

Reliability of a Graded Exercise Test for Assessing Recovery From Concussion

John J. Leddy, MD,* John G. Baker, PhD,† Karl Kozlowski, PhD,‡
Leslie Bisson, MD,* and Barry Willer, PhD§

Objective: To evaluate a graded treadmill test for retest reliability (RTR) and interrater reliability (IRR) in the evaluation of the physiologic effects of symptom exacerbation from concussion.

Design: Prospective case series (RTR) and blinded rater assessment of 10 actors portraying patients with and without symptom exacerbation (IRR).

Setting: University Sports Medicine Concussion Clinic.

Participants: For RTR, 21 refractory concussed patients (11 athletes and 10 nonathletes) and 10 healthy subjects; for IRR, 32 raters representing a variety of health care disciplines.

Intervention: For RTR, a Balke protocol treadmill test to symptom exacerbation before and after 2 to 3 weeks. For IRR, video recordings of actors during the treadmill test viewed by raters blinded to condition.

Main Outcome Measures: For RTR, agreement of the tests for maximal heart rate (HR), systolic blood pressure, diastolic blood pressure, and rating of perceived exertion. For IRR, presence or absence of symptom exacerbation and the symptom exacerbation HR.

Results: Raters achieved a sensitivity of 99% for identifying actors with symptom exacerbation and a specificity of 89% for ruling out concussion symptoms and agreed on 304 of 320 observations (accuracy of 95%). The intraclass correlation coefficient for the symptom exacerbation HR was large at 0.90 (95% confidence interval, 0.78-0.98). The treadmill test had good RTR for maximum HR (intraclass correlation coefficient, 0.79) but not for systolic blood pressure, diastolic blood pressure, or rating of perceived exertion.

Conclusions: The Balke exercise treadmill protocol has very good IRR and sufficient RTR for identifying patients with symptom exacerbation from concussion.

Key Words: traumatic brain injury, treadmill, symptoms, physiology, heart rate

(*Clin J Sport Med* 2011;21:89-94)

INTRODUCTION

Readiness of athletes to return to play (RTP) after concussion is a challenging issue for clinicians. Consensus conferences on concussion in sport^{1,2} advise that athletes limit cognitive and physical activity until asymptomatic at rest. Athletes may then RTP if their symptoms do not return during a stepwise program of increasing intensity of physical activity and contact, which includes light aerobic exercise, sport-specific exercise, noncontact training drills, and full-contact practice. It is suggested that athletes progress through 1 stage every 24 hours, but if there is exacerbation of symptoms during any step in the process, then athletes are advised to revert to the previous step. These guidelines provide a RTP model that is relatively consistent but which nevertheless is vague about the level and extent of exercise at each stage, relies heavily on clinical judgment, and has not been tested for reliability.

The Zurich Consensus Conference on Concussion in Sport specifically raised the issue of whether provocative exercise testing is useful in guiding the RTP decision in athletes who are asymptomatic at rest.² A provocative exercise test, much like that which cardiologists use to assess the functional condition of patients with heart disease, could potentially supplement or even replace the recommended stepwise approach to RTP. The requirement that the concussed athlete who is asymptomatic at rest exercise to maximum without exacerbation of symptoms before RTP recognizes the physiologic basis of concussion, which is supported by evidence of cerebral and whole-body physiological dysfunction after concussion.³ Provocative exercise testing provides the opportunity to determine the physiologic parameters of symptom exacerbation such as heart rate (HR) and blood pressure. This would allow the clinician not only to identify the athlete who is not ready to RTP but also to determine the level and extent of recovery of the concussed athlete.

If provocative exercise testing is to be recognized as a procedure for establishing the readiness of concussed athletes to RTP, then it must be demonstrated to be safe and reliable when used with athletes recovering from concussion. In a previous study, we safely used a provocative treadmill exercise test that followed a standard protocol in patients with ongoing symptoms due to post-concussion syndrome (PCS).⁴

Submitted for publication March 11, 2010; accepted September 20, 2010.
From the Departments of *Orthopaedics and the Sports Medicine Institute;
†Rehabilitation Medicine and Nuclear Medicine; ‡Exercise and Nutrition
Sciences; and §Psychiatry, University at Buffalo, State University of New
York, Buffalo, New York.

The authors report no professional relationships with companies or manufacturers who will benefit from the results of the present study.

Corresponding Author: John J. Leddy, MD, FACS, FACP, University Sports
Medicine, 160 Farber Hall, 3435 Main St, Buffalo, NY 14214 (e-mail:
leddy@buffalo.edu).

Copyright © 2011 by Lippincott Williams & Wilkins

We have since used provocative exercise testing on many athletes and nonathletes who experienced their concussion recently, who became asymptomatic at rest, and who wanted to know if they were safe for RTP or return to work. We have not encountered issues with safety because exercise testing is always terminated when symptom exacerbation occurs.

A provocative exercise test to assess concussion should meet minimum standards for reliability. There are 2 important aspects of reliability. The first is that a standardized approach to assessment should yield much the same results when a patient is assessed more than once. This type of reliability is called retest reliability (RTR). Retest reliability refers to the extent to which a test consistently and reproducibly measures the attributes it was intended to measure on repeated administration.⁵ The second important aspect of reliability is interrater reliability (IRR), which evaluates the consistency of results among different raters.⁵ Reliability is important for health care assessment procedures because the results of any assessment should be consistent from test to test and among different observers so as to inform clinical decisions rationally. Hence, the purpose of this study was to evaluate a standardized graded treadmill exercise test for RTR and IRR in determining RTP status of concussed athletes and nonathletes.

METHODS

Retest Reliability of the Treadmill Exercise Test

The sample included a consecutive sample of 21 athletes and nonathletes who presented to the University at Buffalo Concussion Clinic with symptoms of concussion. The diagnosis of concussion and of PCS, for those with prolonged symptoms, required 2 elements: fulfill the World Health Organization *International Classification of Disease, Tenth Revision* criteria⁶ of symptoms at rest for ≥ 6 weeks but < 52 weeks (by study physician interview) and demonstrate symptom exacerbation during 2 graded treadmill exercise tests. Only subjects at low cardiac risk according to the American College of Sports Medicine were deemed eligible.⁷ We used concussion subjects who had prolonged symptoms because our previous study showed that one would not expect spontaneous recovery over the period of test-retest assessment.⁴ This meant that a second test was unlikely to be influenced by patient recovery. The 21 concussed subjects (11 men and 10 women) were 29.8 ± 14.8 years old (range, 15-54 years) and were on average 33.2 weeks post injury (range, 6-36 weeks). Eleven of the 21 were athletes. Four of 11 athletes and 4 of 10 nonathletes had a history of 1 or more prior concussions. We also tested a group of 10 healthy subjects (4 men and 6 women) with an average age of 26.5 ± 8.2 years (range, 18-45 years). To control for the effects of previous physical activity on exercise test performance, all 10 controls were sedentary (defined as < 30 minutes of structured aerobic exercise < 2 days per week) for the previous 6 weeks. One of the 10 was an athlete. None of the healthy control subjects had a history of concussion. The healthy subjects were asked to exercise to exhaustion. The study was approved by the University at Buffalo Health Sciences Institutional Review Board, and all subjects provided written informed consent.

Subjects performed an incremental treadmill exercise test following a standard Balke protocol⁸ to the first sign of symptom exacerbation. The treadmill speed was set at 3.3 mph at 0.0% incline. After 1 minute, the grade increased to 2.0% while maintaining the same speed. At the start of the third minute and each minute thereafter, the grade increased by 1.0%, maintaining the speed at 3.3 mph. Blood pressure (BP, sphygmomanometer) was measured every 2 minutes, and HR (Polar 810i T61 HR Monitor; Polar, Inc, Kempele, Finland) and rating of perceived exertion (RPE, Borg scale) were measured every minute. Subjects were asked every minute whether they were experiencing any change in their health condition. The test was terminated at the report of exacerbation of concussion symptoms, and the HR, systolic BP (SBP), and diastolic BP (DBP) were recorded for the threshold of symptom exacerbation. After test termination, subjects were monitored for 60 minutes for safety purposes.

All subjects were retested 2 to 3 weeks after the first test using the exact same procedure. Concussed subjects and normal control subjects were asked to avoid exercise during this period.

Interrater Reliability of the Treadmill Exercise Test

Ten athletic actors were assigned symptoms and characteristics of 10 representative cases of patients assessed in the clinic. Actors were provided with a detailed description of the case they represented, including the symptoms and approximate HR for symptom exacerbation (if appropriate). Six actors (3 women) adopted characteristics of patients observed with symptom exacerbation and 4 actors (2 women) adopted characteristics of patients without symptom exacerbation. Video recordings of the 10 actors were produced for the purpose of creating educational materials for training physicians and athletic trainers to assess RTP using the Balke treadmill protocol. For this reason, it was essential that all actors accurately portray the experience of symptom exacerbation. It was also essential that the most complex cases be represented. For example, experience has taught us that previous symptom(s) reported at rest are rarely the symptom(s) that become exacerbated during the provocative exercise test. Fortunately, the experience of undergoing an exercise test in terms of HR and BP is likely the same whether someone is a patient or an actor. As well, exercise can induce symptoms independent of concussion,⁹ and 2 of the actors developed symptoms just from the exercise experience. A further complexity comes from the identification of the precise moment of symptom exacerbation. Some athletes hide their symptoms as long as possible, and one must rely on observation of facial expressions to determine approximate time of symptom exacerbation.

The 10 videos of actors completing the provocative exercise test were viewed by a total of 32 raters blinded to the conditions being portrayed. To reproduce as closely as possible the treadmill test performed on clinic patients, the raters were also provided with the HR and BP results that had been recorded every 2 minutes during the treadmill tests from the original cases. Each treadmill test was stopped for 1 of 2 reasons: (1) because of the development of concussion symptoms, or (2)

when actors reached voluntary exhaustion. Raters observed the entire video and rated each actor as with or without symptom exacerbation. For those actors rated with symptom exacerbation, raters recorded the HR at which symptom exacerbation occurred. In some instances, the actors continued the test despite the appearance of exercise-related symptoms, thus mimicking the clinical situation where exercise itself produced symptoms rather than the concussion. The raters were divided into 5 groups based on discipline: team physicians ($n = 5$), athletic trainers ($n = 18$), nurses ($n = 2$), emergency medical services technicians ($n = 5$), and chiropractors ($n = 2$). All raters had experience with assessment of concussion with professional and high level amateur athletes; however, none of the raters were specifically trained in the assessment of concussive symptoms using a graded exercise test.

Statistical Analysis

The intraclass correlation coefficient (ICC) informs us as to how likely it is that we would obtain exactly the same value on a second testing.¹⁰ For the treadmill RTR, the ICC (SPSS version 16.0; SPSS, Inc, Chicago, Illinois) was used to determine the extent of agreement of maximal HR, SBP, DBP, and RPE between the 2 tests conducted at the beginning (time 1) and at the end (time 2) of either 2 or 3 weeks. For IRR, the ICC was used to determine the extent of agreement among all the 32 raters and among the 5 rater groups. The presence or absence of symptoms from concussion during the test was used as the primary rating. The symptom onset HR of the actors portraying symptom exacerbation ($n = 6$) was also evaluated for agreement. A 2-way random effects model analysis of variance was specified with absolute agreement, rather than only consistency selected for the type of ICC, which is referred to as ICC (2, 1).¹⁰ The 95% confidence intervals (CIs) were calculated for each ICC value.

RESULTS

Subjects with concussion symptoms exercised 10 to 20 minutes (average, 15.4 minutes) to the point of symptom

exacerbation, whereas control subjects exercised from 12 to 20 minutes (average, 16.4 minutes) to reach voluntary exhaustion.

Retest Reliability of the Treadmill Exercise Test

Table 1 presents the means, SD, and ICCs with confidence intervals for the physiologic measures at maximum exercise (ie, symptom exacerbation for concussed subjects and exhaustion for normal controls). Using Cohen's criteria for effect size,¹¹ the ICCs of 0.79 and 0.64 for maximum HR for concussed subjects and healthy controls, respectively, represent moderate values (in reference to 0.2 for a small, 0.5 for a moderate, and 0.8 for a large value).

For maximum SBP, the concussed subjects showed much more variability, indicated by the ICC value of 0.37, a small value with a relatively wide confidence interval. The mean maximum SBP for the concussed subjects was significantly greater at time 2 versus time 1. The healthy controls showed much greater agreement for maximum SBP, as indicated by both the ICC value and the narrower confidence interval, even with a smaller sample size.

The ICC value of 0.42 for RPE among the concussed subjects is considered a small value, whereas 0.8 for the controls would be considered large. For DBP, the small value of 0.20 for concussed subjects contrasts with the moderate value of 0.52 for healthy controls.¹¹ Maximum DBP at times 1 and 2 increased for some and decreased for others, confirming our clinical observation that DBP is not reliably associated with symptom exacerbation.

Interrater Reliability of the Treadmill Exercise Test

Symptom Exacerbation Versus No Symptom Exacerbation

Overall, the 32 raters made 190 of 192 correct identifications of actors with symptom exacerbation for a sensitivity of 99% and 114 of 128 correct identifications of actors without symptom exacerbation for a specificity of 89%. Twenty of 32 raters were in complete agreement with each of their ratings

TABLE 1. Means, Standard Deviations, ICC, and 95% Confidence Intervals At Symptom Exacerbation Onset for Concussed Subjects ($n = 21$) and At Maximum Exercise for Healthy Controls ($n = 10$) At Time 1 and Time 2 for the Treadmill Test

| | Time 1, Mean (SD) | Time 2, Mean (SD) | Difference (Time 1 – Time 2), <i>F</i> Test | ICC (95% CI) | Difference (ICC > 0), <i>F</i> Test |
|--|-------------------|-------------------|--|-------------------|--|
| Maximum HR | | | | | |
| PCS | 144.7 (23.3) | 148.1 (28.6) | $F = 0.9, P < 0.36$ | 0.79 (0.55-0.91) | $F = 8.3, P < 0.00$ |
| Controls | 175.9 (15.3) | 175.7 (19.3) | $F = 0.0, P < 0.97$ | 0.64 (0.02-0.90) | $F = 4.2, P < 0.02$ |
| Maximum SBP | | | | | |
| PCS | 141.1 (15.8) | 150.6 (16.1) | $F = 6.4, P < 0.02$ | 0.37 (0.00-0.67) | $F = 2.5, P < 0.03$ |
| Controls | 149.8 (20.8) | 148.6 (21.3) | $F = 0.2, P < 0.71$ | 0.90 (0.66-0.98) | $F = 17.8, P < 0.00$ |
| Maximum DBP | | | | | |
| PCS | 79.7 (9.1) | 80.3 (9.4) | $F = 0.7, P < 0.80$ | 0.20 (–0.27-0.58) | $F = 1.5, P < 0.20$ |
| Controls | 74.0 (12.5) | 70.8 (9.0) | $F = 0.9, P < 0.37$ | 0.52 (–0.10-0.85) | $F = 3.1, P < 0.05$ |
| Maximum reported perceived exertion | | | | | |
| PCS | 12.9 (2.2) | 13.8 (2.5) | $F = 3.2, P < 0.09$ | 0.42 (0.03-0.71) | $F = 2.6, P < 0.02$ |
| Controls | 17.3 (2.1) | 16.6 (2.1) | $F = 3.1, P < 0.11$ | 0.80 (0.38-0.95) | $F = 10.4, P < 0.00$ |

HR, heart rate; DPB, diastolic blood pressure; ICC, intraclass correlation coefficient; PCS, post-concussion syndrome; SBP, systolic blood pressure.

of the 10 cases (200 ratings), whereas the remaining 12 were in agreement with 104 of 120 ratings. Overall, the agreement of ratings was 304 of 320 for an accuracy of 95%. The positive and negative predictive values were 93% and 98%, respectively. One of the actors portraying an athlete without symptom exacerbation but with exercise-induced headache accounted for 9 of the 16 incorrect ratings.

For all 32 raters, the overall ICC was significant at 0.83, a large effect size (95% CI, 0.70-0.94).¹¹ The value for the mean of the correlations between each pair of raters for all 32 raters (0.83) was similar to the estimate of the ICC value.

Table 2 presents the ICC values (with 95% CI) and values for the mean of the correlations of ratings between pairs of raters for the 5 groups. The values for 4 of the 5 groups of raters are considered to be large. The value for the EMS raters is a moderate value. The values for the mean interrater correlations for pairs of ratings are similar to the ICC values. The 95% CIs for the estimates of the ICC reflect relatively small sample sizes of the groups of raters.

Ratings of Symptom Exacerbation Heart Rate

The mean value for ratings of symptom exacerbation HR was different for each actor. The SDs of the estimates for each actor ranged from 3.6 to 7.9 with a mean of 5.3. For the 2 false-negative ratings, the mean HR rating for that actor was substituted for the missing value.

The ICC for the HR estimates by the 32 raters for the 6 actors with symptom exacerbation was 0.90 (95% CI, 0.78-0.98), a large value.¹¹ The value for the mean of the correlations between each pair of raters for all 32 raters was 0.93, similar to the estimate of the ICC value.

Table 3 presents the ICC values (with 95% CI) and values for the mean of the correlations of ratings between pairs of raters for the 5 groups. The value for the ICC for the 2 nurses is moderate, whereas the values for the other groups are large.¹¹ The 95% CIs for the estimates of the ICC, especially for the 2 nurses, reflect the relatively small sample sizes of each group. The values for the mean interrater correlations for pairs of ratings are similar to the ICC values. Inspection of the ratings for the nurse group showed a very low HR estimate for 1 of the cases.

DISCUSSION

We used the Balke treadmill protocol to assess symptom exacerbation in persons with concussion and found it to be

a safe procedure that increases exercise intensity in a gradual manner. Our data show that this provocative treadmill test has a high degree of IRR (95%) among various groups of raters for distinguishing symptom exacerbation in those not recovered from concussion from those ready to RTP. The raters achieved a sensitivity of 99% for identifying those who portrayed symptom exacerbation and a specificity of 89% for identifying those who did not portray symptom exacerbation. All raters had considerable experience with assessment of sport-related concussion. The treadmill test had good RTR for assessment of maximum HR in concussed subjects and moderate reliability for maximal SBP; however, concussed subjects showed considerably more variability in SBP and in RPE than healthy controls. Diastolic BP is usually stable or declines as exercise intensity and duration increase in nonconcussed subjects.¹² The DBP was not a reliably reproducible measure, especially among concussed subjects, likely because concussed subjects were in a state of sympathetic nervous system activation that is commonly observed after concussion.³ This could affect arterial compliance, which would be reflected in an abnormal exercise DBP response. The exercise treadmill test used in this study thus has very good interrater and sufficient maximum HR RTR for identifying patients with symptom exacerbation due to concussive effects. What this means for the clinician is that she/he can be confident that the treadmill test is consistent in its ability to measure HR in concussed athletes on repeated administration and that different observers are able to reliably identify the symptom exacerbation threshold in athletes who are not physiologically ready to RTP. The physiologic data may be useful to clinicians if, for example, the HR is observed to be unusually high during a low level of exertion. In this circumstance, the clinician should consider that the athlete's physiology has not yet returned to a pre-injury level, and therefore, the athlete may need more time to recover to physiological homeostasis. Reliability is important because it means the clinician has confidence that the test results reflect the state of the patient and not variability inherent in repeated testing.

The research on reliability of physiological variables during repeated graded treadmill exercise tests is limited. Fielding et al¹³ found good RTR for measures of maximum HR, BP, and RPE over 5 repeated tests using the Bruce protocol. Using a modified Balke treadmill protocol, Tonino and Driscoll¹⁴ reported no significant differences in maximum oxygen uptake between 3 repeated exercise tests, although they did not find good reliability for submaximal HR. Fernhall

TABLE 2. Intraclass Correlation Coefficient (ICC), Mean Interrater Correlation, and Confidence Intervals for Ratings of Symptom Exacerbation Versus No Symptom Exacerbation (n = 10) for 5 Groups of Raters

| | ICC | Difference From Zero | Mean Interrater Correlation | Confidence Interval (95%) |
|----------------------------|-------|-------------------------------|-----------------------------|---------------------------|
| Team physicians (n = 5) | 0.80 | $F(9, 36) = 19.6, P < 0.000$ | 0.79 | 0.59-0.93 |
| Athletic trainers (n = 18) | 0.85 | $F(9, 153) = 99.4, P < 0.000$ | 0.84 | 0.71-0.95 |
| Nurses (n = 2) | 0.80 | $F(9, 9) = 9.0, P < 0.002$ | 0.80 | 0.41-0.95 |
| EMS (n = 5) | 0.70 | $F(9, 36) = 12.8, P < 0.000$ | 0.69 | 0.46-0.90 |
| Chiropractors (n = 2) | 1.00* | — | 1.00 | 1.00-1.00 |

*Both Chiropractors were in complete agreement with their ratings.

TABLE 3. Intraclass Correlation Coefficient (ICC), Mean Interrater Correlation, and Confidence Intervals for Ratings of Symptom Exacerbation Threshold HR (n = 6) for 5 Groups of Raters

| | ICC | Difference from Zero | Mean Interrater Correlation | Confidence Interval (95%) |
|----------------------------|------|-------------------------------|-----------------------------|---------------------------|
| Team physicians (n = 5) | 0.93 | $F(5, 20) = 54.9, P < 0.000$ | 0.95 | 0.79–0.99 |
| Athletic trainers (n = 18) | 0.93 | $F(5, 85) = 233.5, P < 0.000$ | 0.94 | 0.82–0.99 |
| Nurses (n = 2) | 0.69 | $F(5, 5) = 5.2, P < 0.048$ | 0.72 | –0.09–0.95 |
| EMS (n = 5) | 0.87 | $F(5, 20) = 40.5, P < 0.000$ | 0.91 | 0.65–0.98 |
| Chiropractors (n = 2) | 0.97 | $F(5, 5) = 78.1, P < 0.000$ | 0.91 | 0.85–1.00 |

et al¹⁵ reported high reliability for maximum HR in 14 adolescents with developmental delay who completed 2 Balke–Ware treadmill protocols 1 week apart. Finch et al¹⁶ used 2 blinded raters to assess the IRR of the distance walked on a constant speed treadmill test to a Borg “hard” effort level 3 times, a week apart, in 15 subjects with post-polio syndrome. The RTR was 0.85, and the IRR was 0.91 with adherence to a standardized protocol. Lamb et al¹⁷ assessed RTR of the Borg scale in 16 male athletes during 2 identical treadmill tests over a period of 2 to 5 days. The Borg scale reliability decreased as exercise intensity increased, but the HR responses at each treadmill stage did not vary significantly over trials. Thus, there is evidence that treadmill test physiological measurements are reliable in different patient populations. There are no data to our knowledge, however, on the reliability of a symptom exacerbation threshold during these tests.

Although symptom reports, skilled clinical interviews, and neuropsychological (NP) testing are essential for the evaluation and monitoring of patients after concussion, the primary component of the RTP decision process is the ability of the athlete to exercise to the level of his/her sport without exacerbation of symptoms. According to the Zurich guidelines, the assessment process should occur over several days, although they allow for same-day RTP for adult athletes under certain circumstances.² Use of a standardized and reliable provocative exercise protocol could increase the accuracy of the RTP decision and the confidence of the clinician and athlete with the decision.

We propose that the response to incremental (provocative) exercise testing can help with the RTP decision. Provocative exercise testing should only be performed, however, in patients whose rest symptoms have resolved and for whom a determination is being made as to fitness to return to sport or activity⁴ because experimental animal data show that premature voluntary exercise within the first week after concussion impairs cognitive performance,¹⁸ an observation corroborated by some human data.¹⁹ Provocative exercise testing should therefore be administered at the appropriate time after concussion.

This study represents an attempt to standardize the process of provocative testing by determining the reproducibility of a treadmill test that challenges the physiology of patients with concussion. The primary limitation of the study is that there is no established “gold standard” for the diagnosis of concussion. We are thus left with a de facto standard that athletes have recovered once they can exercise to maximum levels without developing symptoms.^{1,2} Another possible weakness is the use of actors performing the treadmill test in

the assessment of IRR. We originally intended to produce videos of actual cases but found ourselves confronted by technical and ethical limitations, including interactions between the assessors and the patients, which made it obvious that the patient had PCS. We chose to use actors primarily because it allowed us to create the videos during the same period with consistency in terms of data (HR, BP, etc) presentation. Others have successfully used actors to establish IRR of health indicators for most of the same reasons.²⁰ We nevertheless recognize that actual patients may vary in terms of how symptom exacerbation is experienced and therefore may influence reliability across raters. We were encouraged, however, by the fact that raters made some errors with cases, where symptoms developed as a result of the exercise itself at exercise testing completion. Finally, even though the exercise protocol and measurements (potential sources of random error) were controlled, considerable random error due to biological variation can occur when conducting repeated physiological tests.²¹ Although the natural history of PCS is one of gradual resolution over time, we attribute the variability of the HR and BP responses in the concussed subjects to ongoing physiologic dysfunction³ because our previous research has shown that, absent treatment, the period between the 2 treadmill tests was too short to significantly affect symptom reports or exercise tolerance in this group.⁴

The clinical use of a standardized and reliable provocative treadmill exercise test is illustrated by the example of a high school athlete who has sustained his second concussion within the previous year and who presents to your office for clearance to resume contact sport. His symptoms were of greater severity and of longer duration after the most recent concussion. He reports that after 12 days of cognitive and physical rest, he is asymptomatic at rest and he has reached baseline performance on a computerized NP test. His mother confirms that he has not reported any symptoms, and she thinks he is back to normal, but she wants to know if he is ready to play in the next game, which is in 5 days. Using the Balke protocol on your office treadmill, the athlete exercises at progressively increasing exercise intensities to exhaustion without symptom exacerbation. You conclude, and inform the athlete and his mother, that he is physiologically recovered and can safely RTP. Conversely, if the athlete developed signs or symptoms that stopped the test before peak exertion, you have objective information that he is not physiologically ready and will need more recovery time. The most commonly reported symptoms indicating that the concussion is not resolved are worsening headache and/or a sensation that the head feels “full.” A comparison of the HR at the point of symptom

exacerbation to the athlete's theoretical maximum HR gives you a good indication of how close the athlete is to full recovery. If close to full recovery, the test could be repeated in a few days. The test can be performed in a physician's office, an athletic training or physical therapy facility, hospital clinic, or a health club, provided that staff have been trained in treadmill test administration and that there is medical supervision in proximity.

In conclusion, the Balke exercise treadmill test has very good IRR and sufficient maximum HR RTR for identifying patients with symptom exacerbation due to concussion. Symptom reports alone are nonspecific and highly variable, and NP test performance at rest improves in most patients, even in those with ongoing symptoms. The symptom exacerbation threshold during the exercise test in our opinion adds an important and more objective element to help clinicians make the RTP decision in athletes. The Balke treadmill test is safe (it has also been used to assess aerobic capacity in patients with cardiac⁸ and orthopedic²² limitations), can be performed by physicians and athletic trainers with equipment readily available, and does not require specialist interpretation. It thus may be applied in a variety of clinical settings with good reliability, relative ease, and at reasonable cost.

ACKNOWLEDGMENTS

The authors are very appreciative of the financial assistance of the Robert Rich Family Foundation, the Buffalo Sabres Foundation and the New York State Athletic Trainers' Association.

REFERENCES

1. McCrory P, Johnston K, Meeuwisse W, et al. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. *Clin J Sport Med*. 2005;15:48–55.
2. McCrory P, Meeuwisse W, Johnston K, et al. Consensus statement on concussion in sport, 3rd International Conference on Concussion in Sport, held in Zurich, November 2008. *Clin J Sport Med*. 2009;19:185–200.
3. Leddy JJ, Kozlowski K, Fung M, et al. Regulatory and autoregulatory physiological dysfunction as a primary characteristic of post concussion syndrome: implications for treatment. *Neurorehabilitation*. 2007;22:199–205.
4. Leddy JJ, Kozlowski K, Donnelly JP, et al. A preliminary study of subsymptom threshold exercise training for refractory post-concussion syndrome. *Clin J Sport Med*. 2010;20:21–27.
5. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. *Stroke*. 2007;38:1091–1096.
6. Boake C, McCauley SR, Levin HS, et al. Diagnostic criteria for post-concussional syndrome after mild to moderate traumatic brain injury. *J Neuropsychiatry Clin Neurosci*. 2005;17:350–356.
7. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. In: Franklin BA, Whaley MH, Howley ET, eds. 6th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2006:22–29.
8. Ades PA, Grunwald MH. Cardiopulmonary exercise testing before and after conditioning in older coronary patients. *Am Heart J*. 1990;120:585–589.
9. Gaetz MB, Iverson GL. Sex differences in self-reported symptoms after aerobic exercise in non-injured athletes: implications for concussion management programmes. *Br J Sports Med*. 2009;43:508–513.
10. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
11. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
12. Palatini P, Mos L, Mormino P, et al. Intra-arterial blood pressure monitoring in the evaluation of the hypertensive athlete. *Eur Heart J*. 1990;11:348–354.
13. Fielding RA, Frontera WR, Hughes VA, et al. The reproducibility of the Bruce protocol exercise test for the determination of aerobic capacity in older women. *Med Sci Sports Exerc*. 1997;29:1109–1113.
14. Tonino RP, Driscoll PA. Reliability of maximal and submaximal parameters of treadmill testing for the measurement of physical training in older persons. *J Gerontol*. 1988;43:M101–M104.
15. Fernhall B, Millar AL, Tymeson GT, et al. Maximal exercise testing of mentally retarded adolescents and adults: reliability study. *Arch Phys Med Rehabil*. 1990;71:1065–1068.
16. Finch LE, Venturini A, Mayo NE, et al. Effort-limited treadmill walk test: reliability and validity in subjects with postpolio syndrome. *Am J Phys Med Rehabil*. 2004;83:613–623.
17. Lamb KL, Eston RG, Corns D. Reliability of ratings of perceived exertion during progressive treadmill exercise. *Br J Sports Med*. 1999;33:336–339.
18. Griesbach GS, Hovda DA, Molteni R, et al. Voluntary exercise following traumatic brain injury: brain-derived neurotrophic factor upregulation and recovery of function. *Neuroscience*. 2004;125:129–139.
19. Majerske CW, Mihalik JP, Ren D, et al. Concussion in sports: post-concussive activity levels, symptoms, and neurocognitive performance. *J Athl Train*. 2008;43:265–274.
20. Rosengren DB, Hartzler B, Baer JS, et al. The video assessment of simulated encounters-revised (VASE-R): reliability and validity of a revised measure of motivational interviewing skills. *Drug Alcohol Depend*. 2008;97:130–138.
21. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998;26:217–238.
22. Hagberg JM. Exercise assessment of arthritic and elderly individuals. *Baillieres Clin Rheumatol*. 1994;8:29–52.