Determination of peak cardiorespiratory fitness parameters in children : which averaging method should we use?

Joel Blanchard
Samuel Blais
Philippe Chetaille
Michele Bisson
François P. Counil
Thelma H. Girard
Sylvie Lafrenaye
Frederic Dallaire

Determination of peak cardiorespiratory fitness parameters in children: which averaging method should we use?

Running head:

- Averaging methods for peak cardiorespiratory fitness parameters in children

Joel Blanchard[1,3], BSc, Samuel Blais[1], BSc, Philippe Chetaille[2], MD, Michele Bisson[2], PhD, François P. Counil[1], MD PhD, Thelma H. Girard[2], Sylvie Lafrenaye[1], MD PhD, Frederic Dallaire[1], MD PhD

[1]Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, and Centre de recherche du Centre hospitalier universitaire de Sherbrooke, Sherbrooke, Canada; [2]Department of Pediatrics, Centre mère-enfant Soleil, CHU de Québec, Université Laval, Quebec City, Canada; and [3]Faculty of Physical Activity Sciences, Université de Sherbrooke, Sherbrooke, Canada

Abstract

Purpose

The purpose of this study was to identify which averaging methods most accurately measures peak cardiorespiratory fitness (CRF) parameters [peak O_2 uptake (VO_2), peak O_2pulse and peak respiratory exchange ratio (RER)] in a sample of healthy children and adolescents.

Method

In this cross-sectional multicenter study, we recruited 278 healthy children aged 12 to 17 years old. We compared the mean peak value of three CRF parameters using the recommended averaging methods (30-second block average) with alternative averaging methods such as moving averages or shorter smoothing periods. We also assessed averaging methods for accuracy by individually reviewing breath-by-breath scatter plots.

Results

The 30-second block average method resulted in a lower mean peak VO_2 and in an increased proportion of underestimated peak values. Using a 30-second moving average significantly increased mean peak values which increased accuracy. Similar results were found for peak RER and peak O_2pulse.

Conclusion

The currently recommended averaging method (30-second block average) increased the risk of misinterpretation of peak CRF values in children, and that using a moving average approach decreased misinterpretation and increased accuracy.

Keywords

Exercise Test – Averaging method – Adolescent – Child

Introduction

During cardiopulmonary exercise testing (CPET), modern metabolic carts will automatically analyse and compute gas exchange to provide breath-by-breath (B-by-B) O_2 uptake (VO_2) and CO_2 production (VCO_2) in relation with ventilation, workload and heart rate. There is physiologic and technical variability in B-by-B recordings. To smooth curves and average B-by-B variations, most authors recommend reporting cardiorespiratory measurements averaged over 20-30 seconds (ATS/ACCP, 2003; Geithner et al., 2004; Mahon and Marsh, 1992; Paridon et al., 2006; Prioux et al., 1997).

There are recommendations to divide a cardiopulmonary exercise test in equal blocks of 20 or 30 seconds and average B-by-B data points within each of these blocks, thus creating a data point for each 30-second interval of recording. (ATS/ACCP, 2003; Balady et al., 2010) This has been often called a "rolling average" but in this article, we will refer to this average method as a "block average". There are several other ways of averaging data points. Among them, we have been using a moving average similar to what is often used in the financial sector. In this method, each B-by-B data point represents the center of a sliding time window and an average of all data points within that time window is calculated. In this article, we will refer to this approach by using the term "moving average".

There are specific challenges in determining peak VO_2 in children, notably because most children fail to reach a plateau of VO_2 at maximal effort (Armstrong et al., 1996; Rowland and Cunningham, 1992). In preliminary data analysis from a previous study in healthy children (Blanchard et al., 2018), we have empirically observed that cardiorespiratory fitness (CRF) peak parameters measurements differed depending on the method and duration used to average data points, especially when no plateau of VO_2 was observed at peak exercise.

In the adult CPET literature, there are recommendations to use a 20-second or 30-second block average (ATS/ACCP, 2003; Balady et al., 2010). It was argued that averaging would minimize the noise of the B-by-B and that 20- or 30-second intervals would be an adequate balance between high precision and variability (long sampling intervals will be less variable, but also less precise) (Myers et al., 2009). For children, the American Heart Association recommends averaging CRF parameters over 30 seconds, without specifying the averaging method (Paridon et al., 2006). These recommendations are expert opinions and we found no study comparing the various methods of averaging in children.

The purpose of this study was to identify which averaging methods (block average or moving average) and which time interval (15-, 20- or 30-second time interval) most accurately measures peak CRF parameters in a sample of healthy children and adolescents. Based on our observations, we hypothesised that the currently recommended 30-second block average results in lower peak VO_2 in children not reaching a plateau of VO_2 at peak exercise, and that using a 30-second moving average results in less underestimation of peak CRF in pediatric CPET. We also compared how averaging methods influenced calculated CRF parameters such as peak respiratory exchange ratio (RER) and peak O_2pulse.

Methods

Population and recruitment

In this cross-sectional multicenter study, we prospectively recruited 240 healthy children aged 12 to 17 years old. Subjects included in this study were also included in a previous study on reference values for CRF parameters in children (Blanchard et al., 2018). As previously described, participants were recruited in local schools from the region of Sherbrooke and Quebec City (Canada). An additional sample of 38 female adolescents aged 15 to 17 years old from a study on pain perception (unpublished results) was also included in the study. Subjects from the later study underwent the same CPET protocol at the Sherbrooke site.

Our inclusion criteria were: healthy children and adolescents between 12 and 17 years of age. Children were excluded for the following reasons: medical conditions forbidding intense exercise, history of exercise discomfort awaiting medical investigation, musculoskeletal, cardiovascular or pulmonary condition limiting exercise performance (e.g. muscular dystrophy, congenital heart disease, asthma, etc.), and current medication influencing CRF, including inhaled corticosteroids and beta agonists. Participants and parents or guardians gave their written consent to participate. The institutional research Ethics Board approved and monitored this research project in both centers.

Cardiopulmonary exercise testing

A symptom-limited CPET progressive ramp protocol was performed on two different electronically-braked cycle ergometers depending on the center: the VIAsprint™ 150P bicycle (CareFusion, New Jersey, USA) and the Corival bicycle (Lode, Groningen, Netherlands). Participants were equipped either with a face mask (7450 Series Silicone V2™ Oro-Nasal, Hans Rudolph, Shawnee, USA) or a mouth piece connected to a Vmax Encore Metabolic Cart (Vmax Encore Metabolic Cart, Sensormedic, San Diego, CA) or an Ultima™ CardiO2 gas exchange analysis system (MGC Diagnostics, St Paul, USA).

The workload was individualized to achieve maximum exertion within 8 to 12 minutes with increasing workload as described previously (Blanchard et al., 2018). Selection of the workload ramp was based on predicted values for height (Godfrey et al., 1971) and then tailored according to the participants level of habitual physical activity. Participants who did not achieve maximal effort within 8 to 12 minutes were either excluded or asked to be retested later. Participants were actively encouraged to keep a constant pedaling rate between 60 to 80 rpm. The test was preceded by a 2-minute rest phase (seated on the cycle ergometer without pedaling) followed by a 3-minute warm-up of unloaded pedaling. The exercise phase was followed by a 3-minute recovery period of slow pedaling (40 rpm) at 10% of the predicted maximal workload. The recovery period was initiated if any of the following events occurred: plateauing of the VO_2 for > 30 seconds, participant asked to stop the exercise phase (for any reason), or inability to maintain constant pedaling > 50 rpm. Except for one subject who was subsequently excluded, participants asking to stop the test all showed subjective signs of intense effort and had difficulty pedaling at > 50 rpm.

We continuously measured 12-lead electrocardiogram, oxygen saturation and breath-to-breath gas exchanges (flow, O_2 uptake and CO_2 production). B-by-B VO_2 ($ml \cdot min^{-1}$), VCO_2 ($ml \cdot min^{-1}$) and V_e ($L \cdot min^{-1}$) were then computed. For each subject we considered that a plateau of VO_2 was reached if the peak VO_2 did not increase by >150 $ml \cdot min^{-1}$ in the last 30 seconds of exercise (analysed on B-by-B data points), or if there was a clear plateauing of the data by visual inspection. Each test was supervised by a physician and an experienced respiratory therapist (Sherbrooke) or kinesiologist (Quebec City).

Peak CRF calculation

We calculated the peak CRF values using two different methods: a block average and a moving average. The block average method consists of dividing the test in a series of fixed time periods (often 30 seconds) in which all B-by-B measures are averaged into one single measurement. A 30-second block average would then transform a 10-minute test in 20 data points each representing an average of 30 seconds of reading. Blocks that overlapped two test stages were not considered in the analysis.

The moving average can be seen as a moving time window. Each B-by-B data point represents the center of a time window and the arithmetic mean of all data points within that time windows is calculated. In a 30-second moving average, for each breath, an averaged point representing all data points for the preceding 15 seconds and the following 15 seconds is calculated. Figure 1 shows an example of an exercise test with important B-by-B variability and its effect on a 30-second block average curve and a 30-second moving average curve.

We applied these two averaging methods to each CRF parameter over intervals of 15, 20 and 30 seconds. The peak value of each parameter of each subject was then determined as the highest averaged data point during the exercise phase. Hence, peak values of each CRF parameter were calculated according to six averaging methods: 15-second, 20-second and 30-second block averages, and 15-second, 20-second and 30-second moving averages.

Visual assessment of peak CRF parameters

To verify if an averaging method correctly estimated the peak value, three blinded reviewers with experience in exercise physiology (F.D., J.B. and F.-P.C.) reviewed individual scatter plots of the B-by-B measurements for a subgroup of 47 consecutive subjects. On each plot, a single blinded reference line representing the peak value computed with one of the average methods was displayed. Figure 2 demonstrates examples of B-by-B scatter plots that the reviewers faced. The plots are from the same subject and the black dotted line represents the peak value as estimated by two different averaging methods.

Each reviewer was asked to subjectively determine if the peak value shown by the black line was accurate, overestimated, or underestimated, according to the B-by-B data points. In this example, all reviewers considered that the peak VO_2 was accurate in panel A, and they all agreed that it was overestimated in Panel B. This evaluation was performed for all the selected maximal CRF parameters for each averaging method and interval.

Statistical Analysis

We used SAS for Windows version 9.4 for all analyses (SAS Institute Inc, Cary, NC). Mean and standard deviation were computed for each peak CRF parameter according to each averaging method. The paired Student t-test was used to compare the mean of differences of the various averaging methods to that of the 30-second block average method for each peak CRF parameter. The proportion of correctly estimated peak values according to the different averaging methods was compared using the chi-square test. The same analyzes were performed during the comparison of the mean values for peak VO_2 according to the presence or not of a plateau. A p-value <0.05 was considered statistically significant.

Results

Two hundred and eighty-six subjects meeting inclusion criteria were invited to participate. Four were excluded for a suspicion of chronic pulmonary disease on spirometry. A total of 282 tests were performed. Four participants were subsequently excluded: one for non-sustained ventricular arrhythmia during exercise testing, one because of an obvious submaximal test and two for not reaching peak exercise within 8-12 minutes. A total of 278 participants (40.3% male) were thus included in the final analysis.

The mean age was 14.56 ± 1.6 years. Mean height and body mass were 165 ± 8 cm (range 143 to 184 cm) and 57 ± 12 kg (range 34 to 95 kg), respectively. Mean BMI-for-age Z score was 0.22 ± 1.00 (range –2.57 to 2.87).

The mean peak VO_2 varied significantly when estimated by the six different averaging methods. Table 1 shows the mean peak values for VO_2, O_2pulse and RER as well as the proportion of correctly estimated peak values, according to the averaging methods. For peak VO_2, the block average methods yielded lower peak values compared to the moving average ($p<0.05$). The mean peak VO_2 for the 30-second block average was ~48 $ml \cdot min^{-1}$ lower than that of the 30-second moving average ($p<0.05$). Differences between the 30-second moving average and the 30-second block average were as high as 300 $ml \cdot min^{-1}$. Similar results were seen for the 15- and 20-second intervals.

For peak VO_2, the 30-second block average had the lowest proportion of accurately estimated values: only 29.0% were correctly estimated according to individual manual review, mostly because the peak value was thought to be underestimated. The averaging methods that yielded the highest proportion of correctly estimated peak values were the 30-second moving average (82.2%), and the 15-second block average (80.9%). These proportions were statistically significantly higher than that of the 30-second block average. Overall, the block average method tended to underestimate peak values compared to moving average. Also, shorter intervals tended to overestimate peak values compared to longer time intervals.

We obtained similar results for the peak O_2pulse and RER (Table 1). The 30-second moving average allowed the highest proportion of correctly estimated values for both peak RER and peak O_2pulse. The 30-second block average had a statistically significantly lower proportion of correctly estimated value and a statistically significantly lower mean value compared to the 30-second moving average.

We observed a plateauing of VO_2 at maximal exercise in only 66 subjects (23.7%). When results were analysed according to the presence or absence of a plateau (Table 2), we observed that the percentage of correctly estimated average was significantly higher when a plateau was reached (92% for the 30-second moving average and 63% for the 30-second block average). When no plateau was reached, the proportion of correctly estimated peak value was lower: 73% for the 30-second moving average and only 24% for the 30-second block average).

Discussion

In this study, we found that the 30-second moving average significantly decreased the proportion of misinterpretations of peak measurements in children. For peak VO_2, the various averaging methods influenced the mean peak value by almost 100 ml·min^{-1}. The 30-second block average currently recommended approach in adults (ATS/ACCP, 2003; Balady et al., 2010) significantly underestimated peak VO_2 value while the 15-second and the 20-second moving averages overestimate this parameter commonly. Although the 15-second and the 20-second block averages offer better results, they still have a high proportion of misinterpretations for peak VO_2 measurements in children. Similar results were found for peak O_2pulse and peak RER. To our knowledge, we are the first to assess which averaging methods most accurately interpret peak values in children.

This study was undertaken after observing significant differences in peak values according to the averaging method and time interval when raw data from the metabolic carts were re-analysed in a previous study (Blanchard et al., 2018). Breath-by-breath system continuously measures airflow, which increase the probability of registering extreme data (Wasserman et al., 2005). To account for this, some averaging is necessary and the time interval must be such that the true physiological signal is distinguishable from the noise generated from large inter-breath variations (Potter et al., 1999). However, it has been recommended that the time intervals should be no larger than 30 seconds so that the data are not overly smoothed (Myers et al., 1990).

Although we showed that the 30-second moving average maximized accuracy in interpreting peak values, no averaging method was perfect, especially when no clear plateauing of the peak values was observed at maximal exercise. Using the 30-second moving average as the default method to establish peak CRF parameters value would likely increase accuracy, but we believe that manual review of the B-by-B data to ensure the peak value is correctly estimated is essential for accurate results.

Our results also highlight the hazard of using adults' guidelines in pediatric populations. We observed that only 24% of healthy children could reach a plateau of VO_2. Despite the fact that the criterion used (150 $ml \cdot min^{-1}$) could have had a significant impact on this result, similar proportions have been consistently observed by others (Armstrong et al., 1995; Armstrong et al., 1996; Armstrong et al., 1991). There are also many other physiological differences that will affect CRF results in children, compared to adults [relative increase in stroke volume and heart rate, effect of puberty, tolerance to lactate increase, tolerance to high intensity exercise and exertion, etc. (Paridon et al., 2006; Rowland, 2005; Turley and Wilmore, 1997)]. It is therefore essential that clinicians and researchers approach analysis and interpretation of children CPET by keeping in mind pediatric specificities such as averaging methods and reference values (Blanchard et al., 2018).

This study has limitations. The moving average method may not be available in all metabolic cart analysis commercial software. It is however not computationally complex and could be easily integrated in software updates. In cases where the block average method is the only option, time intervals shorter than 30 seconds decreased overestimation in our population. There is no true gold standard in determining which averaging methods were accurate or not. We relied on subjective assessment made by three observers with experience, but different background and expertise (one kinesiologist, one pediatric cardiologist, and one pediatric respirologist). Furthermore, this analysis was done on a subset of 47 participants. This was a multicenter study in which two different metabolic carts were used by two separate teams, with a possibility of variation for the collection of raw data. All tests were however done in research settings with strict standardization and calibration of equipment.

Conclusion

We showed that the 30-second block average currently recommended averaging method in adults increased the risk of misinterpretation of peak CRF values in children, and that using a moving average approach decreased underestimations and increased accuracy. Accurate interpretation of CRF in children may increase sensitivity and specificity to detect change in aerobic capacity over time and abnormal CRF in children with chronic disease.

References

Armstrong, N., Kirby, B., McManus, A., & Welsman, J. (1995). Aerobic fitness of prepubescent children. *Annals of human biology, 22*(5), pp. 427-441.

Armstrong, N., Welsman, J., & Winsley, R. (1996). Is peak VO2 a maximal index of children's aerobic fitness? *International journal of sports medicine, 17*(05), pp. 356-359.

Armstrong, N., Williams, J., Balding, J., Gentle, P., & Kirby, B. (1991). The peak oxygen uptake of British children with reference to age, sex and sexual maturity. *European journal of applied physiology and occupational physiology, 62*(5), pp. 369-375.

ATS/ACCP. (2003). ATS/ACCP statement on cardiopulmonary exercise testing. *American journal of respiratory and critical care medicine, 167*(2), p 211.

Balady, G. J., Arena, R., Sietsema, K., Myers, J., Coke, L., Fletcher, G. F., . . . Gulati, M. (2010). Clinician’s guide to cardiopulmonary exercise testing in adults. *Circulation, 122*(2), pp. 191-225.

Blanchard, J., Blais, S., Chetaille, P., Bisson, M., Counil, F. P., Girard, T. H., . . . Dallaire, F. (2018). New Reference Values for Cardiopulmonary Exercise Testing in Children. *Medicine & Science in Sports & Exercise*

Geithner, C. A., Thomis, M. A., Eynde, B. V., Maes, H. H. M., Loos, R. J. F., Peeters, M., . . . Beunen, G. P. (2004). Growth in Peak Aerobic Power during Adolescence. *Medicine & Science in Sports & Exercise, 36*(9), pp. 1616-1624. doi:10.1249/01.mss.0000139807.72229.41

Godfrey, S., Davies, C., Wozniak, E., & Barnes, C. A. (1971). Cardio-respiratory response to exercise in normal children. *Clinical science, 40*(5), pp. 419-431.

Mahon, A., & Marsh, M. (1992). Reliability of the rating of perceived exertion at ventilatory threshold in children. *International journal of sports medicine, 13*(08), pp. 567-571.

Myers, J., Arena, R., Franklin, B., Pina, I., Kraus, W., McInnis, K., & Balady, G. (2009). Recommendations for clinical exercise laboratories. *Circulation, 119*(24), pp. 3144-3161.

Myers, J., Walsh, D., Sullivan, M., & Froelicher, V. (1990). Effect of sampling on variability and plateau in oxygen uptake. *Journal of Applied Physiology, 68*(1), pp. 404-410.

Paridon, S. M., Alpert, B. S., Boas, S. R., Cabrera, M. E., Caldarera, L. L., Daniels, S. R., . . . Rhodes, J. (2006). Clinical stress testing in the pediatric age group: a statement from the American Heart Association Council on Cardiovascular Disease in the Young, Committee on Atherosclerosis, Hypertension, and Obesity in Youth. *Circulation, 113*(15), pp. 1905-1920.

Potter, C., Childs, D., Houghton, W., & Armstrong, N. (1999). Breath-to-breath “noise” in the ventilatory and gas exchange responses of children to exercise. *European journal of applied physiology and occupational physiology, 80*(2), pp. 118-124.

Prioux, J., Ramonatxo, M., Mercier, J., Granier, P., Mercier, B., & Prefaut, C. (1997). Changes in maximal exercise ventilation and breathing pattern in boys during growth: A mixed cross-sectional longitudinal study. *Acta Physiologica, 161*(4), pp. 447-458.

Rowland, T. (2005). *Children's exercise physiology*: Human Kinetics Champaign, IL.

Rowland, T. W., & Cunningham, L. N. (1992). Oxygen uptake plateau during maximal treadmill exercise in children. *Chest, 101*(2), pp. 485-489. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1735277

Turley, K. R., & Wilmore, J. H. (1997). Cardiovascular responses to treadmill and cycle ergometer exercise in children and adults. *Journal of Applied Physiology, 83*(3), pp. 948-957.

Wasserman, K., Hansen, J. E., Sue, D. Y., Stringer, W. W., & Whipp, B. J. (2005). Principles of exercise testing and interpretation: including pathophysiology and clinical applications. *Medicine & Science in Sports & Exercise, 37*(7), pp. 52-131.

Table 1. Mean values for estimated peak CRF parameters according to different averaging methods

Parameters and averaging method	Mean [95% CI]	Mean of differences ± SD	Proportion of correctly estimated average (%)
Peak VO_2 ($ml•min^{-1}$)			
30-second block average	2239 [2170 –2309]	Reference	29.0 %
30-second moving average	2287 [2218 –2359]	48.48 ± 47.07*	82.2 %*
20-second block average	2278 [2209 –2349]	39.40 ± 50.45*	68.1 %*
20-second moving average	2316 [2245 –2386]	76.06 ± 58.13*	53.9 %
15-second block average	2299 [2227 –2370]	59.04 ± 51.34*	80.9 %*
15-second moving average	2338 [2267 –2409]	98.51 ± 67.90*	36.2 %
Peak O_2pulse ($ml•beat^{-1}$)			
30-second block average	11.92 [11.53 –12.30]	Reference	57.5 %
30-second moving average	12.13 [11.75 –12.51]	0.212 ± 0.235*	97.9 %*
20-second block average	12.12 [11.74 –12.50]	0.203 ± 0.288*	63.8 %
20-second moving average	12.29 [11.91 –12.67]	0.373 ± 0.302*	34.0 %*
15-second block average	12.24 [11.85 –12.62]	0.322 ± 0.314*	59.6 %
15-second moving average	12.44 [12.06 –12.93]	0.525 ± 0.407*	29.8 %*
Peak RER			
30-second block average	1.19 [1.18 ± 1.20]	Reference	72.3 %
30-second moving average	1.21 [1.20 ± 1.22]	0.023 ± 0.023*	85.1 %*
20-second block average	1.20 [1.19 ± 1.22]	0.015 ± 0.018*	68.1 %
20-second moving average	1.22 [1.21 ± 1.23]	0.031 ± 0.025*	80.9 %
15-second block average	1.21 [1.20 ± 1.22]	0.024 ± 0.024*	59.6 %*
15-second moving average	1.23 [1.22 ± 1.24]	0.038 ± 0.030*	72.7 %

* Statistically significantly different ($p<0.05$) than the 30-second block average method

Table 2. Mean values for peak VO_2 according to different averaging methods and the presence or not of a plateau

Averaging method	Mean [95% CI]	Mean of differences ± SD	Proportion of correctly estimated average (%)
Plateau (*n*=66)			
30-second block average	2403 [2243 – 2562]	Reference	62.8 %
30-second moving average	2443 [2282 – 2604]	40.39 ± 45.05*	92.2 %*
No plateau (*n*=212)			
30-second block average	2189 [2112 – 2265]	Reference	24.4 %
30-second moving average	2240 [2162– 2317]	51.00 ± 47.51*	73.3 %*

* Statistically significantly different ($p<0.05$) than the 30-second block average method

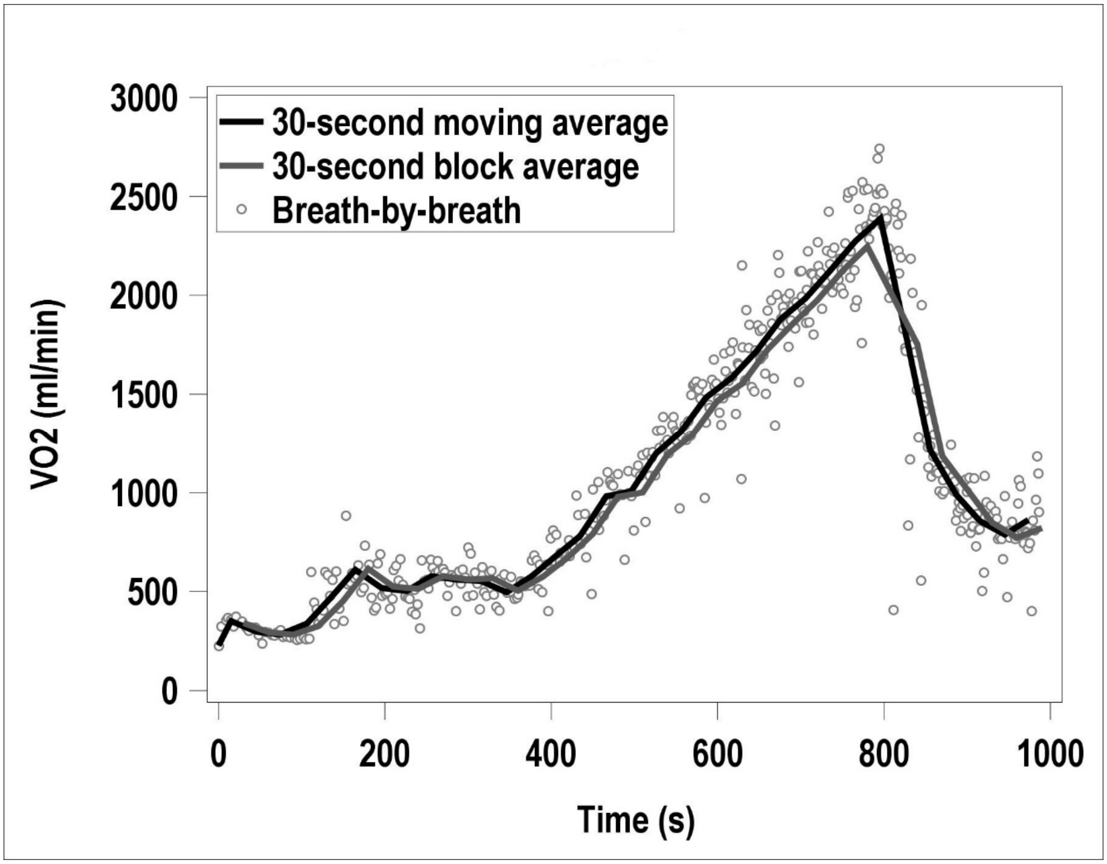

Figure 1. Example of an exercise test with important breath-by-breath (B-by-B) variability and its effect on a 30-second block average curve and a 30-second moving average curve.

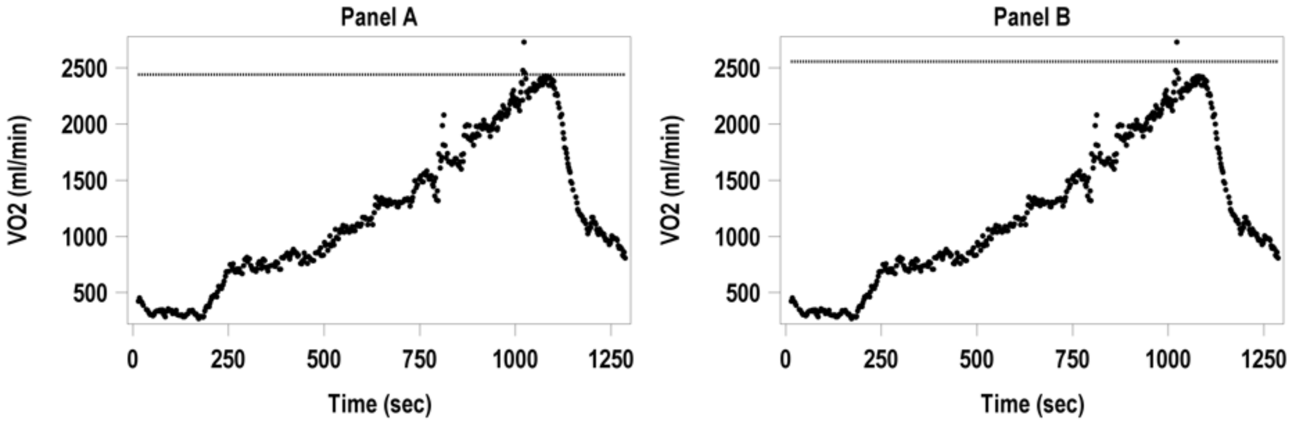

Figure 2. Examples of scatter plots of VO_2 according to time for the same subject. The dotted horizontal line represents the peak values estimated using two averaging methods. For this example, all reviewers considered that the peak VO_2 was accurate in Panel A and overestimated in Panel B.