal interval response scale. It now exists in a weekly as well as a daily version. It has been subjected to considerable validation and is available in 60 languages. It is suitable for routine use as well as in clinical trials. The website lists
a number of key publications and provides other useful details (http://www.ccq.nl).

**COPD Assessment Test**

The COPD Assessment Test (CAT) is the most recent questionnaire to be developed and was developed as a simple tool to facilitate communication between doctor and patient; however, it was developed to the standards set by the US Food and Drug Administration (FDA) for patients reported outcomes (PRO) so that it would be suitable for use in clinical trials\(^{10}\). It introduced Rasch modelling to questionnaire development in respiratory medicine, so despite its small number of items, it covers a range of core effects of COPD and the user can be confident that it includes items with good measurement properties\(^{11}\). It has shown good correlations with the established measures such as the CRQ\(^{12}\), SGRQ\(^{13}\), and CCQ\(^{14}\) and there are already a large number of publications, sufficient for two systematic reviews\(^{15,16}\). Recently it has been shown to be as sensitive as the SGRQ in detecting pharmacological treatment effects\(^{17}\). The website contains a user guide and allows online completion and scoring in 64 languages (www.CATestonline.org).

**Choice of questionnaire**

All of these instruments are well validated and can be recommended. The CRQ and SGRQ are mainly research instruments, and in recent years the SGRQ has been adopted by the pharmaceutical industry and accepted as an outcome measure for COPD trials by both the European Medicines Agency (EMA) and the FDA. Draft FDA guidance for use of the SGRQ is being developed. The CCQ and CAT are suitable for both routine use and research. In the context of clinical trials, the fact that the CAT was developed to meet FDA PRO development guidance may influence some users. The layout of the CAT is more “modern” and clear than the CCQ, but the latter has been used very successfully in many thousands of patients across the world. In terms of measurement properties, although the methodology used for CAT development was more sophisticated and rigorous than with any of the older instruments, there have been insufficient studies to determine whether it is superior to any of them.

**MINIMUM CLINICALLY IMPORTANT DIFFERENCE**

Minimum clinically important difference (MCID) is a poorly understood issue that has been reviewed recently in some depth\(^{18}\), so will not be discussed in detail here, but some points should be made. When health status questionnaires were first developed, no one had any idea what the numbers meant, so MCIDs were created to provide a point of reference. There is still no agreed standard methodology, although one might question whether an MCID based purely on statistical distributions, such as the half the standard deviation MCID, has any validity as a clinical reference point. Despite the fact that a number of methods are used, those that are clinically based usually show similar values; an excellent example of MCID estimation (for the CAT) can be found in Kon et al.\(^{19}\).

One recent area of confusion has been triggered by the pharmaceutical industry’s disappointment that large spirometric gains seen
when two bronchodilators of different classes are combined did not translate into commensurately large PRO gains. There has been talk of “MCIDs being developed with reference to placebo and not relevant to comparisons between two active drugs”. That is completely incorrect and reflects a lack of understanding of how MCIDs have been calculated. The issue of measuring the effect of adding one active treatment to another will be discussed in more detail later, but there is no scientific case for having separate MCIDs for different therapeutic settings.

More difficult is the issue of using different MCIDs for patients with different degrees of severity. As already discussed, a distinction should be drawn between size of change and consequence of the change. This brings in a discussion of what is used as reference to make the value judgment about whether a change or difference in score is clinically significant. No methodology has been proposed for addressing this question, but arguably for most patient populations and most treatments, this may not be important. The MCID is an average estimate (with a standard deviation) obtained in a group of patients. As long as the MCID was obtained in a group of patients similar to those being studied, it should be appropriate to that study population. In fact, there are very few research studies in very severe or very mild populations, and whilst it is scientifically interesting to test the hypothesis that they would need different MCIDs, there appears to be little practical need. In practice, the baseline SGRQ scores in different studies are usually very similar (Table 1).

The final issue around MCIDs concerns the way that scores are reported when reporting treatment effects. Traditionally, the group mean difference has been compared to the MCID and if the difference didn’t reach the MCID, the result was judged not clinically significant. That is a complete misjudgement. The change in PRO scores is usually normally distributed, so the mean and the median are the same. Under these circumstances, for a mean score to improve by more than the MCID, over half of the patients will have had to improve by more than the MCID. Where else in chronic disease is a treatment judged worthwhile only if more than half of the patients improve by a clinically significant amount? A better approach may be to report the number

<table>
<thead>
<tr>
<th>Drug</th>
<th>Author</th>
<th>Journal (year)</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indacaterol</td>
<td>Chapman21</td>
<td>Chest (2011)</td>
<td>46 ± 19</td>
</tr>
<tr>
<td>Tiotropium</td>
<td>Vincken22</td>
<td>European Respiratory Journal (2002)</td>
<td>44 ± 17</td>
</tr>
<tr>
<td>Glycopyrronium</td>
<td>D’Urzo23</td>
<td>Respir Res (2011)</td>
<td>40 ± 18</td>
</tr>
<tr>
<td>Aclidinium</td>
<td>D’Urzo24</td>
<td>Journal of COPD (2013)</td>
<td>47 ± 17</td>
</tr>
<tr>
<td>DVA</td>
<td>Vogelmeier29</td>
<td>Lancet Respir Med (2013)</td>
<td>42 ± 23</td>
</tr>
</tbody>
</table>
of patients who exceed the MCID. This is clinically appropriate because it identifies the additional proportion of patients who benefit, or expresses the result in terms of the increased probability of benefit, both of which clinicians can readily understand. The other advantage is that the responder ratio between treatment groups remains quite stable over a range of cut-off points. A recent study with the SGRQ showed this to be the case across a range of 2-6 units (four being the MCID); reducing the MCID will not improve the calculated treatment effect. One criticism of this approach is that it lumps all patients together as improved if they cross the MCID and may miss patients who improve a lot. However, carrying out a responder analysis for “super responders”, for example those who improve by twice the MCID, can mitigate this.

HEALTH STATUS IN POPULATIONS OF PATIENTS

There are relatively few data concerning disease-specific health status scores in healthy people. A study in working people in Japan showed median SGRQ scores below 10, with very few above 20 (Fig. 4). The same study showed median CAT scores to be around 5, with less than 25% of scores lying above 10. An important finding from that study was that there was very little difference in SGRQ or CAT score between non-smoking healthy subjects and those with airflow limitation, whether defined by the lower limit of normal or FEV1/FVC ratio. Surprising as this may seem, it was not a new observation since it can be found in a Spanish COPD screening study. Controls without COPD had a mean SGRQ score of 8.8, those with a previous diagnosis of COPD had a score of 34.3, whilst those identified by screening spirometry had an SGRQ score of 15.6. Thus it appears that COPD patients are diagnosed primarily because of their symptoms or, put another way around and more provocatively, these two studies suggest that there may not be large numbers of COPD patients who are suffering from untreated symptoms because they are undiagnosed.

As shown in table 1, the mean SGRQ score is very consistent across trials, which may suggest a recruitment bias. Mapping SGRQ to CAT score (permissible because of the consistently high correlation, typically > 0.8), the equivalent CAT score is approximately 18. This can be set against CAT scores obtained in a range of non-clinical trial settings, which are surprisingly similar in quite different countries (Fig. 5). In terms of health status,
patients recruited to clinical trials appear to be quite similar overall to those seen in clinical practice.

**EXACERBATIONS**

Patients who have more exacerbations have worse health status\(^{32}\), importantly unreported exacerbations may have an impact on health status that can be as large as that seen with reported events\(^{33-35}\). Patients who have more frequent treated exacerbations have a faster rate of deterioration of SGRQ score\(^{36}\), so studies on the long-term impact of untreated (and therefore untreated) exacerbations are needed. Use of the CAT has shown that with the onset of an exacerbation, the median change reported in three studies was 6.1 (range 4.7-6.6)\(^{15}\). One study reported a 9.9 unit change in CAT on recovery\(^{37}\), almost five times the individual patient MCID\(^{19}\). Recovery of health status from an exacerbation takes long time. Whilst symptoms may recover quickly\(^{38}\), in a study of 133 patients who had a moderate COPD exacerbation treated with antibiotics, with or without corticosteroids, a large improvement (more than twice the MCID) was seen over the first four weeks, but from week 4-26 the improvement was still almost twice the MCID\(^{39}\). This
continued improvement was seen in all three SGRQ domains. One study of exacerbation recovery used the CAT as a daily diary and appeared to show complete recovery by 12 days, but this is a misapplication of the CAT, which was not designed as a daily diary, but for use at intervals with patients who were not able to reference their previous scores.

The association between exacerbations and health status should also mean that health status scores predict exacerbations, which they do, independent of the effect of a prior history of exacerbations, although not as strongly as the latter. Recently however, a study performed in patients with a history of one or more exacerbations in the previous year showed that the CAT was a good predictor. For example, patients with a CAT score in the range 20-29 had a 70% greater risk of an exacerbation than those with CAT 0-9. Thus it appears that in a group of patients already at increased risk of exacerbations, the CAT score could identify those at relatively lower or higher risk.

**COMORBIDITIES**

Health status questionnaires are not diagnostic instruments so they cannot distinguish between COPD and chronic heart failure. Like all measurement instruments, they need to be used appropriately; for example FEV₁ alone cannot distinguish between COPD and asthma. A small effect of comorbidity on SGRQ has been reported, although a large study of COPD patients in primary care found that SGRQ and CAT scores were not different between patients with zero and 1-2 reported comorbidities, but both scores were slightly worse in patients with three or more. In contrast, the physical component score of the SF-12 (a general health score) was significantly worse in patients with 1-2 vs. no comorbidities and in patients with cardiac comorbidities compared to those with none. Thus it appears that pulmonary disease-specific health status scores may not be greatly affected by comorbidity. However, when used as part of the GOLD assessment scheme, patients categorised symptomatically using the CAT showed that there was a significantly increased risk of metabolic syndrome between GOLD Groups A, B, and D, in that order. The rate of coronary artery disease showed the same pattern, but the differences were not statistically significant; however, the number of cases was relatively small, giving rise to the possibility of a Type 2 statistical error. Rather than being influenced by comorbidity, health status scores may help identify patients at increased risk of cardiovascular disease.

**CLINICAL TRIALS**

Rehabilitation trials have often used the CRQ and shown that the average improvement exceeds the MCID by a wide margin. The SGRQ shows improvements of similar magnitude. This has been interpreted to show that rehabilitation has a large effect on health status and therefore the patient’s HRQoL, but these are at best single-blind studies so a bias may be present. A large number of randomised double-blind controlled trials of pharmacological therapy now report SGRQ data, and a recent network analysis of single-agent bronchodilator therapy shows that newer therapies appear to produce a bigger effect compared than older ones (Fig. 6). As noted earlier, the CAT appears to be as responsive as...
the SGRQ, whether in response to inhaled corticosteroid/long acting beta₂ agonist (ICS/LABA) combination, or long-acting anti-muscarinic (LAMA)/LABA combination.

The placebo response in health status score commonly seen in clinical trials is an interesting and important phenomenon. It may be rapid, within four weeks of patients being randomized to treatment, with no further improvement in SGRQ total score, or show little immediate improvement over the first two weeks, but a progressive and eventually large improvement over one year. The effect may be sustained for up to 30 months. It appears to be largely due to the Symptoms and Impacts domains of the SGRQ (Fig. 6), rather than Activity, a pattern that is seen in both clinical trials (TOwards a Revolution in COPD Health [TOC]) and in cohort studies. It is also greater in patients recruited from low-medium rather than high socioeconomic countries. The mechanisms are not
clear. In part it may be due to a placebo effect – i.e. the placebo makes the patients feel better; a regression to the mean effect due to the patients being more ill when recruited to a study; or a Hawthorne effect – a change in the behaviour of the patient (or clinician) when being observed, for example better adherence to concomitant treatment. Something similar may occur in routine practice but has not been explored.

DISEASE PROGRESSION

The Inhaled Steroids in Obstructive Lung Disease (ISOLDE) study showed that SGRQ scores worsened at a measureable rate and that this deterioration was faster in the Activity and Impacts domains than in the Symptoms domain\(^5^2\). Subsequently, the TORCH study showed that, even in placebo-treated patients, there was a maintained improvement in SGRQ Symptoms score (Fig. 6), whereas the Activity and Impacts domains showed progressive deterioration after 12 months. A very similar pattern was seen in an observational cohort study over five years\(^5^0\), the first follow-up measurement being made one year after the patients entered the study. This suggests that the improvement in Symptoms domain may not be due simply to a clinical trial artefact. Unfortunately, no details are available about which items in the Symptoms component showed this sustained improvement.

The TORCH study showed that worsening of SGRQ score was correlated with rate of FEV\(_1\) decline\(^5^3\); this was also seen in the Hokkaido cohort\(^5^0\). In the latter study, only patients classified as rapid decliners showed a clear deterioration in SGRQ score (total score 1.33 units/year), whereas those who maintained their FEV\(_1\) showed a small but statistically significant improvement (total score –0.94 units/year). There also appears to be an age effect. In the TORCH study, for a given deterioration in FEV\(_1\), the health status of younger patients (age < 65) deteriorated significantly more slowly than in older patients\(^5^3\).

In terms of treatment to prevent health status deterioration, the ISOLDE study suggested that the rate of worsening of SGRQ score might be altered by ICS\(^5^0\). Subsequently the Towards a Revolution in COPD Health (TORCH)\(^5^4\) and Understanding Potential Long-term Improvements in Function with Tiotropium (UPLIFT)\(^4^9\) studies showed that treatment with ICS/LABA or LAMA produced an improvement in health status that was maintained for three years or more. Setting aside the findings in ISOLDE, there is as yet no other direct evidence that treatment alters the measured rate of worsening (although clearly it must do that, in view of the sustained benefit seen in TORCH and UPLIFT trials). In part, this may be because differential drop out causes a healthy survivor effect in placebo-treated patients\(^5^5,5^6\). This dropout is “missing not at random”, so various approaches have been used to impute missing data, including multiple imputation or weighting using propensity scores. The problem is that they can take into account factors that occurred before withdrawal to predict changes and events that occurred after the patients withdrew, but whilst the dropout may have been informative and give an indication of future changes, some of these may be unpredictable.

Deterioration of SGRQ score may also be a predictor of future health events. In an analysis of data from the Evaluation of COPD Longitudinally to Identify Predictive Surrogate
Endpoints study (ECLIPSE) study, patients whose health status deteriorated over the first year were more likely to have an exacerbation, be admitted to hospital, or die than those whose SGRQ was unchanged or improved during the first year. This effect was most marked in younger patients (online supplement figures). This predictive effect was delayed (it took a further year for a difference in outcome to be seen) and suggests that rapidly worsening health status is reflecting greater disease, which then results in an increased risk of adverse events, not immediately but in the near-term (1-2 years).

**WHAT ADDED VALUE DOES HEALTH STATUS MEASUREMENT OFFER THE CLINICIAN?**

The advent of the CCQ and the CAT has made health status measurement available in routine practice, whether primary or secondary care. Despite its small number of items, the CAT is the most sophisticated of all the instruments available for COPD and there is a rapidly growing body of data about its performance. It also seems to behave in a very similar way to the SGRQ. So, what is the value of routine health status measurement?

The Global Initiative for Chronic Obstructive Lung Disease (GOLD) suggests that, as an aid to treatment decisions, a comprehensive measure of health status impairment is preferable to measurement of breathlessness alone. The need for such assessments is clear. The FEV1 is a very poor predictor of breathlessness, exercise capacity, and health status in individual patients and patients with COPD appear to underestimate the severity of their condition.

Clinical decisions should be informed by measurements and the better informed the clinician is, the better should be their decisions. This has been tested in a randomised controlled study using the CAT in a simulated primary care setting. The study had some design weaknesses and one of the clinical scenarios tested was inappropriate, but it did suggest that knowledge of the CAT score improved assessment of COPD issues. The CAT scores, coupled with other symptom assessments, are currently being used to guide treatment in a randomised controlled trial of step-up step-down therapy in COPD. The CAT and CCQ both have predictive validity for future adverse health events, but no studies in COPD have tested whether treatment decisions based on a prognostic score alter outcome. However, that might not matter since worse outcomes are associated with worse health status scores, and treatments directed at symptom improvement also improve outcomes such as exacerbations.

In practical terms, what is likely to be the greatest benefit of health status measurement in COPD? In my view, this will be in terms of monitoring. No measurement (whether FEV1, CAT or CCQ) has sufficient repeatability between assessments to reliably detect a symptomatic gain that still requires clinical assessment of benefit; however, repeated measurements over time are reliable within an individual. It is now well established that patients differ in terms of FEV1 decline and patients whose FEV1 declines faster have a faster deterioration in health status. Currently there are no long-term data about changes in CAT and CCQ over time, although these are now being collected. One study of a two time-point measurement of CAT scores shows clear
evidence of a regression to the mean effect so it has little value in this context\textsuperscript{62}. Mapping CAT scores from SGRQ scores in a number of longitudinal studies suggests that the average CAT change will be around +1 unit per year. Certainly a worsening of two or more units per year is an indication for exploring reasons for more rapid progression, especially in patients less than 70 years of age. Factors could include: faster loss of lung function (including progression of emphysema); frequent exacerbations that may be unreported; poor adherence with treatment; poor inhaler technique. There is also good evidence to support the use of health status measurement to establish recovery from an exacerbation (which, of course, requires a baseline measurement when stable before the event). If the patient recovers by less than four units on the CAT (even in non-hospitalised episodes), this requires careful re-evaluation and, in the absence of any obvious cause, points to the need for rehabilitation.

CONCLUSION

Health status measurements have an established role in clinical trials to measure clinical efficacy. The availability of shorter instruments brings the ability to measure health status into routine care. The CAT and CCQ are both easy to administer and take a maximum of three minutes to complete and score (unlike the FEV\textsubscript{1}) and provide important information that can aid clinical management. The FEV\textsubscript{1} is essential for diagnosing COPD, but thereafter its value diminishes. On the other hand, there is now a good argument for measurement of health status in routine care, since this is should to lead to an improved HRQoL for patients.

REFERENCES


