

# Original Article

## Accuracy of the stair-climbing test using maximal oxygen uptake as the gold standard\*

Daniele Cristina Cataneo<sup>1</sup>, Antonio José Maria Cataneo<sup>2</sup>

### Abstract

**Objective:** To determine the accuracy of the variables related to the fixed-height stair-climbing test (SCT) using maximal oxygen uptake ( $\dot{V}O_2$  max) as the gold standard. **Methods:** The SCT was performed on a staircase consisting of 6 flights (72 steps; 12.16 m total height), with verbal encouragement, in 51 patients. Stair-climbing time was measured, the variables 'work' and 'power' also being calculated. The  $\dot{V}O_2$  max was measured using ergospirometry according to the Balke protocol. We calculated the Pearson linear correlation (r), as well as the values of p, between the SCT variables and  $\dot{V}O_2$  max. To determine accuracy, the  $\dot{V}O_2$  max cut-off point was set at 25 mL/kg/min, and individuals were classified as normal or altered. The cut-off points for the SCT variables were determined using the receiver operating characteristic curve. The Kappa statistic (k) was used in order to assess concordance. **Results:** The following values were obtained for the variable 'time': cut-off point = 40 s; mean =  $41 \pm 15.5$  s;  $r = -0.707$ ;  $p < 0.005$ ; specificity = 89%; sensibility = 83%; accuracy = 86%; and  $k = 0.724$ . For 'power', the values obtained were as follows: cut-off point = 200 w; mean =  $222.3 \pm 95.2$  w;  $r = 0.515$ ;  $p < 0.005$ ; specificity = 67%; sensibility = 75%; accuracy = 71%; and  $k = 0.414$ . Since the correlation between the variable 'work' and  $\dot{V}O_2$  max was not significant, that variable was discarded. **Conclusion:** Of the SCT variables tested, using  $\dot{V}O_2$  max as the gold standard, the variable 'time' was the most accurate.

**Keywords:** Exercise test; Respiratory function tests; Heart function tests; Spirometry; Ergometry.

\* Study carried out at the *Júlio de Mesquita Filho Universidade Estadual Paulista* – UNESP, Paulista State University – at Botucatu School of Medicine *Hospital das Clínicas*, Botucatu (SP) Brazil.

1. PhD, Assistant Professor in the Department of Thoracic Surgery at the *Júlio de Mesquita Filho Universidade Estadual Paulista* – UNESP, Paulista State University – at Botucatu School of Medicine, Botucatu (SP) Brazil.

2. Tenured Professor, Head of the Department of Thoracic Surgery at the *Júlio de Mesquita Filho Universidade Estadual Paulista* – UNESP, Paulista State University – at Botucatu School of Medicine, Botucatu (SP) Brazil.

Correspondence to: Daniele Cristina Cataneo, Departamento de Cirurgia e Ortopedia – UNESP – Distrito de Rubião Jr., s/n, CEP 18618-970, Botucatu, SP, Brasil. Phone 55 14 38116091. E-mail: dcataneo@fmb.unesp.br/danykt@yahoo.com

Submitted: 16 February 2006. Accepted, after review: 3 August 2006.

## Introduction

One of the various clinical applications of cardiopulmonary exercise testing (CPET) is to evaluate patient physical condition, which is equivalent to the cardiopulmonary reserve, prior to surgery.<sup>(1-4)</sup> Such testing has come to be considered the gold standard for predicting surgical risk.<sup>(4-8)</sup> Within this context, the best variable used is maximal oxygen uptake ( $\dot{V}O_2$  max), measured via spirometry during exercise. The  $\dot{V}O_2$  max corresponds to the highest volume of oxygen consumed during exercise and is therefore considered a maximal test. The  $\dot{V}O_2$  max seems to be the best indicator of exercise capacity.<sup>(6)</sup> However, despite being very useful, spirometry during exercise is available in few hospitals.

Another type of CPET, which is considered submaximal and can also be used for this purpose, is the stair-climbing test (SCT), which does not lead the individual to maximal exertion. The SCT was initially described in 1955.<sup>(9)</sup> However, to date, it has not been properly standardized. The most significant studies of SCTs started, in the 1960s, with their standardization in terms of the number of flights.<sup>(10,11)</sup> In the mid 1980s, there were studies that compared the SCT and spirometry, questioning the latter for being a static test and for only evaluating pulmonary function.<sup>(12,13)</sup> In the 1990s, the SCT parameters were correlated with  $\dot{V}O_2$  max, showing that the results were comparable in predicting post-operative complications although there was still some difficulty in determining which SCT variable to use.<sup>(14-17)</sup> At the end of the last century, one group of authors<sup>(17)</sup> measured the height of the staircase in meters rather than in number of steps; however, no special attention was given to the variable 'time' since, in almost all previous studies, patients were oriented to climb the stairs at their own pace, with no verbal encouragement.<sup>(11-18)</sup>

According to the literature, in most SCT studies, the height of the staircase is not duly determined, stair-climbing time is not standardized, and the variable 'power' is not properly considered. Consequently, there is no clearly defined SCT, nor can we draw comparisons among those studies, due to their differences in methodology. In view of this, and since the test is affordable and accessible to all the population, as well as the fact that it can be carried out in any hospital that has a staircase of the minimum height, we believe that the SCT could

be used in Brazil, since spirometry during exercise testing is not widely available. Therefore, our objective was to standardize the SCT in our facility by defining the total height of the staircase in order to correlate the various variables obtained in this test with  $\dot{V}O_2$  max and subsequently test the accuracy of the variables with the highest correlation, using  $\dot{V}O_2$  max as the gold standard.

## Methods

The study was approved by the Ethics in Research Committee of the Botucatu School of Medicine *Hospital das Clínicas*, in Botucatu, Brazil. At the study outset, we contacted patients who were scheduled for spirometry. Those who agreed to participate in the study gave written informed consent. Eligible patients were all of those who had previously scheduled spirometry, whatever the indication (either clinical or preoperative), and had no difficulties in walking. The exclusion criteria were the same as those for spirometry during exercise<sup>(19,20)</sup>: acute disease, systolic arterial pressure greater than 200 mmHg and diastolic arterial pressure greater than 110 mmHg, decompensated heart failure, infarction less than 40 days prior, decompensated COPD, complete obstruction of the left branch on the electrocardiogram, and difficulties in walking (orthopedic, neurological, or vascular changes). Those patients who were unable to climb all the steps of the staircase or were unable to undergo spirometry during exercise were also excluded. A brief anamnesis was then carried out, together with a complete physical examination and an electrocardiogram. Subsequently, the patients underwent the SCT and, on another date, spirometry during exercise.

The SCT was performed in the shade using a staircase consisting of 6 flights and having a 30° inclination. Each flight had 12 steps (totaling 72 steps), and each step measured 16.9 cm (total height, 12.16 m). The patients were instructed to climb all the steps in the shortest time possible, with standardized verbal encouragement at each flight, always given by the same examiner. Between the flights, the patients took two or three steps on level ground, trying to maintain their speed, at which point the examiner asked them if everything was fine. The test was interrupted only due to fatigue, limiting dyspnea, chest pain, or exhaustion. The time

taken to climb the total height was designated the stair-climbing time and is expressed in seconds (s). Each patient underwent the test only once.

The work performed by the patient in order to climb the staircase was calculated, in joules, using the following formula:

$$W = m \times g \times h \quad (1)$$

where  $W$  is work,  $m$  is patient weight in kilograms,  $g$  is the acceleration due to gravity ( $9.8 \text{ m/s}^2$ ), and  $h$  is the height of the staircase in meters (12.16 m). We obtained the values for the variable 'power', expressed in watts (w), by dividing the work performed by the stair-climbing time.

The  $\dot{V}O_2$  max was measured using a Quinton stress test system (Q4500, Quinton Instruments, Seattle, WA, USA) coupled to a treadmill, in a climate-controlled environment.<sup>(19,20)</sup> Heart rate and respiratory frequency were monitored throughout the test, as were arterial blood pressure, oxygen saturation, and electrocardiogram with 12 leads. Although all of the variables of the spirometry during exercise test were measured, the only one used was  $\dot{V}O_2$  max, expressed in mL/kg/min. We used the Balke protocol,<sup>(21)</sup> which is an incremental protocol and is indicated for individuals with comorbidities. The test would be interrupted if there were the following developments: a drop in systolic arterial pressure greater than 10 mmHg in relation to rest; angina; symptoms related to the central nervous system (ataxia, dizziness, or lightheadedness); signs of low perfusion (cyanosis or pallor); technical difficulties in monitoring the electrocardiogram or the arterial blood pressure; sustained ventricular tachycardia; ST-segment elevation greater than 2 mm; ST-segment depression greater than 3 mm; patient request; fatigue; dyspnea; wheezing; cramps; claudication; bundle branch block or slowed conduction; increasing chest pain; or hypertensive response.

For the statistical analysis, we used the Pearson linear correlation coefficient ( $r$ ) to determine the correlation between  $\dot{V}O_2$  max and the SCT variables, calculating the respective  $p$  values. Sensitivity, specificity, accuracy, and likelihood ratio were calculated for the SCT variables that showed a significant correlation with  $\dot{V}O_2$  max, which was used as the gold standard. The  $\dot{V}O_2$  max cut-off point used in order to classify patients as normal or altered was set at 25 mL/kg/min. The cut-off points for the SCT variables were determined using the receiver oper-

ating characteristic curve.<sup>(22)</sup> The Kappa statistic ( $k$ ) was used in order to assess concordance.<sup>(23,24)</sup>

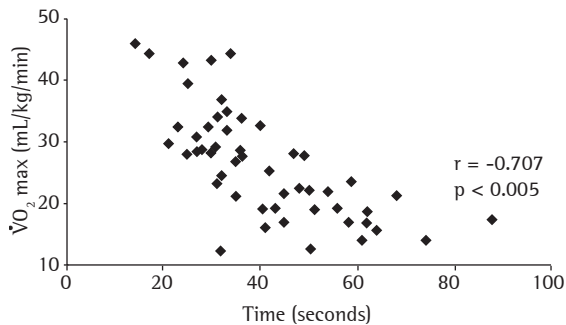
## Results

The SCT was performed in ninety-eight volunteer patients. One patient was excluded for not being able to climb the entire 12.16-m staircase. Spirometry during exercise was performed in fifty-one of the patients (thirty males and twenty-one females, ranging in age from 18 to 77 years; mean,  $52 \pm 16$  years). Since  $\dot{V}O_2$  max was considered the gold standard, forty-six patients were excluded for not reporting for the spirometry during exercise test, and one patient was excluded for not being able to walk on the treadmill at minimum speed. None of the patients needed to interrupt the SCT. The  $\dot{V}O_2$  max ranged from 12.4 mL/kg/min to 46.1 mL/kg/min (mean,  $26.5 \pm 8.85$  mL/kg/min).

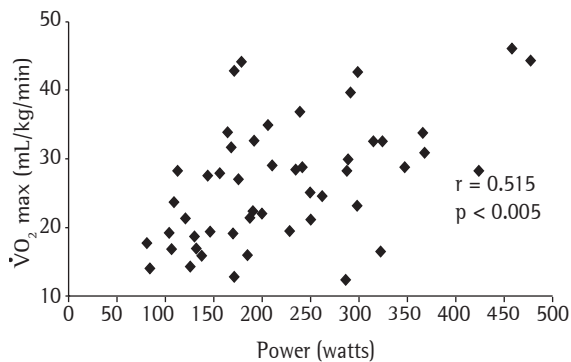
The variable 'time' ranged from 14 to 88 s (mean,  $41 \pm 15.5$  s). There was a significant negative linear correlation between  $\dot{V}O_2$  max and the variable 'time' ( $r = -0.707$ ,  $p < 0.005$ ) (Figure 1). The variable 'power' ranged from 83.3 to 476.7 w (mean,  $222.3 \pm 95.2$  w). There was a significant linear correlation between  $\dot{V}O_2$  max, ( $r = 0.515$ ,  $p < 0.005$ ) (Figure 2). Since there was no significant correlation between work and  $\dot{V}O_2$  max, that SCT variable was not used to calculate accuracy and was discarded. The cut-off points for the variables 'time' and 'power', determined using the receiver operating characteristic curve, were 40 s and 200 w, respectively. For the variable 'time', we obtained a sensitivity of 83%, a specificity of 89%, an accuracy of 86%, and  $k = 0.724$  (good concordance) (Table 1). The positive likelihood ratio was 7.54, and the negative likelihood ratio was 0.19. For the variable 'power', we obtained a sensitivity of 75%, a specificity of 67%, and accuracy of 71%, and  $k = 0.414$  (moderate concordance) (Table 2). The positive likelihood ratio was 2.27, and the negative likelihood ratio was 0.37.

## Discussion

The use of CPET prior to surgery can detect changes in oxygen transport that would only be discovered if metabolic demand increased in the intra-operative or post-operative periods. Spirometry during exercise, which can be used for this purpose, is available in few health care facilities, and, in addi-



**Figure 1** – Linear correlation between  $\dot{V}O_2$  max and stair-climbing time.



**Figure 2** – Linear correlation between  $\dot{V}O_2$  max and stair-climbing power.

**Table 1** – Distribution of the individuals by  $\dot{V}O_2$  max (gold standard) and by stair-climbing time, considering sensitivity, specificity, accuracy, and Kappa statistic concordance.

Time	$\dot{V}O_2$ max		Total
	Altered	Normal	
Altered	20	3	23
Normal	4	24	28
Total	24	27	51

Sensitivity = 83%; Specificity = 89%; Accuracy = 86%; Kappa = 0.724.

tion to the high cost of the equipment, there are few professionals who are trained in its performance, especially in severe patients. Furthermore, patients need time to learn to walk on the treadmill and to overcome their phobia so that they can cope with the mask. Therefore, despite being highly efficient and considered the gold standard for predicting surgical risk by most authors,<sup>[3-6]</sup> CPET is far from being feasible, especially given the current status of the public health care system in Brazil.

**Table 2** – Distribution of the individuals by  $\dot{V}O_2$  max (gold standard) and by stair-climbing power, considering sensitivity, specificity, accuracy and Kappa statistic concordance.

Power	$\dot{V}O_2$ max		Total
	Altered	Normal	
Altered	18	9	27
Normal	6	18	24
Total	24	27	51

Sensitivity = 75%; Specificity = 67%; Accuracy = 71%; Kappa = 0.414.

The ideal test for pre-operative investigation should be simple, cheap, and widely available. The SCT presents these characteristics and, still in the new millennium, has been tested in rich countries in order to evaluate cardiopulmonary training.<sup>(18,25,26)</sup> Considering that the presence of cardiovascular or pulmonary diseases limits exercise capacity, it is not surprising that patients with cardiopulmonary disorders have difficulty in climbing stairs, and that the degree of limitation is proportional to the degree of impairment of the cardiopulmonary function.<sup>(16)</sup> Nor is it surprising that, on the contrary, patients who climb multiple floors in quick steps and have no symptoms have a good cardiopulmonary reserve.<sup>(14)</sup> However, it is essential to use a precise, universal, and accurately determined variable that can make it possible to standardize the SCT, as has already been done for the walk test. The attempts to stratify the post-operative complications only by the number of floors or steps climbed have been frustrating at times.<sup>(14)</sup> In addition, the step is not a universal unit of measurement. Therefore, the ideal would be to measure the height reached in meters rather than in flights or floors. Similarly, we would have difficulty in applying the SCT if height was considered as a variable since very tall stairs, which are not available in all facilities including ours, would be needed. Nevertheless, if we consider a minimum height of 12 m as a constant, it would be possible to use time as a variable. Lesser heights might not be as useful, since it has been shown that 50% of the patients that are unable to climb at least 12 m present complications.<sup>(18)</sup>

The variable 'time' must be accurately determined and there must be verbal encouragement during the stair climbing so as to prevent the patients from walking at their own pace. Measuring the time without giving any verbal encouragement, in addition to not being a means of accurately

assessing the real physical capacity of the individual, also alters the other SCT parameters, such as power and estimated  $\dot{V}O_2$ , that have been used by other authors to date.<sup>(14,18)</sup> Ultimately, it is necessary to use the variable 'time' in order to calculate these other variables. It would be the same as applying the walk test with no verbal encouragement: a patient capable of walking 600 m might walk only 300 m.

As for the variable 'power', other authors<sup>(14)</sup> have used the following formula to calculate the amount of work done in order to climb a staircase:

$$W = \text{height of the step (m)} \times \text{number of steps} / \text{t (min)} \times \text{mass (kg)} \times 0.1635 \quad (2)$$

where t is time. In this formula, what is actually calculated is not work but rather power, expressed in watts. This is the formula for power seen in a more complicated way. However, eleven years later, there are authors<sup>(18)</sup> who continue to refer to this variable as 'work' rather than using the classical formula for power. It is not necessary to include the variable 'time' in order to calculate work. Indeed, the variable 'work' was not found to correlate with  $\dot{V}O_2$  max but with power, which depends on time. In order to calculate the estimated  $\dot{V}O_2$  max for the SCT, those same studies used the variable 'power'.

$$\dot{V}O_2 \text{ (mL/min)} = 5.8 \times m \text{ (kg)} + 151 + (10.1 \times P) \quad (3)$$

where P is power. Therefore, time remains an important variable in estimating  $\dot{V}O_2$  max. Time must be measured as accurately as are mass and height, and, if there is no verbal encouragement, it will never be sufficient. As can be seen in Table 3, the authors of the studies in the literature have not concerned themselves with verbal encouragement or with the accurate measurement of time.

When we designed this experiment, we believed that the correlation with power was much stronger than the correlation with time since the authors of the studies evaluated, despite misnaming the former, gave it considerable weight. Ultimately, we found that the best parameter was the variable 'time', which is obtained in a direct manner. Nevertheless, power must be determined, since it is the basis for estimating  $\dot{V}O_2$  max.

Based on the results obtained under our experimental conditions, we can state that the patients who took less than 40 s to climb the 12 m have a high probability of presenting a  $\dot{V}O_2$  max above

**Table 3** – Dimensions of the staircase by author.

Authors	St h	St N	St/f	St/F	f	F	t	VE
Van Nostrand <sup>(11)</sup>	n	n	n	n	2	n	n	n
Colman <sup>(12)</sup>	n	n	n	n	2	n	n	n
Bolton <sup>(13)</sup>	n	127	n	n	n	5	y	y
Olsen <sup>(14)</sup>	17.40	125	n	n	n	5	n	n
Holden <sup>(15)</sup>	17.00	44	11	n	n	n	n	n
Pollock <sup>(16)</sup>	18.50	n	n	18	n	5	n	n
Pate <sup>(17)</sup>	17.50	n	n	21	n	5	y	n
Girish <sup>(25)</sup>	16.51	n	9	18	14	7	n	y
Brunelli <sup>(18)</sup>	15.50	n	11	n	16	n	y	y
Cataneo (current)	16.90	72	12	24	6	3	y	y

h; height in centimeters; N: number; St: steps; f: flights; F: floors; t: time; VE: verbal encouragement; n: no or no information; y: yes.

25 mL/kg/min, and that those who took more than 40 s have a high probability of presenting a  $\dot{V}O_2$  max below 25 mL/kg/min. As for the variable 'power', we can state that when it is higher or lower than 200 w, the probability of the  $\dot{V}O_2$  max being greater or smaller than 25 mL/kg/min, respectively, is also high, although we will err more frequently than we would if we used the variable 'time'.

Those patients who do not reach the 12 m, with time being considered infinite, should be carefully evaluated in order to detect and try to correct any changes in the oxygen transport system, as previously demonstrated by other authors.<sup>(18)</sup>

The primary objective of the present study was to test the accuracy of the parameters obtained in the fixed-height SCT with verbal encouragement rather than to determine cut-off points in order to identify which would be the patients at high or low risk of surgery. The next step, which is already underway through a prospective study, is to determine the cut-off points for the patients at high, medium and low risk of surgery through the analysis of the post-operative complications correlated with the SCT parameters.

## Acknowledgements

We would like to thank Prof. Dr. Lídia Raquel de Carvalho, from the Department of Biostatistics at the *Universidade Estadual Paulista* (UNESP, Paulista State University), for performing the statistical analysis. We would also like to thank Rafael de Camargo Paccanaro, a resident in General Surgery



at the *Universidade Estadual Paulista* (UNESP, Paulista State University) *Hospital das Clínicas*, for conducting the stair-climbing tests.

## References

1. Shoemaker WC, Appel PL, Kram HB. Role of oxygen debt in the development of organ failure, sepsis and death in high risk surgical patients. *Chest*. 1992;102(1):208-15.
2. Beckman EH. Complications following surgical operations. *Surg Gynecol Obstet*. 1914;18:551-5.
3. Olsen GN. The evolving role of exercise testing prior to lung resection. *Chest*. 1989;95(1):218-25.
4. Schuurmans MM, Diacon AH, Bolliger CT. Functional evaluation before lung resection. *Clin Chest Med*. 2002;23(1):159-72.
5. Eugene J, Brown SE, Light RW, Milne NE, Stemmer EA. Maximum oxygen consumption: a physiologic guide to pulmonary resection. *Surg Forum*. 1982;33:260-2.
6. Weisman IM. Cardiopulmonary exercise testing in the preoperative assessment for lung resection surgery. *Semin Thorac Cardiovasc Surg*. 2001;13(2):116-25.
7. Win T, Jackson A, Sharples L, Groves AM, Wells FC, Ritchie AJ, et al. Cardiopulmonary exercise tests and lung cancer surgical outcome. *Chest*. 2005;127(4):1159-65.
8. Weisman IM, Zeballos RJ. Clinical exercise testing. *Clin Chest Med*. 2001;22(4):679-701.
9. Gaensler EA, Cugell DW, Lindgren I, Verstraeten JM, Smith SS, Strieder JW. The role of pulmonary insufficiency in mortality and invalidism following surgery for pulmonary tuberculosis. *J Thorac Surg*. 1955; 29(2):163-87.
10. Souders CR. Clinical evaluation of the patient for thoracic surgery. *Surg Clin N Am*. 1961;41:545-56.
11. Van Nostrand D, Kjelsberg MO, Humphrey EW. Preresectional evaluation of risk from pneumonectomy. *Surg Gynecol Obstet*. 1968;127(2):306-12.
12. Colman NC, Schraufnagel DE, Rivington RN, Pardy RL. Exercise testing in evaluation of patients for lung resection. *Am Rev Respir Dis*. 1982;125(5):604-6.
13. Bolton JW, Weiman DS, Haynes JL, Hornung CA, Olsen GN, Almond CH. Stair climbing as an indicator of pulmonary function. *Chest*. 1987;92(5):783-8.
14. Olsen GN, Bolton JW, Weiman DS, Hornung CA. Stair climbing as an exercise test to predict the postoperative complications of lung resection: two years experience. *Chest*. 1991;99(3):587-90.
15. Holden DA, Rice TW, Stelmach K, Meeker DP. Exercise testing, 6 min walk, and stair climb in the evaluation of patients at high risk for pulmonary resection. *Chest*. 1992;102(6):1774-9.
16. Pollock M, Roa J, Benditt J, Celli B. Estimation of ventilatory reserve by stair climbing: a study in patients with chronic airflow obstruction. *Chest*. 1993;104(5):1378-83.
17. Pate P, Tenholder MF, Griffin JP, Eastridge CE, Weiman DS. Preoperative assessment of the high-risk patient for lung resection. *Ann Thorac Surg*. 1996;61(5):1494-500.
18. Brunelli A, Al Refai M, Monteverde M, Borri A, Salati M, Fianchini A. Stair climbing test predicts cardiopulmonary complications after lung resection. *Chest*. 2002;121(4):1106-10.
19. European Respiratory Society. Clinical exercise testing with reference to lung diseases: indications, standardization and interpretation strategies. ERS Task Force on Standardization of Clinical Exercise Testing. *Eur Respir J*. 1997;10(11):2662-89.
20. Neder JA, Nery LE. Teste de exercício cardiopulmonar. *J Pneumol*. 2002; 28(Supl 1):S166-S206.
21. Balke B, Ware RW. An experimental study of physical fitness of Air Force personnel. *US Armed Forces Med J*. 1959;10(6):675-88.
22. Martinez EZ, Louzada-Neto F, Pereira BB. A curva ROC para testes diagnósticos. *Cadernos Saúde Coletiva*. 2003;11(1):7-31.
23. Altman DG. *Practical statistics for medical research*. London: Chapman & Hall;1991.
24. Mould RF. Sensitivity and specificity. In: Mould RF, editor. *Introductory medical statistics*. 3rd ed. Bristol: IOP;1998. p.232-8.
25. Girish M, Trayner E Jr, Dammann O, Pinto-Plata V, Celli B. Symptom-limited stair climbing as a predictor of postoperative cardiopulmonary complications after high-risk surgery. *Chest*. 2001;120(4):1147-51.
26. Kinasewitz GT, Welch MH. A simple method to assess postoperative risk. *Chest*. 2001;120(4):1057-8.