Long-term reproducibility of Borg scale estimates of breathlessness during exercise

RACHEL C. WILSON AND P. W. JONES
Department of Medicine, Division of Physiological Medicine, St George’s Hospital Medical School, London

(Received 16 May/5 October 1990; accepted 26 October 1990)

SUMMARY

1. The intensity of breathlessness in normal subjects during exercise was measured on seven occasions over a 40-week study period to assess the long-term repeatability of Borg scale estimates of breathlessness.

2. In all subjects there was a significant correlation (P=0.0001) between breathlessness and minute ventilation. Minute ventilation measured at each work rate did not differ between the seven exercise tests (P>0.05).

3. There was no significant difference between the mean Borg scores (measured with respect to a given level of ventilation) in 5 of the 7 test weeks (P>0.05). The slope of the relationship Borg score/minute ventilation did not differ between the seven exercise tests (P>0.05).

4. Breathlessness estimation was highly reproducible both after 1 week and after 40 weeks of the study (both P>0.05).

5. The duration without testing between consecutive tests did not affect reproducibility: the mean Borg score was equally reproducible after an interval of 22 weeks without testing as after an interval of 1 week (P>0.05).

6. The Borg scale provides a reliable technique for studying the sensation of breathlessness over extended time periods.

Key words: breathlessness, exercise, perception, reproducibility.

Abbreviations: FEV_{1}, forced expiratory volume in 1 s; FVC, forced vital capacity; \bar{V}_{E}, minute ventilation.

INTRODUCTION

It has been shown that within an individual, estimates of perceived breathlessness during exercise are reproducible over a period of a few days [1] and over a period of 1 or 2 weeks [2–4]. Although many studies of breathlessness involve serial measurements taken over a longer period, to our knowledge there are no reported observations of the stability of breathlessness measurements on exercise over a period exceeding 2 weeks. Adams et al. [5] recorded that the repeatability of visual analogue scale measurements of breathlessness during hypercapnia was poor after an interval of a year. Since, however, the sensation experienced with hypercapnia is very different to that during exercise [5, 6] caution must be used when extrapolating between the two experimental states. The aim of this study was to establish the repeatability of Borg scale estimates of breathlessness in a group of normal subjects over a 40-week period, with increasing intervals between studies.

METHODS

Subjects

Seven healthy volunteers (six males and one female), aged between 23 and 31 years, were used. All gave informed consent. The subjects had no history of cardiorespiratory disease and all had normal forced expiratory volume in 1 s (FEV_{1}) and forced vital capacity (FVC). Their mean FEV_{1}/FVC ratio was 83%.

Experimental design

Each subject performed an incremental exercise test on a cycle ergometer on seven occasions over a 40-week study period. The exercise tests were performed during week 1 (WK1), week 2 (WK2), week 3 (WK3), week 4 (WK4), week 6 (WK6), week 18 (WK18) and week 40 (WK40) of the study. Each exercise test consisted of 4 min of cycling at 25 W, followed by a 1 min increment progressive exercise test of 8 min duration. Each workload increment was 20 W for the males and 15 W for the female.
Study protocol

Subjects were naive of the design and objectives of the experiments. None of the subjects had previous experience of breathlessness scaling. Before each exercise test, subjects were familiarized with the apparatus and read a detailed description of the nature of the sensation to be measured. All subjects expressed that they felt confidently able to measure the sensation exclusive of other sensations associated with exercise. Over the study period none of the subjects undertook any strenuous exercise training programme.

Objective measurements

Expired gas was collected using a no. 2700 Hans Rudolph valve with mouthpiece and noseclip. Minute ventilation ($V_E$) was measured at 15 s intervals using a Jaeger E.O.S. Sprint exercise system. This consists of a pneumotachograph with digital integration of the flow signal to provide tidal volume.

Subjective measurements

The subjects were asked to quantify their feelings of breathlessness. This was defined to them as the feeling of 'an uncomfortable need to breathe' rather than any other sensation associated with exercise, such as fatigue or their awareness that ventilation had increased. At the end of each minute of exercise subjects were asked to estimate their level of breathlessness, if any, using a modified Borg scale [7] (Fig. 1). A light could be placed at the appropriate position on the scale, by use of a hand-operated rotary switch. No intermediate positions could be selected between points on the scale. After each recording the scale indicator was reset at zero to ensure that every estimation was made independently of the previous estimation.

Statistical analysis

The results were analysed using analyses of variance and co-variance. All comparisons between different exercises were made by analysis of the group data after removal of the inter-subject variance. This was performed by using a linear modelling technique. Statistical significance was assessed by using an $F$-test and significance was accepted at the 1% and 5% level (where stated). The results are expressed as mean and sds. The nature of the breathlessness-ventilation relationship was found to approximate well to a linear regression. Graphical representation of the results uses the mean regression estimates for all subjects obtained from analysis of co-variance. Validation of this statistical analysis has been discussed previously [3]. Differences in mean Borg scores between tests were analysed using the mean Borg score derived from the analysis of co-variance of Borg score against $V_E$, i.e. as a measure of the mean vertical separation of the regression slopes. Thus, changes in mean Borg score between tests were measured with respect to ventilation and not work rate.

RESULTS

Reproducibility of ventilatory response to exercise

The mean $V_E$ of the group data, measured at each work rate, did not differ significantly between the seven exercise tests (analysis of variance, $P>0.05$).

Relationship between Borg score and $V_E$

In each subject there was a significant correlation between breathlessness and $V_E$ (median $r=0.85$, range 0.70-0.95, $P=0.0001$). There were no significant differences in the slopes of the Borg score/$V_E$ relationship obtained in the seven exercise tests ($P>0.05$).

Comparison of mean Borg score of all seven tests

Fig. 2 displays the Borg score/$V_E$ relationship for the grouped data from all subjects and from one individual subject. For clarity, WK1, WK4, WK18 and WK40 only are displayed. Comparison of the mean Borg score at a given $V_E$ showed no significant difference between five of the seven exercise tests ($P>0.05$). The mean Borg score at a given $V_E$ was reduced in WK3 compared with WK1 ($P<0.05$) and in WK6 compared with WK1, WK2, WK4, WK18 and WK40 ($P<0.01$). All other differences between tests were not significant ($P>0.05$). Fig. 3 illustrates the pattern of the changes in mean Borg score. These changes may suggest a small downwards trend in mean Borg score over the first 6 weeks of successive testing, with the trend later being lost or reversed. To test for this, a polynomial regression was fitted to the data. There was no significant second-order component ($P>0.05$). Analysis of the results in the individual subjects revealed that only two subjects showed a pattern of changes over time that could be considered 'typical' of the changes illustrated in Fig. 3 (i.e. a fall followed by a rise in breathlessness). Two others showed a pattern suggesting an entirely opposite trend, i.e. increasing followed by decreasing dyspnoea. One subject showed a general trend downwards over the entire sequence of seven tests, whilst in the remaining two subjects no pattern could be read into the data. Overall there were no

---

**Statistical analysis**

The results were analysed using analyses of variance and co-variance. All comparisons between different exercises were made by analysis of the group data after removal of the inter-subject variance. This was performed by using a linear modelling technique. Statistical significance was assessed by using an $F$-test and significance was accepted at the 1% and 5% level (where stated). The results are expressed as mean and sds. The nature of the breathlessness-ventilation relationship was found to approximate well to a linear regression. Graphical representation of the results uses the mean regression estimates for all subjects obtained from analysis of co-variance. Validation of this statistical analysis has been discussed previously [3]. Differences in mean Borg scores between tests were analysed using the mean Borg score derived from the analysis of co-variance of Borg score against $V_E$, i.e. as a measure of the mean vertical separation of the regression slopes. Thus, changes in mean Borg score between tests were measured with respect to ventilation and not work rate.

**RESULTS**

Reproducibility of ventilatory response to exercise

The mean $V_E$ of the group data, measured at each work rate, did not differ significantly between the seven exercise tests (analysis of variance, $P>0.05$).

**Relationship between Borg score and $V_E$**

In each subject there was a significant correlation between breathlessness and $V_E$ (median $r=0.85$, range 0.70-0.95, $P=0.0001$). There were no significant differences in the slopes of the Borg score/$V_E$ relationship obtained in the seven exercise tests ($P>0.05$).

**Comparison of mean Borg score of all seven tests**

Fig. 2 displays the Borg score/$V_E$ relationship for the grouped data from all subjects and from one individual subject. For clarity, WK1, WK4, WK18 and WK40 only are displayed. Comparison of the mean Borg score at a given $V_E$ showed no significant difference between five of the seven exercise tests ($P>0.05$). The mean Borg score at a given $V_E$ was reduced in WK3 compared with WK1 ($P<0.05$) and in WK6 compared with WK1, WK2, WK4, WK18 and WK40 ($P<0.01$). All other differences between tests were not significant ($P>0.05$). Fig. 3 illustrates the pattern of the changes in mean Borg score. These changes may suggest a small downwards trend in mean Borg score over the first 6 weeks of successive testing, with the trend later being lost or reversed. To test for this, a polynomial regression was fitted to the data. There was no significant second-order component ($P>0.05$). Analysis of the results in the individual subjects revealed that only two subjects showed a pattern of changes over time that could be considered 'typical' of the changes illustrated in Fig. 3 (i.e. a fall followed by a rise in breathlessness). Two others showed a pattern suggesting an entirely opposite trend, i.e. increasing followed by decreasing dyspnoea. One subject showed a general trend downwards over the entire sequence of seven tests, whilst in the remaining two subjects no pattern could be read into the data. Overall there were no
Reproducibility of breathlessness scaling

Fig. 2. Comparison of the Borg score/$V_E$ relationships on exercise between WK1 (■), WK4 (△), WK18 (○) and WK40 (▲) from the grouped data from all subjects (a) and from one individual subject (b). There was no significant difference between the slopes of the Borg score/$V_E$ relationship or between the mean Borg scores of the four exercise tests ($P > 0.05$).

consistent trends and the apparent trends in the grouped data were due to random variation within and between subjects. All of the mean changes were small, as can be seen in the results for the subject illustrated in Fig. 2(b).

Comparison of maximum Borg score and $V_E$ of all seven tests

The maximum level of ventilation at the point of maximum exercise was not significantly different between the seven exercise tests ($P > 0.05$). Fig. 4(a) illustrates maximum $V_E$ for each exercise test.

The maximum Borg score at the point of maximum exercise was not significantly different between the seven exercise tests ($P > 0.05$). Fig. 4(b) illustrates maximum Borg score for each exercise test.

Effect of intervals between consecutive tests

There was no significant difference ($P > 0.05$) in mean Borg score between any two exercise tests that were separated by 1 week and there was also no significant difference ($P > 0.05$) between the mean Borg scores of WK18 and WK40 which were separated by an interval of 22 weeks.

DISCUSSION

Magnitude estimates of breathlessness made using a modified Borg scale were highly reproducible after 1 week of the study and equally reproducible when repeated at the end of the 40-week study period. Within the study period the interval without testing between consecutive tests did not affect the reproducibility of the technique.

The high repeatability of the Borg scale estimates of breathlessness after both 1 week and 40 weeks indicates that this technique is equally suited to both short- and long-term studies. The interval without testing between consecutive exercise tests did not affect the reproducibility of the Borg scale. Over the first 4 weeks of the study, when exercise tests were performed at intervals of 1 week, estimates of perceived breathlessness were reproducible between consecutive exercise tests. The estimates were equally reproducible over the 22-week interval between WK18 and WK40. These results suggest that the Borg scale provides a reliable technique for the measurement of breathlessness during consecutive exercise tests irrespective of the length of time between tests. The only question concerning its repeatability arose when serial tests were performed at frequent intervals. Within the 40-week study period only two of the seven tests (in WK3 and WK6) showed a significant difference from the remaining tests. These observations could be interpreted to suggest a downward trend in perceived breathlessness relative to ventilation over the first 6 test weeks when tests were performed frequently at short intervals, followed by an increase in breathlessness back to the original level after a prolonged interval without testing. However, no statistically significant trends could be identified in the group data and there was no pattern in the change over time that was common to all subjects. We conclude that
R. C. Wilson and P. W. Jones

Fig. 4. (a) Maximum $V_E$ for each test week. There were no significant differences over the 7 test weeks ($P > 0.05$).

(b) Maximum Borg scores for each test week. There were no significant differences over the 7 test weeks ($P > 0.05$).

...the small differences between the tests and the apparent trends were due to random variations within and between subjects.

There is recent evidence that subjects may distinguish different respiratory sensations depending on the experimental stimulus [6]. In the current study, breathlessness was defined as 'an uncomfortable need to breathe'. This definition has been used in previous studies by others [5] and ourselves [3]. Other workers who have demonstrated the short-term repeatability of breathlessness measurements defined breathlessness as 'the sense of effort' [4] whilst some workers have simply asked their subjects to measure their breathlessness, without defining the sensation [1, 2]. We cannot say whether the choice of wording influences the stability of the measurements, since no comparative studies have been made. However, in the short term, 'an uncomfortable need to breathe' [3] and 'the sense of effort' [4] were scaled relative to ventilation with good repeatability using both the Borg and the visual analogue scales. We are not aware of published data concerning the long-term repeatability of the visual analogue scales for breathlessness measurement during exercise. One of us has experience of the use of visual analogue scales for breathlessness measurement in a study that lasted several months [8]. In that study, the repeatability of the visual analogue scores was too low to allow analysis of the results and we would recommend that the long-term repeatability of visual analogue scores should be confirmed before studies of long duration are attempted using this scaling technique.

In summary, Muza et al. [4] conclude that the visual analogue scale has good short-term repeatability and may be more sensitive than the Borg scale for detecting subtle differences in breathlessness. We have also shown good repeatability for the visual analogue scale, but found that the Borg scale had slightly higher repeatability when compared in the same subjects over a 2-week period [3]. We have now shown that the Borg scale provides a reproducible method for the measurement of breathlessness during exercise over extended inter-test intervals and long study periods. A body of evidence is accumulating that allows informed decisions concerning the selection of a scaling technique for a particular study.

ACKNOWLEDGMENT

R.C.W. was supported by the Wellcome Trust.

REFERENCES