Reading numbers aloud: a measure of the limiting effect of breathlessness in patients with cancer

Andrew Wilcock, Vincent Crosby, Debra Clarke, Ray Corcoran, Anne Tattersfield

Hayward House Macmillan Specialist Palliative Care Unit, Nottingham City Hospital NHS Trust, Nottingham NG5 1PB, UK
A Wilcock
V Crosby
D Clarke
R Corcoran

Department of Respiratory Medicine
A Tattersfield

Correspondence to:
Dr A Wilcock.

Abstract

Background—Progress in the treatment of breathlessness at rest or on minimum exertion in patients with cancer requires a practical and valid method of measuring symptoms. A study was undertaken to explore the practicality, repeatability, and sensitivity of reading numbers as a form of exercise test in this group of patients.

Methods—Thirty patients with cancer and 30 age matched healthy subjects read numbers aloud as quickly and clearly as they could for 60 seconds. After five readings the maximum number of numbers read and the number read per breath was noted. This procedure was carried out twice in one day and one week later to assess within and between day repeatability. The sensitivity of the test was assessed by making measurements in 13 patients with cancer before and after drainage of their pleural effusion.

Results—The concept was easily understood by all subjects. Twelve patients were unable to complete five readings in all tests due to tiredness. Compared with control subjects patients read fewer numbers in the three tests (87–89% of control) and fewer numbers per breath (59–60% of control). Repeatability was good both within and between days. After drainage of their effusion all patients were less breathless and there was an increase in both the maximum number of numbers read (23%) and the number read per breath (60%).

Conclusions—The number of numbers read and the number read per breath over 60 seconds was practical, easy to carry out, showed good repeatability within and between days and was sensitive to the improvement seen following drainage of a pleural effusion. It may be a useful measure of the limiting effect of breathlessness in this group of patients.

(*Thorax* 1999;**54**:1099–1103)

The treatment of breathlessness in patients with cancer who are short of breath at rest or on minimum exertion is inadequate. New approaches are required and need to be assessed. This requires a quantitative method of assessing breathlessness in these patients since exercise testing on a cycle or treadmill is usually impractical and the repeatability of numerical or visual analogue scales is such that relatively large numbers of patients are necessary for studies of sufficient power.[1] [2] We have therefore explored the use of reading numbers aloud as a form of exercise test to measure the limiting effect of breathlessness in patients who are breathless at rest or on minimum exertion. We have examined the repeatability of the maximum number of numbers that can be read aloud in one minute and the number of numbers read per breath by healthy volunteers and patients with breathlessness due to cancer. The sensitivity of both measures was assessed by making measurements in patients with cancer before and after drainage of their pleural effusions.

Methods

SUBJECTS

All patients were recruited from respiratory or oncology outpatient clinics and from a palliative care unit (Hayward House). Their physical condition, medication, and times of administration remained unchanged during the study. None was cognitively impaired or limited by dysarthria. All subjects gave verbal informed

|									
3	47	43	73	86	36	96	47	36	61
97	74	24	67	62	42	81	14	57	20
16	76	62	27	66	56	50	26	71	7
12	56	85	99	26	96	10	68	27	31
55	59	56	35	64	38	54	82	46	22
16	22	77	94	39	86	49	43	54	82
84	42	17	53	31	57	24	55	6	88
63	1	63	78	59	16	95	55	67	19
33	21	12	34	29	78	64	56	7	82
57	60	86	32	44	9	47	27	96	54
18	18	7	92	46	44	17	16	58	9
26	62	38	97	75	84	16	7	44	99
23	42	40	64	74	82	97	77	54	81
52	36	28	19	95	50	92	26	11	97
37	85	94	35	12	83	39	50	8	30

Figure 1   The grid of numbers
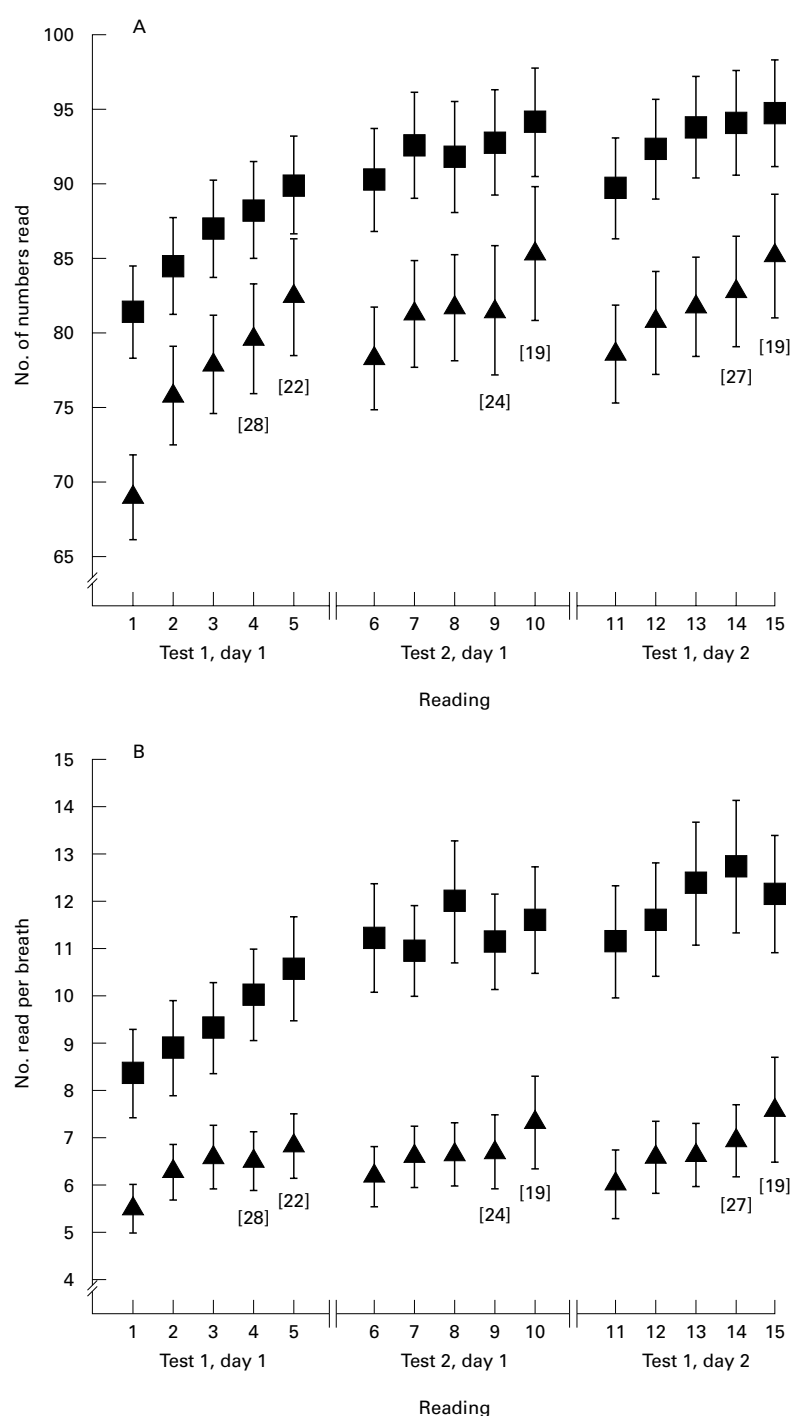
Figure 2   Mean (SE) number read (A) and number read per breath (B) for control subjects (squares) and patients (triangles). Numbers in square brackets denote size of sample if less than 30.

consent and the study was approved by the Nottingham City Hospital ethics committee.

Repeatability study
Thirty healthy control subjects (20 women; mean age 62 years) were recruited from staff or volunteers at Hayward House. None smoked, had respiratory disease, or was limited by breathlessness. Control subjects were chosen to be in the same age range as the patients.

Thirty patients (nine women; mean age 68 years) with evidence of primary or secondary lung cancer were recruited; 24 patients had carcinoma of the lung, two carcinoma of the breast, and the remaining four had malignant thymoma, mesothelioma, melanoma, or sar-

coma. All complained of increasing breathlessness that limited their daily activities since the development of cancer. Eighteen patients also had obstructive airways disease and two had ischaemic heart disease.

Sensitivity study
Thirteen patients (eight women; mean age 69 years) with a pleural effusion due to primary or secondary lung cancer were recruited. Three patients had carcinoma of the lung and three had mesothelioma, the remainder having secondary lung cancer originating from carcinoma of the breast (3), ovary (2), bladder (1), or prostate (1). Two patients also had obstructive airways disease and one each had asthma, valvular heart disease, and cardiac failure. All had breathlessness at rest (7) or on minimum exertion (6) and they underwent pleural aspiration (6) or drainage (7). Their physical condition otherwise remained stable over the period of the study.

MEASUREMENTS
Spirometric tests
Forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC) were measured with the subjects standing as the best of three recordings (Vitalograph, Buckingham, UK) in the repeatability study and as the best of three recordings within 100 ml using a hand held spirometer (Micro Spirometer, Micro Medical Ltd, Rochester, UK) in the sensitivity study.

Numbers reading test
While seated the subjects were given a page containing a grid of numbers (fig 1) and asked to read the numbers aloud and in order as quickly and as clearly as they could. The number of breaths taken and the number of numbers read after 60 seconds was recorded. The procedure was repeated five times using the same grid of numbers each time and the maximum values achieved over 60 seconds was noted. Subjects were allowed to recover between readings and continued when they felt rested.

Breathlessness
Patients were asked to rate their worst and average breathlessness and the degree of trouble or bother it had caused them over the last 24 hours on a numerical rating scale of 0–10 (0 = not breathless at all/no trouble or bother at all; 10 = breathlessness as bad as you can imagine/trouble or bother as bad as you can imagine, respectively).

PROTOCOL
Repeatability study
The tests were carried out in control subjects and patients in an identical manner. After resting for five minutes FEV_1 and FVC were measured. The numbers tests, each consisting of five readings, were carried out twice in one day 30 minutes apart to assess within day repeatability. A further test was carried out at the same time of day one week later to assess between day repeatability.

Table 1   Mean (SE) number of numbers read and number of numbers read per breath over 60 seconds in control and patient groups. Test 1,1 etc indicates day 1, test 1, etc

	Controls	Patients	Mean (95% CI) difference	p value for difference
Number of numbers read				
Total				
Test 1,1	91.4 (3.3)	81.1 (3.4)	10.3 (2.5 to 18.1)	0.03
Test 1,2	95.1 (3.6)	83.8 (3.5)	11.3 (3.1 to 19.5)	0.03
Test 2,1	96.2 (3.5)	83.8 (3.3)	12.4 (4.2 to 20.6)	0.01
Per breath				
Test 1,1	11.0 (1.3)	6.6 (0.6)	4.4 (1.5 to 7.3)	0.003
Test 1,2	11.8 (1.1)	7.1 (0.7)	4.7 (1.9 to 7.6)	0.001
Test 2,1	12.3 (1.2)	7.2 (0.7)	5.1 (2.5 to 7.7)	0.001

Sensitivity study

Prior to drainage of the pleural effusion subjects rested for five minutes and FEV_1 and FVC were measured. The subjects were asked to rate on a scale between 0 and 10 their worst and average breathlessness and the level of trouble or bother that their breathlessness had caused them over the last 24 hours. Subjects then completed five readings of the numbers. Twenty four hours following the aspiration or drainage of the effusion the assessments outlined above were repeated. The volume of effusion removed and any use of additional analgesia was noted. On the basis of the repeatability study nine patients were required to provide a reasonable chance (power 90%; p = 0.05) of detecting a change of 25% in the number of numbers read and of 50% in the number read per breath.

ANALYSIS OF DATA

The highest number obtained from the five readings for each test was used for both the repeatability and sensitivity studies in the analysis. If less than five readings were completed only the number of readings that was common to each test was used—that is, only the first four readings for all tests when the patient failed to complete a fifth test on one occasion. The within and between day repeatability for the maximum number of numbers read and the number of numbers read per breath were assessed as the standard deviation (SD) of the difference between tests and as the intraclass correlation coefficient (the ratio of between subject to total variation) as described by Chinn.[3] The intraclass correlation coefficient allows the repeatability of different scales to be compared and has a maximum value of 1 (perfect repeatability) with values below 0.6 indicating poor repeatability.[3]

Age, FEV_1, and number reading data were compared between control subjects and patients by a *t* test. The difference in the maximum number of numbers read and the number of numbers read per breath were compared before and after drainage of the effusion by paired *t* test. The difference in the ratings of breathlessness over the preceding 24 hours (at worst, on average, and the trouble or bother it caused) was analysed using the non-parametric Wilcoxon signed rank test. Correlations were carried out applying Bonferroni's correction. A p value of <0.05 was regarded as being statistically significant.

Results

The concept of reading numbers was easily understood by all subjects.

REPEATABILITY STUDY

Mean FEV_1 was 99% predicted[5] in the control group and 49% predicted in the patients with a mean FEV_1/FVC ratio of 67%. Six patients were taking no medication whilst the others were taking an inhaled bronchodilator (n = 13), opioid analgesic (n = 17), oral steroid (n = 2), and other medications (n = 11) in constant dose throughout.

All control subjects completed all the readings whereas 12 patients (40%) were unable to complete all five readings in all tests due to tiredness. Six patients stopped after four readings and two after three readings in all tests, whilst a further four patients completed five readings in the first test but managed only four or three readings in subsequent tests.

There was a learning effect during the first test with an increase in the number of numbers read and the number of numbers read per breath over the five readings in both groups. For the patients the largest increase occurred between the first and second readings in the first test (fig 2).

The mean values for the maximum number of numbers read over 60 seconds in the three tests were higher in the control subjects (91.4–96.2) than in the patients (81.1–83.8; table 1) and the same was true for the mean number of numbers read per breath (control subjects 11.8–12.3; patients 6.6–7.2). In the patient group the maximum number of numbers read correlated with the number of numbers read per breath ($r = 0.55$, p = 0.03) and the number of numbers read per breath correlated with FVC ($r = 0.6$, p = 0.004); neither measure correlated with age, FEV_1 % predicted or FEV_1/FVC% in either group.

There was an increase in the mean (95% CI) maximum number of numbers read over 60 seconds both within (3.7 (95% CI 1.9 to 5.5); p<0.001) and between (4.8 (95% CI 3.0 to 6.6); p<0.001) days but not in the number of numbers read per breath within (0.8 (95% CI –0.2 to 1.8)) and between (1.3 (95% CI –0.3 to 2.9)) days (both p>0.1). The maximum number of numbers read was more repeatable than the number of numbers read per breath as judged by the intraclass correlation coefficient (table 2). The number of numbers read was more repeatable within day than between days.

Table 2   Repeatability within and between days, expressed as the standard deviation (SD) of the difference between tests and the intraclass correlation coefficient for the control subjects and patients

	Controls		Patients	
	Within day	Between day	Within day	Between day
Number of numbers read over 60 seconds				
Mean difference between tests	3.7	4.8	2.7	2.7
SD of the difference between tests	4.7	4.9	4.5	6.3
Intraclass correlation coefficient	0.97	0.97	0.98	0.97
Number of numbers read per breath				
Mean difference between tests	0.8	1.3	0.5	0.6
SD of the difference between tests	2.9	4.5	2.0	2.2
Intraclass correlation coefficient	0.95	0.90	0.91	0.85

Table 3   Mean (SE) values of lung function, ratings of breathlessness, number of numbers read and number of numbers read per breath over 60 seconds before and after drainage of the pleural effusion

	Before drainage	After drainage	Mean (95% CI) difference	p value for difference	% change
FEV_1 (l)	0.87 (0.10)	1.04 (0.12)	0.17 (0.06 to 0.28)	0.006	20
FVC (l)	1.09 (0.13)	1.30 (0.16)	0.21 (0.08 to 0.34)	0.004	19
Breathlessness score over the past 24 hours					
At worst	7.4 (0.6)	4.0 (0.5)	3.4 (1.7 to 5.1)	0.005	46
On average	4.5 (0.5)	2.4 (0.4)	2.1 (0.8 to 3.4)	0.008	47
Amount of trouble or bother	6.6 (0.5)	2.3 (0.7)	4.3 (2.3 to 6.3)	0.002	65
No. of numbers read	66.9 (2.5)	82.5 (4.8)	15.6 (9.8 to 21.4)	0.0001	23
No. of numbers read per breath	4.8 (0.8)	7.7 (1.2)	2.9 (1.1 to 4.7)	0.003	60

FEV_1 = forced expiratory volume in one second; FVC = forced vital capacity.

There was an increase in the mean maximum number of numbers read both within (2.7 (95% CI 1.1 to 4.3); p<0.01) and between (2.7 (95% CI 0.4 to 5.1);p<0.05) days but not in the number of numbers read per breath within (0.5 (95% CI –0.3 to 1.3)) or between (0.6 (95% CI –0.2 to 1.4)) days (both p = 0.16). Again the maximum number of numbers read over 60 seconds had a higher intraclass correlation coefficient than the number of numbers read per breath and both were more repeatable within day than between days (table 2).

Repeatability of the number of numbers read over 60 seconds was similar in control subjects and patients, both within and between days. The number of numbers read per breath were generally more repeatable in the control subjects than in the patients, both within and between days (table 2).

SENSITIVITY STUDY
Mean values for FEV_1 and FEV_1/FVC were 37% predicted and 80%. Six patients were taking no regular medication whilst the others were taking a diuretic (n = 4), an inhaled bronchodilator and corticosteroid (n = 2), an opioid (n = 3) and a stable dose of prednisolone, digoxin, diazepam, theophylline and oxygen (n = 1 for each).

Prior to drainage of their effusion two patients could only complete four readings during the test and in one patient this persisted after drainage. There was no correlation between the maximum number of numbers read or the number of numbers read per breath and FEV_1, FEV_1% predicted, FVC, FEV_1/FVC%, or breathlessness scores prior to drainage.

The mean (range) volume of fluid drained from the pleural space was 1840 ml (750–3800). Five patients had additional analgesia after insertion of the chest drain with a non-steroidal anti-inflammatory drug or paracetamol (n = 3) or an opioid (n = 2). None of the patients was limited by chest pain when reading the numbers.

Following drainage of the effusion no patient experienced a complete resolution of symptoms but all reported an improvement in breathlessness and breathlessness scores fell significantly (table 3). Only two patients remained short of breath at rest. FEV_1 and FVC improved following the procedure (table 3) though neither these nor the change in breathlessness scores correlated with the volume of fluid drained. There was an increase in both the maximum number of numbers read (23%; p<0.001) and the number of numbers read per breath (60%; p<0.01) following drainage of the effusion (table 3) although neither correlated with the volume of pleural fluid drained nor with the change in FEV_1, FVC, and breathlessness scores.

Discussion
This study has explored the use of reading numbers aloud to measure the limiting effect of breathlessness in patients with cancer who were breathless on talking or at low levels of exertion. The concept was easily understood, simple to carry out, and well tolerated apart from tiredness which meant that a third of the patients were unable to complete all five readings for all tests.

Compared with the healthy control subjects, the patients with breathlessness read significantly fewer numbers and fewer numbers per breath over 60 seconds supporting the face validity of the test as a measure of the limiting effect of breathlessness. Patients only achieved 59–60% of the number of numbers read per breath and 87–89% of the total number of numbers read compared with the control subjects (table 1; fig 2). The patients were slightly but not significantly older than the control subjects and contained more men. Neither the number of numbers read nor the number read per breath correlated with age or sex, however. The lack of correlation between the number of numbers read and the number read per breath and FEV_1% or FEV_1/FVC% in both groups may be a reflection of the narrow range of age and spirometric values in both patients and control subjects. The number of numbers read per breath correlated with FVC in the patients which suggests that factors such as lung compliance, respiratory muscle strength, and residual lung volume which impact upon FVC may influence the number of numbers read.

Subjects carried out five readings for each test following pilot studies. A learning effect was observed with successive tests with a significant increase in the number of numbers read and a non-significant increase in the number of numbers read per breath. The effect was small in the patient group (maximum improvement 3%) but would need to be taken into account when designing studies. Both measures showed good repeatability both within and between days in both groups with the number of numbers read over 60 seconds the more repeatable measure as judged by the intraclass correlation coefficient (table 2).

Table 4   Intraclass correlation coefficients for the number of numbers read and the number of numbers read per breath using between day repeatability data. Those for the visual analogue (VAS) and numerical rating scales (NRS) were obtained using data from a separate study[2]

	Intraclass correlation coefficient		
		VAS	NRS
No. of numbers read	0.97		
No. of numbers read per breath	0.85		
Breathlessness score over the past 24 hours			
At worst		0.81	0.89
On average		0.83	0.92
Amount of trouble or bother		0.86	0.91

The between day intraclass correlation coefficient for number reading was compared with those obtained for visual analogue and numerical rating scales using data from a previous study involving similar patients with breathlessness due to cancer.[2] The number of numbers read has the highest intraclass correlation coefficient of all three types of measurement (0.97). The value obtained for the number of numbers read per breath (0.85) is similar to those of the visual analogue scale (0.81–0.86) but is less than with the numerical rating scale (0.89–0.92) (table 4).

We have used these data to calculate the sample size needed for future studies and arbitrarily selecting a change equivalent to 50% of that seen following drainage of the pleural effusion. For a within subject between day study nine and 15 patients would be required to reliably (90% power; p = 0.05) detect such a change in the number of numbers read and the number of numbers read per breath, respectively.[5] Data from the previous study[2] suggest that the sample sizes required to detect the equivalent change in worst and average breathlessness and the trouble or bother it causes, respectively, would be 16, 29, and 25 patients using a numerical rating scale and 40, 35, and 42 patients for a visual analogue scale. Although the differences in the number of patients required to provide sufficient power between the three methods are relatively small, they could be important considering the difficulties in recruiting this group of patients into studies.

Following drainage of their effusions all patients had a significant increase in FEV_1, FVC, and an improvement in breathlessness scores of 46–65% (table 3). This was associated with a 23% increase in the number of numbers read and a 60% increase in the number of numbers read per breath. Whilst the increase in the number of numbers read over 60 seconds was proportionally smaller than the increase in breathlessness scores, the former may be the more discriminating test since it produced the most statistically significant effect (p = 0.0001) and was more repeatable. The improvement seen in these two very different measures of breathlessness—that is, numerical rating scales and number reading—provides criterion validity for the use of number reading as an indirect measure of breathlessness. The lack of correlation between improvement in breathlessness scores and the changes in the number of numbers read and number of numbers read per breath, FEV_1, FVC, or FEV_1% may in part be due to the relatively small sample.

Thus, the measurement of the number of numbers read over 60 seconds and the number of numbers read per breath is practical, easy to carry out, shows good repeatability within and between days, and is sensitive to the improvement seen following drainage of a pleural effusion. Sample size calculations suggest that, with the use of number reading to assess the limiting effect of breathlessness, a smaller number of patients would be required than for numerical and visual analogue scales for intervention studies.[2 5] Number reading therefore may provide a useful measure of the limiting effect of breathlessness in patients with cancer who are breathless at rest or on minimum exertion in a study setting. It is only suitable for assessing interventions that do not affect cognition; future studies will use this method to examine the effect of interventions such as oxygen therapy, theophylline, or respiratory muscle training in this group of patients.

1 Wilcock A. The measurement and treatment of breathlessness in patients with cancer. DM Thesis: University of Nottingham, 1997.
2 Crosby V, Wilcock A, Clarke D, *et al*. Repeatability of breathlessness measurements in cancer patients. *Thorax* 1999;**54**:375.
3 Chinn S. Repeatability and method comparison. *Thorax* 1991;**46**:454-6.
4 Quanjer P. Standardised lung function testing. *Bull Eur Physiopathol Respir* 1983;**19**(Suppl 5):1–95.
5 Machin D, Campbell MJ. *Statistical tables for the design of clinical trials*. Oxford: Blackwell Scientific, 1987.