BMJ Open

Comparison between electronic and paper versions of patient-reported outcome measures in subjects with chronic obstructive pulmonary disease: an observational study with a cross-over administration

Koichi Nishimura,[1] Masaaki Kusunose,[1] Ryo Sanda,[1] Yousuke Tsuji,[2] Yoshinori Hasegawa,[3] Toru Oga[4]

For numbered affiliations see end of article.

Correspondence to
Dr Koichi Nishimura;
koichi-nishimura@nifty.com

ABSTRACT

Objectives A wide range of electronic devices can be used for data collection of patient-reported outcome (PRO) measures in subjects with chronic obstructive pulmonary disease (COPD). Although comparisons between electronic and paper-based PRO measures have been undertaken in asthmatics, it is currently uncertain whether electronic questionnaires work equally as well as paper versions in elderly subjects with COPD. The aim of this study was to compare the responses to paper and electronic versions of the Evaluating Respiratory Symptoms in COPD (E-RS) and the COPD Assessment Test (CAT).

Design A randomised cross-over design was used to compare the responses to paper and electronic versions of the two tools. The interval between the two administrations was 1 week.

Setting Electronic versions were self-administered under supervision using a tablet computer at our outpatient clinic (secondary care hospital in Japan) while paper questionnaires completed at home were requested to be returned by mail. It was intended that half of the patients completed the electronic versions of both questionnaires first, followed by the paper versions while the other half completed the paper versions first.

Participants Eighty-one subjects with stable COPD were included.

Results The E-RS total scores (possible range 0–40) were 6.8±7.4 and 5.0±6.6 in the paper-based and electronic versions, respectively, and the CAT scores (possible range 0–40) were 10.0±7.4 and 8.6±7.8. In both questionnaires, higher scores indicate worse status. The relationship between electronic and paper versions showed significant reliability for both the E-RS total score and CAT score (intraclass correlation coefficient=0.82 and 0.89, respectively; both $p<0.001$). However, both the E-RS total and CAT scores were significantly higher in the paper versions ($p<0.05$).

Conclusions In both cases, the two versions of the same questionnaire cannot be used interchangeably even though they have both been validated.

Strengths and limitations of this study

► A randomised cross-over design was used to compare the responses to paper and electronic versions of the Evaluating Respiratory Symptoms (E-RS) in chronic obstructive pulmonary disease (COPD) as well as the COPD Assessment Test (CAT) with a 1-week interval in elderly subjects with COPD.

► Since the CAT's 'original' mode was paper, while the E-RS' 'original' mode was electronic, the present study represents a bidirectional comparison of alternative administration modes rather than a validation of newer electronic versions of paper originals, which is generally the case.

► One of the limitations of our study design might be that the paper versions were completed unsupervised at home while the electronic versions were self-administered under supervision at our outpatient clinic.

► The result might have been influenced by the possible flaw of the study setting since the authors had expected the equivalence of the paper and electronic versions.

INTRODUCTION

Measuring patient-reported outcomes (PROs) has continued to gain importance in the healthcare sciences,[1–3] including respiratory medicine. Chronic obstructive pulmonary disease (COPD) is a major cause of mortality and morbidity globally and is the fourth-leading cause of death in the world.[4] Since many patients with COPD complain of dyspnoea and exertional intolerance, the condition has been one of the model diseases for measuring PROs, such as health-related quality of life.[5] For instance, the Chronic Respiratory Disease Questionnaire was the first published disease-specific tool for measuring

quality of life for subjects with COPD.[6] Although the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines have long recommended assessing disease severity based on the degree of airflow limitation, measured by forced expiratory volume in 1s (FEV_1), the GOLD consensus report has proposed an alternative classification system since 2011.[7] This system comprises a revised 'combined COPD assessment' classification in which symptoms should be assessed either as a dyspnoea measure using the modified Medical Research Council dyspnoea scale, or as a health status measure using the COPD Assessment Test (CAT).[8–11] This means that some of the PRO measuring tools have been recommended for use in daily clinical practice by the international guidelines. Thus, measures of PROs are currently considered to be essential in patient assessment and clinical research on subjects with COPD.

In the past, paper-based questionnaires played a central role in the data collection of PROs. However, electronic data capture might be preferable for daily diaries and an acceptable option for hospital-based surveys. This mode has been in common practice for a number of years in drug development studies.[12–15] Therefore, it is currently considered that a wide range of electronic devices can be used for the data collection of PROs in subjects with COPD also. Comparisons between electronic and paper-based PRO measures have been undertaken in several diseases and conditions including asthma.[16–24] The majority of the asthmatics in the previous studies were middle aged; however, it is possible that elderly subjects with COPD might be inexperienced in the use of electronic devices. Although there has been evidence that electronic questionnaires work equally as well as paper versions in older adults with Parkinson's disease and cancer,[19 22 23] less is known about subjects with COPD. The International Society for Pharmacoeconomics and Outcomes Research ePRO Good Research Practices Task Force Report has stated that an electronic PRO (e PRO).

questionnaire ought to produce data that are equivalent or superior (ie, higher reliability) to the data provided by the original paper version. Furthermore, measurement equivalence is a function of the comparability of the psychometric properties of the data obtained via the original and adapted administration mode,[12] and that a mixed-mode, electronic and paper-based trial requires a measurement equivalence established between modes.[15] Thus, possible bias between paper and electronic modes of data collection as well as equivalence between them is an important topic in the field of PRO research.

Leidy *et al* published the 14-item Exacerbations of Chronic Pulmonary Disease Tool (EXACT) PRO (known as EXACT-PRO) that is designed specifically to quantify exacerbations in COPD.[25–27] She and her colleagues subsequently reported that the Evaluating Respiratory Symptoms in COPD (E-RS) which uses the 11 respiratory symptom items contained in the 14-item EXACT is a reliable and valid instrument for E-RS severity in subjects with stable COPD, which included elderly persons.[28–30]

The EXACT is designed as a diary to be completed by the study participants on an electronic personal digital assistant, or a handheld electronic device similar in size to a smartphone[26 27] since paper data collection might have been considered to be flawed for a daily diary.[31] Therefore, the original developers of the EXACT recommended using electronic versions to obtain responses although the EXACT diary has also been administered in paper format, with results supporting its validity using this approach.[32 33]

While health status includes symptoms and impact, the EXACT is exclusively concerned with symptoms with the E-RS measuring just the respiratory symptoms. Although it might be commonly accepted that symptoms are one of the essential components of health status in subjects with COPD, the developers of the CAT and E-RS have stated that the two tools are derived from different conceptual frameworks. Therefore, the constructs being measured are different yet related.

On the other hand, the methods used to develop the two measures, both following rigorous instrument development approaches, include similar systematic item reduction. Both the CAT and E-RS are easy to administer in clinical practice due to Rasch analysis psychometric techniques.[9 25] Technically, the CAT is a questionnaire administered cross-sectionally or periodically while the E-RS is a diary, and both are instruments for outcome measures. Compared with the CAT which was designed for clinical use, the E-RS was designed for ease of use by study subjects. While it is less than ideal for clinical use as a daily diary, it could be useful as a quick symptom assessment in the clinic.

The aim of this study was to show equivalence of the responses to paper and electronic versions of the two most widely used questionnaires for patients with COPD. One is the CAT[8 9 11] and the other is the E-RS, which is designed to address the need for a standardised respiratory symptom diary.[28–30] The authors intended to compare the scores or individual item-level responses and ensure the overall comparability of the measurement properties between the paper and electronic versions of these two representative tools.

METHODS
Study design
A randomised cross-over design was used to compare the responses to paper and electronic versions of the E-RS and CAT. Electronic versions were self-administered under supervision using a tablet computer at the outpatient clinic while paper questionnaires completed at home were requested to be returned by mail. Half of the participants completed the electronic versions of both questionnaires first, followed by the paper versions, while the other half completed the paper versions first (figure 1). The interval between the two administrations was 1 week.
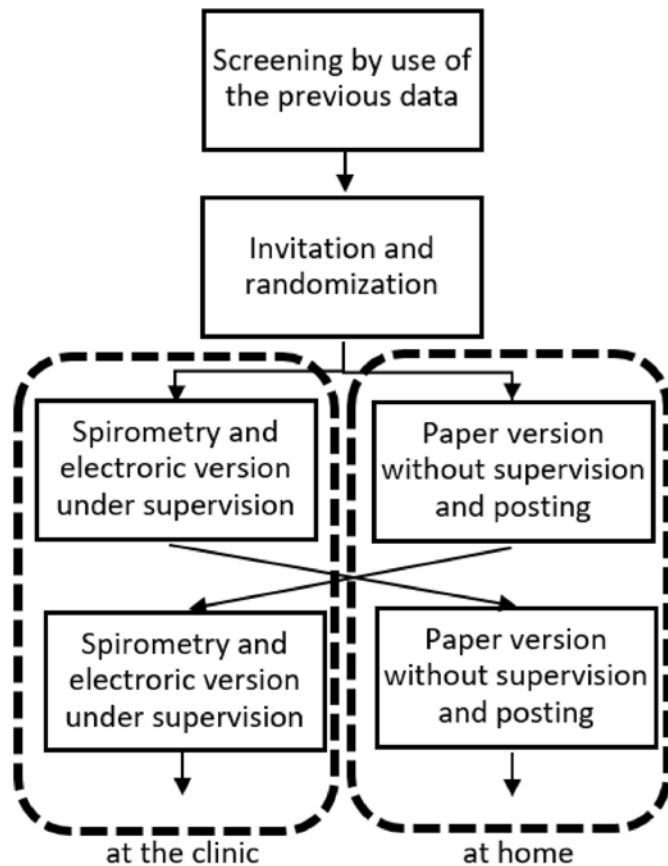
Figure 1 The study design.

Setting

After being screened according to previously collected spirometric data, patients with clinically stable COPD, as defined by the GOLD, were invited to participate during routine scheduled health assessment visits to our outpatient clinic in the Department of Respiratory Medicine at the National Center for Geriatrics and Gerontology. If randomised to the paper version first, they were first given the measure with instructions during the visit, and asked to take it home and return it by mail. They then visited the clinic for the electronic administration a week later. The instructions regarding when to answer the questions and how to return by mail were provided together with a reply envelope. Supervisors at the clinic opened the first page of the electronic version and handed the tablet to the participants. They showed them how to select their desired options, input the answer to the examination question and proceed to the next page. After the participants completed the last page, they returned the tablet to the supervisors. The change of global health between the first and second administration was also assessed using a five-point scale, and they were analysed at a time when they were considered to be clinically stable. Participants whose questionnaires contained incomplete or inappropriate responses such as missing items and double checks were excluded from the analysis. In this situation, participants who had dropped out once were invited to take part again according to the study design.

Participants

Criteria for inclusion in the present study were: (1) age over 50 years; (2) smoking history of more than 10 pack-years; (3) postbronchodilator FEV_1/forced vital capacity (FVC) ratio of less than 0.7; (4) no obvious abnormal shadows on chest X-rays which could influence lung function; (5) absence of any other active lung disease; (6) absence of uncontrolled comorbidity and (7) no changes in treatment regimen during the preceding 4 weeks. Exclusion criteria were: (1) a history of asthma and (2) an exacerbation of COPD within the preceding 3 months. All patients had more than 3 months of outpatient management before entry into the study to avoid any subsequent changes caused by new medical interventions.

The participants underwent spirometry using a spirometer (CHESTAC-8800; Chest, Tokyo, Japan) prior to entering their responses into a tablet at the clinic. According to the method described by the American Thoracic Society (ATS)/ European Respiratory Society (ERS) Task Force in 2005,[34] three acceptable spirometric flow-volume curves were recorded with the participant in a sitting position. The highest FEV_1 and the highest FVC values from three attempts were then analysed. The predicted values for FEV_1 and vital capacity were calculated according to the proposal from the Japan Respiratory Society.

Patient-reported outcomes

Disease-specific health status was assessed with a previously validated Japanese version of the CAT.[35] The CAT was originally developed as a paper-based questionnaire consisting of eight items scored from 0 to 5 in relation to cough, sputum, dyspnoea, chest tightness, capacity for exercise and activities, sleep quality and energy levels.[8 9 11] The CAT scores range from 0 to 40, with a score of 0 indicating no impairment.

Although the E-RS includes just the 11 respiratory symptom items from the 14-item EXACT, the entire EXACT was administered in the present study. Scores on the E-RS range from 0 to 40, with higher scores indicating more severe symptoms.[28–30] The RS-Total score represents overall respiratory symptom severity. Three subscales are used to assess breathlessness (RS-breathlessness), cough and sputum (RS-cough and sputum), and chest-related symptoms (RS-chest symptoms). The recall period was 'today' and patients selected the answers that best described their experiences for that day. The Japanese translation was created and provided by the original developers. The present survey was conducted using a paper-based questionnaire or a questionnaire on a tablet computer, with no knowledge of their own previous responses, that is, without informed administration.

The EXACT was originally designed for electronic administration, and the paper version was provided by the developers for use in English. The paper version of the Japanese EXACT was administered in a three-page booklet where six questions are listed on the first page, five on the second page and three on third. The

participants can see all of their responses on the paper versions while they can only see some of their responses when completing the measures electronically. This is a limitation as the two modes of administration do not provide the same experience to the user.

Electronic devices

The electronic Japanese versions of the E-RS and CAT were developed on a non-profit basis by technical staff at Hoshi Iryo-Sanki. Both electronic versions were designed to be accessed via a hand-held tablet computer (17 cm x 24 cm, iPad). The item questions and response options are identical to the paper versions of the questionnaires. A large font-size is used to enhance clarity and readability. As for the electronic version of the CAT, the first three questions are displayed on the screen. When the responses to the three questions have been input, they scroll upwards and are replaced by the next three questions. As for the electronic version of the EXACT, each question is displayed on a separate screen with the following page automatically appearing after the answer options for the current question have been entered. It is neither possible to move from one item to the next without answering the question nor to choose two answers for the same question. The programme also allows the user to correct or change previous answers by using the 'back' button. A summary screen showing the calculated scores appears following the final question and there is one additional 'Thank You' screen on completion of the questionnaire. Therefore, the questionnaires comprise 16 screens in the EXACT and three screens in the CAT. Although the copyright holder of the EXACT has published the EXACT e-Diary Certification Programme on its website, the electronic version we used in the present study unfortunately did not meet the criteria in their programme, since the electronic version was individually administered only once in the present study to get an answer for the E-RS but not for the EXACT-PRO diary, and this certification programme had not been included in the licence agreement.

Patient and public involvement

Patients and the public were neither involved in the development of the research question, the design of this study, nor the recruitment to and conduct of the study.

Written informed consent was obtained from all patients before the study. The authors received permission to use the Japanese EXACT in the study entitled 'A validation study of an electric version of the Japanese EXACT', which aims to 'develop an electric version of the Japanese EXACT'.

Statistical methods

All results are expressed as means±SD. Calculating Cronbach's coefficient alpha enabled us to assess the internal consistency. The score distribution of the PROs was evaluated by histograms and the Shapiro-Wilk test. Relationships between two sets of data were analysed by Spearman's rank correlation tests. Concordance between the

two methods was examined by intraclass correlation coefficient (ICC) analysis. The significance of between-group differences was determined by the Wilcoxon-signed rank test. A $p < 0.05$ was considered to be statistically significant. Statistical analyses were performed using Bell Curve for Excel (Social Survey Research Information).

RESULTS

Of the 130 participants initially enrolled, data obtained from 49 were excluded from the analysis for the following reasons: inappropriate answers in the paper version with missing items or multiple answers in 15, wrong completion date entered by participants in 10, non-return of the paper questionnaire in 7 and other violations in 17. Finally, 81 patients with stable COPD were included in this analysis. The majority were male (87.9%), the average age and FEV_1 were 75.6±5.9 years and 1.71±0.58 L (70.1±21.5%pred) (table 1). Fifty subjects completed the electronic versions first, followed by the paper versions, while the remaining 31 participants completed the paper versions first. The patient characteristics including age, sex and disease severity were not significantly different between these two study groups.

The frequency distribution histograms of the scores obtained from each instrument are shown in figure 2. The normality of the score distributions of the CAT and E-RS scores obtained from the electronic and paper versions was rejected using the Shapiro-Wilk test ($p < 0.001$, all). The E-RS total and CAT scores were both skewed toward the milder end of the respective scales. The best possible score ('floor effect') on the E-RS was noted in 17 subjects (21.0%) in the paper version and 22 (27.2%) in the electronic version. For the CAT, it was five subjects (6.2%) in the paper version and 7 (8.6%) in the electronic version (table 2).

The internal consistency of each questionnaire was assessed with Cronbach's alpha coefficient (table 2). The internal consistency of the CAT and E-RS total scores obtained from the electronic and paper versions was high (Cronbach's coefficient α=0.92–0.94). The internal consistency of Breathlessness and Chest Symptoms scores of the E-RS (RS-breathlessness and RS-chest symptoms) was also high (alpha=0.90–0.95) and that of the RS-cough and sputum ranged from alpha=0.76 to alpha=0.81 regardless of version used. The E-RS total scores (possible range 0–40) were 6.8±7.4 and 5.0±6.6 in the paper-based and electronic versions, respectively, and the CAT scores (possible range 0–40) were 10.0±7.4 and 8.6±7.8 (table 2). The relationship between the electronic and paper versions showed significant reliability in both the E-RS total score and CAT score (ICC=0.82 and 0.89, respectively; both $p < 0.001$) (table 3). However, both the E-RS total and CAT scores were significantly higher in the paper version ($p < 0.05$) (table 3). Correlation coefficients of E-RS total and CAT scores obtained from the paper and electronic versions together with other clinical variables are shown in table 1. Physiological measures had

Table 1 Demographic details and correlations with the E-RS total and CAT scores obtained from paper and electronic versions (Spearman's rank correlation coefficient)

Variable	Units	Mean	SD	Correlations with E-RS total score obtained from		Correlations with CAT score obtained from	
				Paper version	Electronic version	Paper version	Electronic version
Age	years	75.6	5.9	0.00	−0.12	0.02	−0.04
BMI	kg/m^2	22.1	3.0	0.04	−0.05	−0.04	−0.04
Cumulative smoking	pack-years	54.6	31.7	0.15	0.09	0.13	0.12
FVC	Litres	3.06	0.75	−0.21	−0.18	−0.30*	−0.32*
FVC	% pred	99.0	21.1	−0.30*	−0.25*	−0.30*	−0.37*
FEV$_1$	Litres	1.71	0.58	−0.42*	−0.39*	−0.50*	−0.44*
FEV$_1$	% pred	70.1	21.5	−0.47*	−0.45*	−0.49*	−0.46*
FEV$_1$/FVC	%	55.3	11.2	−0.45*	−0.42*	−0.44*	−0.29*
CAT score obtained from paper version		10.0	7.4	0.81*	0.64*	–	–
CAT score obtained from electronic version		8.6	7.8	0.76*	0.72*	–	–
E-RS total score obtained from paper version		6.8	7.4	–	–	0.81*	0.76*
E-RS total score obtained from electronic version		5.0	6.6	–	–	0.64*	0.72*

*P<0.05.
BMI, body mass index; CAT, the COPD Assessment Test; E-RS, the Evaluating Respiratory Symptoms in COPD; FEV$_1$, forced expiratory volume in 1 s; FVC, forced vital capacity.

weak or modest correlations with E-RS total and CAT scores except FVC and E-RS total score. Almost all of the Spearman's rank correlation coefficients were similar between the paper and electronic versions. The E-RS total scores were also well correlated with the CAT scores both in the paper (Spearman's rank correlation coefficient (Rs), 0.81) and electronic versions (Rs, 0.72).
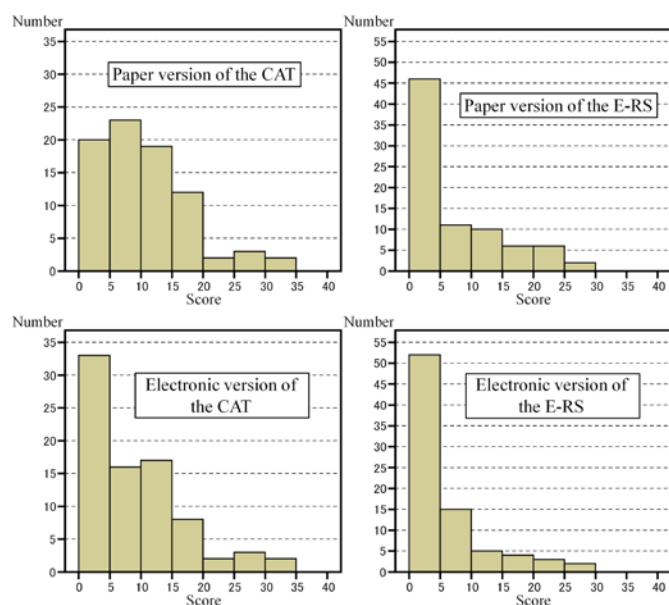

Figure 2 The frequency distribution histograms of responses to paper and electronic versions of the Evaluating Respiratory Symptoms in COPD (E–RS) total and the COPD Assessment Test (CAT) scores. COPD, chronic obstructive pulmonary disease.

DISCUSSION
Strengths
This is the first cross-sectional study to directly compare responses to the electronic and paper-based versions of the E-RS and CAT in elderly subjects with COPD. Two disease-specific tools that are easy to administer were examined in the present study. While the CAT was originally developed as a paper questionnaire to measure health status, the E-RS was specifically developed as an electronic measuring tool to measure the respiratory symptoms. First, agreement between paper and electronic versions, evaluated using an ICC were 0.82 for the E-RS total score and 0.89 for the CAT score. The correlations between the versions of both tools were moderate to strong. Therefore, the relationship of the overall scores between electronic and paper versions might be acceptable. Second, both the E-RS Total and CAT scores were significantly higher in the paper versions. It is unclear whether the electronic version of the CAT is underestimating health status compared with its paper counterpart, or the paper version is overestimating. Similarly, for the E-RS, it is unclear if the paper version is overestimating respiratory symptoms or the electronic version is underestimating. In a comparison of the score distribution between the paper and electronic versions, there was a slight skew towards the mild end of the scale in the electronic version of both tools. The possible difference could be explained by the skewed score of the electronic versions of the E-RS and CAT. The mean difference observed between modes of administration for the CAT score was 1.4 in the present study, which is smaller than the value reported to be the

Table 2 Internal consistency and score distribution

Patient-reported outcomes	Possible score range	Items (n)	Version	Cronbach's α coefficient	Score distribution Mean	SD	Median	Max	Min	Floor effect, %
E-RS total score	0–40	11	Paper	0.94	6.8	7.4	3	27	0	21.0
			Electronic	0.94	5.0	6.6	3	28	0	27.2
RS-breathlessness	0–17	5	Paper	0.95	3.1	4.3	0	15	0	51.9
			Electronic	0.91	2.3	3.4	0	14	0	51.9
RS-cough and sputum	0–11	3	Paper	0.76	1.9	1.8	2	6	0	29.6
			Electronic	0.81	1.5	1.8	1	8	0	40.7
RS-chest symptoms	0–12	3	Paper	0.91	1.7	2.3	0	8	0	50.6
			Electronic	0.90	1.2	2.0	0	9	0	60.5
CAT score	0–40	8	Paper	0.92	10.0	7.4	9	33	0	6.2
			Electronic	0.94	8.6	7.8	7	31	0	8.6

In both questionnaires, higher scores indicate worse status.
CAT, the COPD Assessment Test; E-RS, the Evaluating Respiratory Symptoms in COPD.

minimum clinically important difference of two points.[36] This suggests that there was no clinically meaningful difference. However, consequently, the authors might fail to demonstrate the equivalence of the measurement properties between the paper and electronic versions.

To the best of our knowledge, in asthmatics, a total of six studies, which compared the responses to electronic and paper versions of PRO measures, have been published.[16–18 20 21 24] Two of these studies were conducted to compare asthma diary completion.[16 20] The other four compared responses to paper-based and electronic versions of generic or disease-specific quality of life questionnaires.[17 18 21 24] Olajos-Clow et al reported that agreement between electronic and paper versions of the Mini Asthma Quality of Life Questionnaire for the overall score was acceptable with no bias but that a small but significant bias was noted in the activity limitation domain, and that generalisability might be limited in the young (12–17 years) and older (>65 years) adults.[24] Juniper et al examined paper and electronic versions of the Asthma Quality of Life Questionnaire, the Asthma Control Questionnaire and the Rhinoconjunctivitis Quality of Life Questionnaire and reported that the significant bias and only modest concordance found provided evidence that

patients might respond differently to questionnaires in different formats.[21] Our findings on subjects with COPD were similar to the previous investigations on asthmatics. Since COPD typically affects the aged and it has been pointed out that the average age of Japanese patients with COPD is higher,[37] the present finding might illustrate a growing problem in Japan.

Which is the better measurement tool for elderly subjects with COPD, a paper or electronic questionnaire? In the case of the CAT and E-RS, the CAT's 'original' mode was paper, while the E-RS' 'original' mode was electronic. So the present study represents a comparison of alternative administration modes rather than a validation of newer electronic versions of paper originals, which is generally the case. The authors have demonstrated score reliability (internal consistency and reproducibility) of each measure across modes, and construct validity of each measure and mode through correlations between measures although there is a small difference in the score distribution between the two versions in each case. However, electronic versions have advantages that should not be overlooked. While missing items and multiple checks are considered to be an inevitable and unavoidable consequence in paper-based questionnaires, these

Table 3 Relationships and differences between paper and electronic versions of the E-RS and CAT

	Spearman's rank correlation coefficient	Intraclass correlation coefficient	P value from Wilcoxon-signed rank test
E-RS total score	0.76 (p<0.001)	0.82 (p<0.001)	<0.001
RS-breathlessness	0.79 (p<0.001)	0.79 (p<0.001)	0.0066
RS-cough and sputum	0.67 (p<0.001)	0.71 (p<0.001)	0.012
RS-chest symptoms	0.66 (p<0.001)	0.75 (p<0.001)	0.0036
CAT score	0.84 (p<0.001)	0.89 (p<0.001)	<0.001

CAT, the COPD Assessment Test; E-RS, the Evaluating Respiratory Symptoms in COPD.

issues can be eliminated in electronic versions. In fact, we also found that, while almost one-third of paper questionnaires collected were excluded from analysis due to the aforementioned problems, the electronic versions were largely problem free.

Limitations

One of the main limitations of our study design is that the paper versions were completed unsupervised at home and this might have had an effect on the results. Since missing items and incorrectly completed questionnaires might be inevitable in paper-based versions, the presence or absence of a supervisor and the site, at the clinic or at home should be adequately acknowledged in the present study. This might explain the possible imbalance between the two administrations with 50 subjects completing the electronic versions first and 31 participants completing the paper versions first.

Second, since it was difficult to offer an ideal situation for the simultaneous administration of both the CAT and E-RS, the result might have been influenced by the possible flaw of the study setting. Since the authors had expected the equivalence of the paper and electronic versions, this matter in the study design had not been considered important. For example, the CAT should be answered at the clinic under supervision while the EXACT diary should be completed at home before going to bed. Although some researchers advocate the use of electronic methods of data collection to ensure that data are captured as per the study protocol with unsupervised data collection, electronic modes were given with guidance from supervisors in the present study. The study interval may be also controversial since the recall period is likely to be several weeks in the CAT and a day in the E-RS that is originally derived from a diary. Furthermore, the use of an electronic data-capturing device in Japanese which has not been passed by the EXACT e-Diary Certification Programme might have undermined the effectiveness of the E-RS.

Third, there might be concerns about the possibility of selection bias and generalisations of these results might not be warranted. We recruited only patients who could attend our outpatient clinic on a regular basis. It is likely that we did not include sufficient numbers of those patients without any subjective symptoms who were unaware of having COPD, or patients who could not regularly attend our clinic due to heavy physical burden. This single-centre study was also limited by the small number of participants and distinct male preponderance of the subjects, even though it contains most of the patients with stable COPD seen in our hospital during the study period. Our study included predominantly men since numbers of women with COPD were, in fact, quite low in Japan at the time. Thus, the study reflected the reality of clinical COPD in our population.

Conclusions

Three main conclusions can be drawn from our findings. The first is the consistent reliability across modes of administration for the two measures. Internal consistency levels were high and correlations between the two modes (reproducibility) were also high. The relationships between the total scores for the electronic and paper versions of both tools showed significant reliability. Second, however, both the E-RS Total and CAT scores were significantly higher in the paper versions. There were significant, systematic score differences between modes that might be due to the measures. Third, in the case of both the E-RS and the CAT, paper and electronic versions of the same questionnaire cannot, therefore, be used interchangeably even though both versions of each tool have been validated.

Author affiliations
[1]Department of Respiratory Medicine, National Center for Geriatrics and Gerontology, Obu, Aichi, Japan
[2]Hoshi Iryo-Sanki Co. Ltd, Tokyo, Japan
[3]National Hospital Organization, Nagoya Medical Center, Nagoya, Japan
[4]Department of Respiratory Medicine, Kawasaki Medical School, Kurashiki, Okayama, Japan

REFERENCES

1. DeMuro C, Clark M, Doward L, *et al*. Assessment of pro label claims granted by the FDA as compared to the EMA (2006-2010). *Value Health* 2013;16:1150–5.
2. Gnanasakthy A, Mordin M, Evans E, *et al*. A review of patient-reported outcome labeling in the United States (2011-2015). *Value Health* 2017;20:420–9.
3. Kyte D, Duffy H, Fletcher B, *et al*. Systematic evaluation of the patient-reported outcome (pro) content of clinical trial protocols. *PLoS One* 2014;9:e110229.
4. Vogelmeier CF, Criner GJ, Martinez FJ, *et al*. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. gold executive summary. *Am J Respir Crit Care Med* 2017;195:557–82.
5. Ekström M, Sundh J, Larsson K. Patient reported outcome measures in chronic obstructive pulmonary disease: which to use? *Expert Rev Respir Med* 2016;10:351–62.
6. Guyatt GH, Berman LB, Townsend M, *et al*. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987;42:773–8.

7 Vestbo J, Hurd SS, Agustí AG, *et al*. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013;187:347–65.

8 Jones PW, Brusselle G, Dal Negro RW, *et al*. Properties of the COPD assessment test in a cross-sectional European study. *Eur Respir J* 2011;38:29–35.

9 Jones PW, Harding G, Berry P, *et al*. Development and first validation of the COPD assessment test. *Eur Respir J* 2009;34:648–54.

10 Jones PW, Price D, van der Molen T. Role of clinical questionnaires in optimizing everyday care of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2011;6:289–96.

11 Jones PW, Tabberer M, Chen W-H. Creating scenarios of the impact of COPD and their relationship to COPD assessment test (CAT™) scores. *BMC Pulm Med* 2011;11:42.

12 Coons SJ, Gwaltney CJ, Hays RD, *et al*. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (pro) measures: ISPOR ePRO good research practices Task force report. *Value Health* 2009;12:419–29.

13 Ganser AL, Raymond SA, Pearson JD. Data quality and power in clinical trials: a comparison of ePRO and paper in a randomized trial. In: Byrom B, Tiplady B, eds. *ePro electronic solutions for patient-reported data*. Surrey: Gower Publishing Limited, 2010: 49–77.

14 Coons SJ, Eremenco S, Lundy JJ, *et al*. Capturing patient-reported outcome (pro) data electronically: the past, present, and promise of ePRO measurement in clinical trials. *Patient* 2015;8:301–9.

15 Eremenco S, Coons SJ, Paty J, *et al*. Pro data collection in clinical trials using mixed modes: report of the ISPOR pro mixed modes good research practices Task force. *Value Health* 2014;17:501–16.

16 Berg J, Dunbar-Jacob J, Rohay JM. Compliance with inhaled medications: the relationship between diary and electronic monitor. *Ann Behav Med* 1998;20:36–8.

17 Bushnell DM, Martin ML, Parasuraman B. Electronic versus paper questionnaires: a further comparison in persons with asthma. *J Asthma* 2003;40:751–62.

18 Caro JJ, Caro I, Caro J, *et al*. Does electronic implementation of questionnaires used in asthma alter responses compared to paper implementation? *Qual Life Res* 2001;10:683–91.

19 Hurria A, Mohile S, Gajra A, *et al*. Validation of a prediction tool for chemotherapy toxicity in older adults with cancer. *J Clin Oncol* 2016;34:2366–71.

20 Hyland ME, Kenyon CA, Allen R, *et al*. Diary keeping in asthma: comparison of written and electronic methods. *BMJ* 1993;306:487–9.

21 Juniper EF, Langlands JM, Juniper BA. Patients may respond differently to paper and electronic versions of the same questionnaires. *Respir Med* 2009;103:932–4.

22 Morley D, Dummett S, Kelly L, *et al*. An electronic version of the PDQ-39: acceptability to respondents and assessment of alternative response formats. *J Parkinsons Dis* 2014;4:467–72.

23 Morley D, Dummett S, Kelly L, *et al*. Evaluating the psychometric properties of an e-based version of the 39-item Parkinson's disease questionnaire. *Health Qual Life Outcomes* 2015;13:5.

24 Olajos-Clow J, Minard J, Szpiro K, *et al*. Validation of an electronic version of the mini asthma quality of life questionnaire. *Respir Med* 2010;104:658–67.

25 Jones PW, Chen W-H, Wilcox TK, *et al*. Characterizing and quantifying the symptomatic features of COPD exacerbations. *Chest* 2011;139:1388–94.

26 Leidy NK, Wilcox TK, Jones PW, *et al*. Development of the exacerbations of chronic obstructive pulmonary disease tool (exact): a patient-reported outcome (pro) measure. *Value Health* 2010;13:965–75.

27 Leidy NK, Wilcox TK, Jones PW, *et al*. Standardizing measurement of chronic obstructive pulmonary disease exacerbations. reliability and validity of a patient-reported diary. *Am J Respir Crit Care Med* 2011;183:323–9.

28 Leidy NK, Murray LT, Monz BU, *et al*. Measuring respiratory symptoms of COPD: performance of the EXACT- respiratory symptoms tool (E-RS) in three clinical trials. *Respir Res* 2014;15:124.

29 Leidy NK, Sexton CC, Jones PW, *et al*. Measuring respiratory symptoms in clinical trials of COPD: reliability and validity of a daily diary. *Thorax* 2014;69:424–30.

30 Miravitlles M, Chapman KR, Chuecos F, *et al*. The efficacy of aclidinium/formoterol on lung function and symptoms in patients with COPD categorized by symptom status: a pooled analysis. *Int J Chron Obstruct Pulmon Dis* 2016;11:2041–53.

31 Stone AA, Shiffman S, Schwartz JE, *et al*. Patient non-compliance with paper diaries. *BMJ* 2002;324:1193–4.

32 Ehsan M, Khan R, Wakefield D, *et al*. A longitudinal study evaluating the effect of exacerbations on physical activity in patients with chronic obstructive pulmonary disease. *Ann Am Thorac Soc* 2013;10:559–64.

33 Vanaparthy R, Mota P, Khan R, *et al*. A longitudinal assessment of sleep variables during exacerbations of chronic obstructive pulmonary disease. *Chron Respir Dis* 2015;12:299–304.

34 Miller MR, Hankinson J, Brusasco V, *et al*. Standardisation of spirometry. *Eur Respir J* 2005;26:319–38.

35 Tsuda T, Suematsu R, Kamohara K, *et al*. Development of the Japanese version of the COPD assessment test. *Respir Investig* 2012;50:34–9.

36 Kon SSC, Canavan JL, Jones SE, *et al*. Minimum clinically important difference for the COPD assessment test: a prospective analysis. *Lancet Respir Med* 2014;2:195–203.

37 Fukuchi Y, Nishimura M, Ichinose M, *et al*. Copd in Japan: the nippon COPD epidemiology study. *Respirology* 2004;9:458–65.