Biological quality control for exercise testing

S M Revill, M D L Morgan

Abstract

Background—A study was undertaken to evaluate the use of a biological quality control programme for a computer controlled, breath-by-breath exercise test system over a 2.5 year period.

Methods—One healthy volunteer performed a regular progressive treadmill test with breath-by-breath measurements of oxygen uptake ($\dot{V}O_2$), carbon dioxide output ($\dot{V}CO_2$), ventilation ($\dot{V}E$), and the ECG heart rate (HR). Following a familiarisation period, five consecutive tests were performed and the measurements at peak exercise were averaged to give baseline values. All tests were compared with these values.

Results—A total of 35 tests were recorded. The within subject standard deviation for measurements at peak exercise were 52 ml/min for $\dot{V}O_2$, 74 ml/min for $\dot{V}CO_2$, 3.1 l/min for $\dot{V}E$, and 3 beats/min for HR. The mean (SE) percentage variation in measurements at peak exercise compared with the baseline values was +0.37 (0.30)% for $\dot{V}O_2$, –0.10 (0.39)% for $\dot{V}CO_2$, –0.88 (0.52)% for $\dot{V}E$, and +1.2 (0.26)% for HR. The variability present in measurements made during high and moderate intensity exercise (73% $\dot{V}O_2$peak) was not significantly different ($p > 0.05$). During the study period the quality control detected a fault on the oxygen analyser which was not apparent from the automatic calibration.

Conclusions—Regular quality control using a healthy volunteer allows all components of the breath-by-breath system to be checked simultaneously and in a manner which is consistent with its clinical use. This practice can highlight faults not detected by automatic calibration.

(*Thorax* 2000;55:63–66)

Department of Respiratory Medicine, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK
S M Revill
M D L Morgan

Correspondence to:
Dr S M Revill

Biological quality control for exercise test systems is recommended[1] but the practicality, simplicity, and ease with which it can be incorporated into the quality control programme of a clinical service department is not widely documented. Previous papers have addressed the issue of exercise quality control using steady state values with mixing chamber exercise systems.[2] However, there are now many breath-by-breath systems in routine service use and few published data concerning the variability and reliability of such systems. Gross technical faults are usually immediately evident but smaller changes in analyser output may be computer corrected during the calibration procedure whilst an erroneous set of results from a human subject may go undetected.

The computer controlled, breath-by-breath machines are complex multi-component systems and therefore assessment of the response of the equipment working as an integrated unit is important. Additionally, since clinical assessment of the maximum exercise response is a common feature it is also appropriate to examine equipment performance under similar operating conditions. In this paper we report the results from a biological quality control programme over a 2.5 year period which highlights the issues of equipment stability and the identification of technical malfunction.

Methods

A progressive incremental exercise test was performed regularly by a single member of staff over a period of 2.5 years. Since we have a particular interest in walking exercise prescription and regularly use the treadmill to assess maximum functional performance, the quality control test was performed on the treadmill (LE 3000, Erich Jaeger (UK) Ltd). During periods of regular use the quality control test was carried out on a weekly basis but, when the exercise equipment was not used for a period of time, the quality control test was suspended until the week prior to the next appointed test. The Balke protocol[3] was individualised to elicit a measurement of peak oxygen uptake ($\dot{V}O_2$peak) from the human control subject with five two minute increments in treadmill gradient at a constant running speed of 9 km/h. The initial gradient of zero was increased in steps of 2%. Breath-by-breath measurements of oxygen uptake ($\dot{V}O_2$), carbon dioxide output ($\dot{V}CO_2$), and ventilation ($\dot{V}E$) were made with a computerised exercise system (OxyconBeta, Erich Jaeger (UK) Ltd). The heart rate (HR) was measured from a single lead ECG signal (Graseby Medical, Herts, UK).

The $\dot{V}O_2$ and $\dot{V}CO_2$ were reported at STPD and the $\dot{V}E$ at BTPS. During the exercise tests the turbine flowmeter was held in place using a rubber mouthpiece whilst a noseclip was used to seal the nose. The quality control test was performed at the same time of day and with a three hour interval following the last meal. An effort was made to follow a consistent diet on the test days.

Following several practice tests, five tests were performed over a short period of time (eight days) and measurements at the final work rate (final 30 seconds of the 8% gradient) averaged to provide baseline values. All tests were then compared with the baseline values and the absolute change plotted on a graph in order to visualise trends.

Table 1   Mean (SD) oxygen uptake ($\dot{V}O_2$), carbon dioxide output ($\dot{V}CO_2$), ventilation ($\dot{V}E$) and heart rate (HR) at peak exercise (final 30 s of 8% gradient) and differences in the variability at moderate (M; 0% gradient, 73% of $\dot{V}O_2$peak) and high (H; $\dot{V}O_2$peak) intensity exercise (n = 33)

	$\dot{V}O_2$ (ml/min)	$\dot{V}CO_2$ (ml/min)	$\dot{V}E$ (l/min)	HR (beats/min)
Peak exercise				
Mean (SD)	3072 (52)	3307 (74)	101.1 (3.1)	172 (3)
Coefficient of variation	1.7%	2.2%	3.1%	1.7%
Variability (H vs M)				
Mean difference	14.7	−16.8	−0.8	−0.7
95% CI	−17.6 to 47.0	−52.6 to 19.0	−1.8 to 0.2	−2.3 to 0.9
p value	0.36	0.36	0.11	0.38

EQUIPMENT

The Oxycon Beta system uses an infrared CO_2 analyser, a fast response paramagnetic O_2 analyser, and a turbine meter to measure airflow at the mouth. The expired gas is sampled continuously from a side port in the mouthpiece turbine valve. The computer uses an algorithm to correct for the delay in gas analysis thus ensuring that the breath-by-breath flow and gas analysis signals are aligned. Following a warm up period of one hour, the exercise system was calibrated according to the manufacturers' recommendations—that is, electronic system check, two point gas calibration (including response and time delay measurements), and three litre syringe calibration. The system has internal temperature and barometric pressure transducers to enable the correction of gas volumes.

The validity of the system had been examined in a previous experiment by comparison with a "gold standard" Douglas bag system. Five healthy volunteers had performed repeat symptom limited maximal exercise tests on each system on separate days. The 95% confidence intervals for the differences between systems at peak exercise were −177 to +24.2 ml/min for $\dot{V}O_2$peak, −224 to +338 ml/min for $\dot{V}CO_2$, and −23 to +15 l/min for $\dot{V}E$. Additionally, the manufacturers made available unpublished data (Keizer *et al*, University of Limberg) from a comparison of the breath-by-breath system with the Douglas bag method in a large group of healthy individuals which showed small and non-significant differences when the systems were connected in series and also between day measurements.

ANALYSIS OF DATA

The data are presented as mean (SD) unless otherwise stated. Differences between high and moderate intensity exercise were compared using the paired Student's *t* test and the 5% level of significance was adopted throughout.

Results

Thirty five tests were performed over a period of 2.5 years. The mean (SD) measurements at the highest work rate (8% gradient) are shown in table 1 and the absolute changes from the baseline values for each test are illustrated in fig 1. The figure includes two measurements where the variation in $\dot{V}O_2$ was >6.5% whilst the variation in HR and $\dot{V}CO_2$ was <2%. These outlying points suggested there was a fault on the O_2 analyser. In addition, it had been noted that the gas calibration had been drifting off, although it was still within the limits accepted by the computer and therefore not sufficient to trigger a warning alarm. At this point the com-
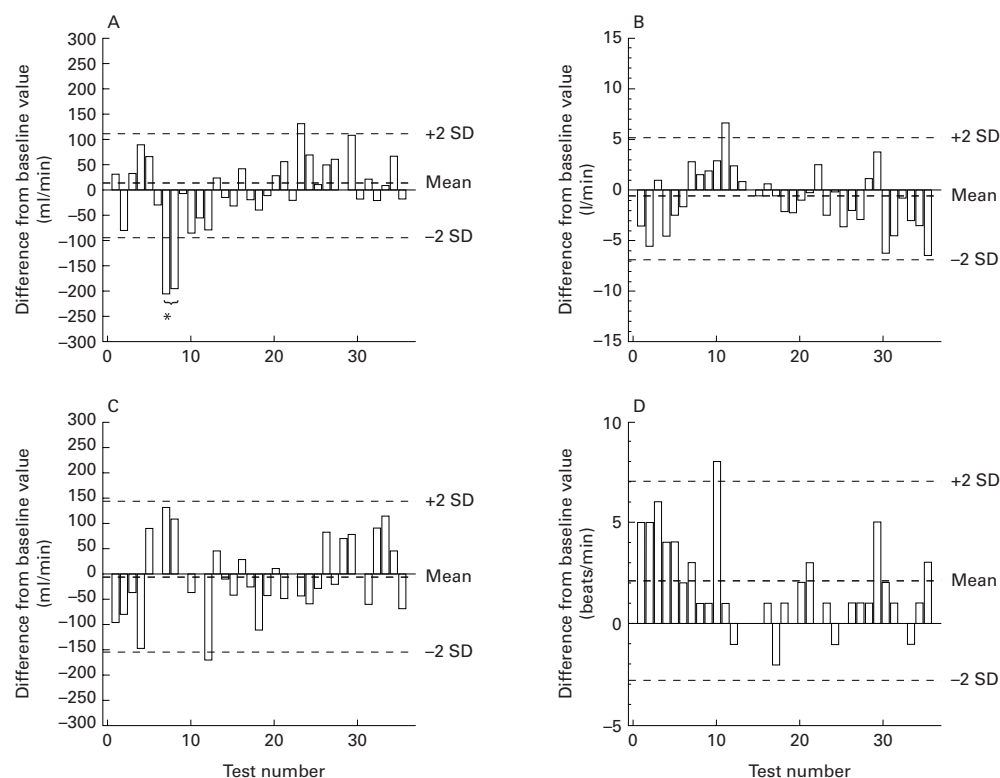

Figure 1   Differences from baseline values for the four key exercise variables: (A) oxygen uptake, (B) ventilation, (C) carbon dioxide output, and (D) heart rate. The mean difference and limits of agreement (±2SD) are shown (n = 33). Asterisk indicates a technical fault on the oxygen analyser.
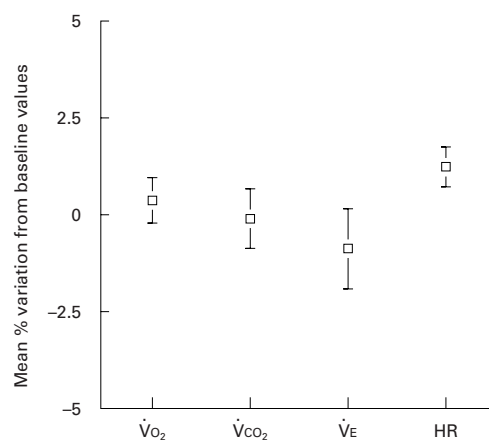
Figure 2   Mean (95% confidence intervals) percentage variation from the baseline values for each of the four key exercise variables: oxygen uptake ($\dot{V}O_2$), carbon dioxide output ($\dot{V}CO_2$), ventilation ($\dot{V}E$), and heart rate (HR).

pany engineer was called in and checks on the analyser voltage output revealed that a replacement analyser was required. The data from these tests have not been included in the statistical analysis.

The mean (SE) percentage variation in measurements at peak exercise compared with the baseline values was +0.37 (0.30)% for $\dot{V}O_2$, –0.10 (0.39)% for $\dot{V}CO_2$, –0.88 (0.52)% for $\dot{V}E$, and +1.2 (0.26)% for HR (fig 2). The variability present in measurements made during moderate exercise (running speed 9 km/h, gradient 0%, submaximal $\dot{V}O_2$ equivalent to 73% $\dot{V}O_2$peak) and high intensity exercise (speed 9 km/h, gradient 8%) was not significantly different (p>0.05; table 1).

Discussion
The aim of this paper was to describe a simple quality control programme in practice which was easy to perform and demonstrated the effectiveness of a biological control to ensure the reliability of an integrated, computerised breath-by-breath exercise system. The validity of the system was initially ensured by comparison with a "gold standard", a manual Douglas bag system. It is important to check the validity of the breath-by-breath systems although it is likely this will become increasingly difficult as computerised systems replace the Douglas bag equipment. For clinical service departments it is worthwhile fostering good relations with a local university department of physiology or sports science where such systems are usually in frequent use. Arranging the systems in series minimises biological variation. Additionally, the manufacturers of breath-by-breath systems should make available data from their own validity checks, which preferably should be performed by an independent assessor.

The type of quality control programme we have described records both equipment and biological variability. Using healthy subjects who are practised and familiar with the equipment is likely to minimise biological variability and represents a reasonably stable "test signal". We have shown that, under such conditions, biological variability is very small. It compares favourably with mechanical systems

simulating respiratory gas exchange[4] and is below the variability measured in chronic airflow limitation (3.5% for $\dot{V}O_2$peak)[5] and restrictive lung disease (5.3% for $\dot{V}O_2$peak).[6]

The variation recorded from the current programme was similar to that described by Jones and Kane[2] for repeat measurements in individual healthy subjects over a period of 26 months. It was also noted that the combined variation from a number of subjects was greater than the variation within an individual subject, and that variation increased over prolonged periods of time.[2] The protocol described by Jones and Kane[2] consisted of three steady state workloads with a mixing chamber configuration. For our quality control programme with the breath-by-breath system we felt it was appropriate to examine the response of the equipment under conditions that were similar to its usual clinical operation; we therefore chose a progressive incremental protocol that exercised the subject to maximal capacity. Additionally, we examined the variability in maximal values since the purpose of routine clinical exercise testing is to establish a patient's maximal functional capacity.

The data from the present study and from others[2] suggest that variations in $\dot{V}O_2$, $\dot{V}CO_2$, and HR of <5% can be expected from a carefully executed biological quality control programme. The variability in $\dot{V}E$ is slightly wider, although the current data suggest that this can still be <7% for an individual. The European guidelines suggest that a variation of >7% from the previous value for $\dot{V}O_2$ should engender a system wide reassessment.[1] In this study we have shown that regular recording and statistical analysis of the four key variables—that is, $\dot{V}O_2$, $\dot{V}CO_2$, $\dot{V}E$, and HR—enables the repeatability and limits of agreement of an individual exercise system to be established and provides a comprehensive system of control. Additionally, performance of a biological quality control on a weekly basis will aid personnel training and familiarisation with the exercise system which is important if several members of staff use the equipment on a rotational basis.

The importance of maintaining graphs to display the change in each of the key variables was highlighted by the large fall in $\dot{V}O_2$ on two occasions without corresponding changes in the other variables. A graph allows visualisation of trends over a period of time which enable the operator to judge whether the performance of one or more components is deteriorating. In the current study it was clear that the output from the oxygen analyser was failing on two consecutive tests, although this had not been immediately apparent from the automatic calibration (further deterioration of the output would have triggered the computer alarm system). A repeat test performed the following day can confirm whether changes outside the limits of agreement are more than a rogue result. The graphical display and statistical analysis of biological quality control data is an option that manufacturers should consider when writing or upgrading a system's software.

It is unlikely that weekly tests of approximately 10 minutes will constitute a physiologi-

cal training stimulus; however, mechanical efficiency may improve and this was suggested by the higher heart rate response over the first five tests. In order to establish baseline values from which all subsequent tests will be compared it may be important to perform a few tests initially until the control subject feels comfortable with the chosen protocol and adjustments can still be made. The programme described in this paper could easily be adapted for the cycle ergometer since this will be the usual mode of exercise for clinical testing in many laboratories. By using a human control subject the stability of the ergometer (or treadmill) is assessed alongside that of the gas analysis since changes in the heart rate response (after elimination of other causes) are likely to reflect changes in ergometer calibration.

In conclusion, we have shown that a regular biological quality control programme for a computer controlled, breath-by-breath exercise system is feasible and represents a valuable resource for monitoring equipment function. The variability in a healthy control subject was less than the variability reported for lung disease. Performance of the quality control on a weekly basis (or the week prior to an appointed test when the equipment is used less regularly)

provides sufficient density of data for the maintenance of quality control graphs which enable the operator to recognise rogue results quickly and to visualise trends over a longer period of time. Such a programme provides an integrated and comprehensive assessment where all components of the system are checked simultaneously and in a manner which is consistent with its clinical use. A biological quality control programme represents good scientific practice, it may act as an aid to staff training, and can highlight equipment faults not apparent on the automatic calibration of a computerised system.

1 Roca J, Whipp BJ. Clinical exercise testing with reference to lung diseases: indications, standardization and interpretation strategies. *Eur Respir J* 1997;**10**:2662–89.
2 Jones NL, Kane JW. Quality control of exercise test measurements. *Med Sci Sports* 1979;**11**:368–72.
3 American College of Sports Medicine. *Guidelines for exercise testing and prescription*. 5th ed. Baltimore: Williams & Wilkins, 1995.
4 Huszczuk A, Whipp BJ, Wasserman K. A respiratory gas exchange simulator for routine calibration in metabolic studies. *Eur Respir J* 1990;**3**:465–8.
5 Cox NJM, Hendriks JCM, Binkhorst RA, *et al.* Reproducibility of incremental maximal cycle ergometer tests in patients with mild to moderate obstructive lung diseases. *Lung* 1989;**167**:129–33.
6 Marciniuk DD, Watts RE, Gallagher CG. Reproducibility of incremental maximal cycle ergometer testing in patients with restrictive lung disease. *Thorax* 1993;**48**:894–8.