

Compliance with peak expiratory flow readings affects the within- and between-reader reproducibility of interpretation of graphs in subjects investigated for occupational asthma

Jean-Luc Malo, MD,^a André Cartier, MD,^a Hebert Ghezze, PhD,^a
and Moira Chan-Yeung, MD^b Montreal, Quebec, and Vancouver,
British Columbia, Canada

Serial recording of peak expiratory flow (PEF) has been proposed as an interesting means in the investigation of occupational asthma (OA).^{1,2} The use of this means of assessment has lately been the object of reappraisal and was presented in a recent statement.²

The advantages include its ability to assess patient's lung function outside the laboratory in a natural setting, the simplicity of measurement with inexpensive and handy devices, the feasibility in nonspecialized centers where specific challenges cannot be performed, and its usefulness in differentiating between nonoccupational and work-related asthma. The limitations of this type of testing include patient's motivation and honesty, organizational difficulties in having patient return to work, and requirement for rather long monitoring. Among these limitations, patient compliance, an aspect that has recently been assessed in two prospective studies,^{3,4} has been shown to be marginally superior to 50% only. It is unknown, however, whether this could have affected the within- and between-reader reproducibility of interpretation of graphs. We therefore decided to pool the data obtained from these two studies to assess the within- and between-reader reproducibility of interpretation.

Abbreviations used

OA: Occupational asthma

PEF: Peak expiratory flow

METHODS

Subjects

The 34 subjects included in this study were previously described. This includes 13 subjects investigated at the Vancouver General Hospital³ and 21 investigated at Sacré-Coeur Hospital.⁴ There were 31 men and 3 women, and their mean age was 39 ($\pm$ 12 SD) years. There were 22 atopic subjects (defined by the presence of at least one immediate reaction to a battery of 15 inhalant allergens as determined by the prick method). The mean duration of exposure at work was 14 ($\pm$ 11 SD) years, and the mean duration of symptoms was 2.3 ($\pm$ 2.4 SD) years. The suspected causal agents were: flour (n = 3), isocyanate (n = 7), Western red cedar (n = 9), formaldehyde (n = 2), methacrylate (n = 1), morphine (n = 1), cereals (n = 1), clam (n = 1), hen (n = 1), inks (n = 2), dyes (n = 1), cloth (n = 1), printing material (n = 1), a metal product (n = 1), soldering (n = 1), and an unidentified chemical product (n = 1). Nineteen subjects were using inhaled steroids, seven used inhaled bronchodilator on an as-needed basis, and eight did not rely on any medication. Seventeen subjects were referred by the Workers' Compensation Board.

Protocol

Each subject was provided with a VMX Mini-Log peak flow meter (Clement Clarke Int., Columbus, Ohio) with detailed verbal and written instructions concerning its use. Each subject was asked to record his or her PEF on six (Vancouver General Hospital) to eight (Sacré-Coeur Hospital) occasions during the day. Three blows were solicited on each occasion. Subjects were asked to record the three readings in a diary with the date and time of measurements, as well as symptoms and medication required. Except for β_2 -agonists, which were taken on demand, the antiasthma medication was kept

From ^athe Department of Chest Medicine, Sacré-Coeur Hospital, Montreal, Canada; and ^bthe Department of Medicine, Vancouver General Hospital.

Supported in part by the Centre Québécois d'Excellence en Santé Respiratoire. Members of the Canadian Network of Centers of Excellence. Dr. Malo is a research scholar with the Fonds de la Recherche en Santé du Québec.

Reprint requests: Jean-Luc Malo, MD, Department of Chest Medicine, Sacré-Coeur Hospital, 5400 West Gouin, Montreal, Quebec, Canada H4J 1C5.

J Allergy Clin Immunol 1996;98:1132-4.

Copyright © 1996 by Mosby-Year Book, Inc.

0091-6749/96 \$5.00 + 0 1/54/77594

TABLE I. Within-reader comparison of interpretation of PEF graphs from self-recorded and stored data

	Self-recorded value	
Stored value	No OA	OA
Reader no. 1		
No OA	26	1
OA	4	3
kappa $\pm$ SEM	0.47 $\pm$ 0.16	
z	2.86	
Reader no. 2		
No OA	26	1
OA	3	4
Kappa $\pm$ SEM	0.60 $\pm$ 0.17	
z	3.55	
Reader no. 3		
No OA	23	3
OA	2	6
Kappa $\pm$ SEM	0.61 $\pm$ 0.17	
z	3.55	

constant during the study period. Subjects were not aware that the PEF values and timing were also being stored in the VMX's microprocessor. The mean duration of monitoring was 23 days for the 13 subjects in Vancouver and 36 days for the 21 subjects investigated at Sacré-Coeur Hospital.

Analysis

Graphs were plotted by using the individual values and the mean, lowest, and highest daily values. This was done by using values written by the subject and the VMX stored values. Graphs were coded in a random fashion and examined in a blinded manner.

Each reader (JLM, AC, and MCY) had to interpret 68 graphs (i.e., 34 graphs with values written by the subject and 34 with stored values) and to code each graph as representing OA or no OA. The criteria for the reader's diagnosis of OA were based on visual analysis of the graphs, which includes lower maximum, mean, or minimum values while at work and/or increased fluctuations occurring on the first day or on consecutive days at work. Kappa statistics were used for comparing within- and between-reader assessment. Briefly, the kappa index assesses the improvement of agreement between readings, which is above and beyond that caused by chance only (for further explanations of the kappa value, see Appendix).

RESULTS

The within-reader concordance of written and stored values is shown in Table I. The concordance of readings was not very satisfactory according to the kappa values. For example, for Reader no. 1, the calculation for kappa statistics according to the Appendix is as follows: Proportion of observed

TABLE II. Between-reader comparison of readings from self-recorded and stored values

	Reader no. 1		Reader no. 3	
	No OA	OA	No OA	OA
Self-recorded values				
Reader no. 2				
No OA	25	2	25	2
OA	1	6	1	6
Kappa $\pm$ SEM	0.74 $\pm$ 0.17		0.74 $\pm$ 0.17	
z	4.35		4.35	
Reader no. 3				
No OA	24	2		
OA	2	6		
Kappa $\pm$ SEM	0.67 $\pm$ 0.17			
z	3.92			
All three: kappa $\pm$ SEM = 0.72 $\pm$ 0.10; z = 7.26				
Stored values				
Reader no. 2				
No OA	28	1	25	4
OA	2	3	0	5
Kappa $\pm$ SEM	0.62 $\pm$ 0.17		0.65 $\pm$ 0.16	
z	4.35		4.03	
Reader no. 3				
No OA	24	1		
OA	6	3		
Kappa $\pm$ SEM	0.36 $\pm$ 0.15			
z	2.34			
All three: kappa $\pm$ SEM = 0.53 $\pm$ 0.10; z = 5.33				

agreement = $(26 + 3) \div 34 = 0.85$; Proportion of expected agreement = $[27 \div 34 \times 30 \div 34] + [4 \div 34 \times 7 \div 34] = 0.72$; Kappa value = $[(0.85 - 0.72)] \div [1.0 - 0.72] = 0.47$. Kappa values were slightly better for Readers nos. 2 and 3. The between-reader concordance of written and stored values is shown in Table II. The kappa values were generally better than for the within-reader comparisons, except for one comparison (Reader no. 1 vs Reader no. 3 for stored values). Compliance tended to be lower in subjects for whom there was disagreement of at least one of the three readers ($66\% \pm 24\%$) than for those in whom there was total agreement ($77\% \pm 26\%$).

DISCUSSION

Studies have shown that compliance with PEF recording is not satisfactory in the investigation of OA.^{3,4} This study shows that this affects the within-subject reproducibility of readings in a significant way. The kappa values for comparison of the self-reported and stored values varied from 0.47 to 0.61 according to the three readers. The within-reader reproducibility of readings was shown to be much higher (kappa values varying from 0.91 to 1.0 for

readings at 2-hour intervals) in a previous study, which used self-reported values read at a 1-week interval by three readers, including two who also participated in this study.⁵

The between-reader comparison of readings for the self-reported values (kappa values varying from 0.67 to 0.74) was comparable to a previously published study (kappa value of 0.77).⁵ However, the kappa values obtained in this study were better with the self-reported values than with the stored values. One possible explanation is that self-reported values were lower than stored values by a mean of 3.4% (by 5% in those referred by the Workers' Compensation Board) while at work and higher by a mean of 1% while away from work; therefore, some tracings might have appeared more suggestive of OA in the case of self-reported values, improving the kappa values. As shown in Table II, more tracings were indeed interpreted as representing OA with self-reported values than with stored values, this being the case for two of the three readers (nos. 1 and 2) and representing twice as many cases (8 vs 4) in the case of Reader no. 1.

In conclusion, compliance with and honesty in PEF monitoring are factors that, among others, affect the interpretation of PEF recording in terms of within- and between-reader reproducibility of readings. This is another pitfall of this approach in the diagnosis of OA, if it must be based on a self-reported assessment only from a consensus made by two or more physicians as previously suggested.⁵ The use of portable spirometers that store data should therefore be advocated. As recently proposed, the ideal apparatus should be able to "record PEF and FEV₁ and store the data with accurate recording of flows."² Serial assessment of airway caliber is an interesting and valuable means for the assessment of asthma and OA.² However, improvement in the methodology of recording is still needed.

We thank Lori Schubert for reviewing the manuscript.

REFERENCES

1. Burge PS. Physiologic assessment of occupational asthma. In: Bernstein IL, Chan-Yeung M, Malo JL, Bernstein DI, editors. *Asthma in the workplace*. New York: Marcel Dekker, 1993:171-88.
2. Moscato G, Godnic-Cvar J, Maestrelli P, Subcommittee on Occupational Allergy of European Academy of Allergy and Clinical Immunology. Statement on self-monitoring of peak expiratory flows in the investigation of occupational asthma. *J Allergy Clin Immunol* 1995;96:295-301.
3. Quirce S, Contreras G, DyBuncio A, Chan-Yeung M. Peak expiratory flow monitoring is not a reliable method for establishing the diagnosis of occupational asthma. *Am J Respir Crit Care Med* 1995;152:1100-2.
4. Malo JL, Trudeau C, Ghezzi H, L'Archevêque J, Cartier A. Do subjects investigated for occupational asthma through serial peak expiratory flow measurements falsify their results? *J Allergy Clin Immunol* 1995;96:601-7.
5. Malo JL, Côté J, Cartier A, Boulet LP, L'Archevêque J, Chan-Yeung M. How many times per day should peak expiratory flow rates be assessed when investigating occupational asthma? *Thorax* 1993;48:1211-7.

APPENDIX

Explanation of the kappa value

Let the data be expressed in the following way.

		Reader B		
		Yes	No	Marginal proportion
Reader A	Yes	a	b	p1
	No	c	d	q1
	Marginal proportion	p2	q2	1.0

where a = the number of subjects classified yes by both observers and b = the number of subjects classified yes by Reader A and no by Reader B, etc.

$p1$ = the proportion of subjects classified yes by Reader A, that is, $p1 = (a + b) \div (a + b + c + d)$ and $q1 = 1 - p1$

$p2$ = the proportion of subjects classified yes by Reader B, that is, $p2 = (a + c) \div (a + b + c + d)$ and $q2 = 1 - p2$

If both readers are completely independent, the expected proportion of agreement (yes by both readers and no by both readers) is called I_e and is given by the product of the corresponding marginals of proportion, that is, $I_e = p1 \times p2 + q1 \times q2$, and the expected number of agreements is: $I_e \times (a + b + c + d)$.

The observed number of agreements is $(a + d)$, and the observed proportion of agreement is $I_o = (a + d) \div (a + b + c + d)$.

The coefficient kappa tries to relate the values of I_o and I_e , that is, the observed agreement and the expected agreement as a result of chance alone. The difference $(1 - I_e)$ is the proportion of cases, which by chance are expected to be classified differently by both observers. It is in these cases that the ability of the readers is to be evaluated. On how many of these cases do the two readers agree? Clearly, the improvement is given by $(I_o - I_e)$, or if the observed proportion is I_o and the expected proportion is I_e , then $I_o - I_e$ gives the excess of agreement or the "real agreement" not caused by chance alone. If one wants to express this difference as a percentage, the denominator should of course be $(1 - I_e)$ or the cases in which chance would indicate a disagreement. Therefore the definition of kappa is: $(I_o - I_e) \div (1 - I_e)$.

So, kappa does not measure agreement per se but assesses improvement of agreement above and beyond that due to chance only.