Measurement Characteristics of Peak Expiratory Flow*

Christina A. Holcroft, ScD; Ellen A. Eisen, ScD; Susan R. Sama, ScD; and David H. Wegman, MD

Study objectives: **To evaluate features of the peak expiratory flow (PEF) test protocol, and to characterize patterns of reproducibility in multiple PEF measurements.**
Design: **Cross-sectional study.**
Setting: **University population.**
Participants: **Two hundred twenty-three healthy adults.**
Interventions: **Participants recorded five PEF measurements in each of five sessions per day for 1 week.**
Measurements and results: **Patterns of within-session variability were characterized using a reproducibility criterion based on a large percentage difference between best trials and evidence of a maneuver-induced bronchospasm (MIB) indicated by successive drops of PEF values in a session. Although the maximum PEF value in a session occurred on the fourth or fifth trial 32% of the time, the change in PEF values was small. Supervision was associated with small improvements in level and reproducibility. Using a cutoff of 5% for defining reproducibility, 15% of all sessions were not reproducible. When averaged over each subject, 9% of the cohort had a mean difference > 5%. Overall, MIB was unusual and observed in 8% of all test sessions; however, MIB was more common among asthmatics and subjects with wheeze, atopy, or allergies than subjects without. By contrast, poor reproducibility was more common among smokers and subjects with cough and phlegm.**
Conclusions: **These results illustrate that it may be unnecessary to supervise all sessions or collect more than three efforts. Results also suggest that reproducibility reflects smoking-related abnormalities, whereas MIB may reflect airways hyperresponsiveness.**
(CHEST 2003; 124:501–510)

Key words: asthma; bronchial spasm; epidemiologic methods; peak expiratory flow; reproducibility of results; respiratory hypersensitivity

Abbreviations: ATS = American Thoracic Society; MIB = maneuver-induced bronchospasm; PEF = peak expiratory flow

Increases in asthma prevalence and mortality during the past decade have led clinicians and epidemiologists to direct more attention to peak expiratory flow (PEF). The increased diurnal variation in PEF, characteristic of asthmatics, has been interpreted as evidence of the increased variability in airway caliber—bronchial hyperresponsiveness—which is the predominant physiologic feature of the disease. Findings of Boezen et al,[1] Higgins et al,[2] Neukirch et al,[3] and Quackenboss et al[4] present a consistent picture in which diurnal amplitude is a marker not only for asthma, but also for the degree of hyperresponsiveness even among nonasthmatics.

The greater circadian variation in PEF suggests that biological variability within test may be at least as large for PEF as for FEV_1. Excessive variability in FEV_1 has been shown to be a good indicator of respiratory difficulty, even in the absence of asthma.[5–8] One of our goals was to find patterns of short-term variability that contain physiologic information. Another aim was to evaluate within-session variability in a relatively unexposed adult population of both asthmatics and nonasthmatics, in order to provide a framework for epidemiologic studies of environmental or occupational hazards.

*From the University of Massachusetts (Drs. Holcroft, Eisen, and Wegman), Lowell; and the Harvard School of Public Health (Dr. Sama), Boston, MA.
This study was supported by National Institutes of Health grant 1-R01-HL51975.
Manuscript received August 8, 2001; revision accepted February 14, 2003.

Correspondence to: Christina Holcroft, ScD, Work Environment Department, UMass Lowell, One University Ave, Lowell, MA, 01854; e-mail: Christina_Holcroft@uml.edu

Until recently, there had been little examination of sources of PEF measurement variability. In fact, in the 1995 update on spirometry standardization, the American Thoracic Society (ATS) states:

> Unlike the FEV_1 obtained from routine spirometry, PEF measurements are more variable and the measurement is often conducted in patients with high variability in their PEF. Although there may be some benefit from using PEF reproducibility to improve subject effort, no specific reproducibility criterion is recommended at this time.[9]

Later that same year, however, Enright and colleagues[10] reported results from a direct examination of within-test reproducibility of PEF in a community-based study. They concluded that the overall reproducibility of PEF was good, with the two best tests within 10 L/min in 73% of sessions, and within 30 L/min in 95% of the sessions. If we assume that the mean PEF is 600 L/min, then the degree of reproducibility may be more similar to FEV_1 than suggested in the ATS statement.[9] Since 40% of maxima occurred on the third trial, the results of Enright and colleagues[10] support the need for three trials.

Some recent attempts have been made to standardize PEF testing on peak flowmeters. Several protocols have been recommended, most notably by US[11] and European respiratory societies,[12] and by Britton[13] and Gannon et al.[14] To improve our understanding of the sources of variability in PEF measurements, we studied an adult population sample in order to contribute information on PEF test performance similar to that used in the development of standards for the forced expiratory maneuver.

Our objectives were to characterize patterns of reproducibility. Because of the repeated test measurements in a session collected for each subject, we were able to examine reproducibility both as a characteristic of a session as well as, more generally, a characteristic of a person. Patterns of short-term variability within a session were characterized using a standard reproducibility criterion based on the difference between the two highest values, as well as a more restrictive criterion proposed by Enright and colleagues[10] to identify a maneuver-induced bronchospasm (MIB). The two measures of short-term variability, *ie*, poor reproducibility and MIB, were then examined in relation to maximum PEF level, and also in relation to other characteristics of individuals, such as asthma, respiratory symptoms, and cigarette smoking.

Materials and Methods

Study Population

In order to examine a generally healthy adult population, participants were recruited from staff and students at the University of Massachusetts Lowell campus in 1996. In order to be eligible for the study, participants had to be > 18 years of age and attend school or work during the day. Staff or students who worked a night shift were ineligible for the study. Because of a limited number of African Americans, researchers also recruited at the Harvard School of Public Health to increase the size of racial subgroups. This study was approved by the UMass Lowell Institutional Review Board. All individuals gave their written informed consent to participate in this study. The data were confidential and used only by the research team.

PEF Diary

Researchers gave each participant a mini-Wright peak flowmeter with a linear scale (Clement Clarke International; Harlow, UK) and a 7-day diary for recording PEF measurements. Researchers instructed participants how to use the flowmeters and supervised participants for several PEF maneuvers. During supervision, if participants were observed using a spitting maneuver or putting their tongue into the mouthpiece, they were instructed not to do so. Participants were told that they could perform the PEF maneuvers either standing or sitting, as long as they were consistent across each session of five PEF measurements. Each day, subjects recorded five PEF measurements (trials) at five times (sessions) during the day. The specific times of day were cued by the events: arising, start of work, lunch, leave work, and bedtime. An attempt was made to supervise at least one session in each person's diary. Individuals were asked to record the actual time of each session. In each test session, we identified the maximal PEF value.

To measure the utility of collecting more than three peak flow maneuvers in a session, we calculated the gain in the maximum PEF value that resulted from observing each of the additional trials. The maximum was determined first from the value of the initial trial only, then from the maximum of the first two trials, then the first three, the first four, and finally from all five trials. Starting with the maximum based on the first two trials, each successive maximum was compared to the one calculated from one less trial. For example, the percentage gain from the (n + 1)th trial was calculated as follows: (maximum based on [n + 1] trials − maximum based on n trials)/(maximum based on n trials).

PEF Variability Parameters

Several different measures of PEF variability were calculated, capturing variation within a test session, across sessions within a day, and across all sessions within a subject.

Within a Test Session: Two measures of within-session variability were examined in detail: a reproducibility requirement based on the difference between the two largest values, and the criterion used by Enright et al[10] to define an MIB. MIB and reproducibility are closely connected concepts. MIB places greater emphasis on the pattern of the drop, and reproducibility places greater emphasis on its magnitude.

Reproducibility was defined as a percentage difference, and calculated as the difference between the two largest values in a session divided by the average of these two values. Nonreproducibility was defined as > 5% difference between the two largest PEF values. Absolute difference, defined in liters per minute as the difference between the two largest values, was also calculated in some instances in comparison to the percentage difference criterion. We used the definition of Enright et al[10] of MIB, a session in which the first PEF value is followed by a drop in the second value (of at least 10 L/min), followed by another drop in the third value (of at least 10 L/min). A session either follows an MIB pattern or does not.

Within a Day: Daily variability was measured by the amplitude/mean percentage, where amplitude was calculated as the difference between the largest and smallest session PEF of each day, and mean was the average across all the sessions in an entire day. For this analysis, each session was characterized by the maximum value of the first three trials. Only days that contained PEF measurements at both arising and bedtime sessions and either lunch or leaving work sessions were considered in the calculation of amplitude.

Within a Person: To characterize short-term variability, an average percentage difference was calculated for each person by averaging the within-session reproducibility across all sessions. Also for each subject, the percentage of sessions with MIB was calculated. High variability for subjects was arbitrarily defined in two ways: (1) those individuals with an average of $> 5\%$ difference across the whole diary, and (2) individuals with at least 15% of all of their test sessions with evidence of MIB. These criteria were chosen to approximately identify the 10% of subjects with the largest variability.

Spirometry

Trained technicians used a Collins bell spirometer (Collins Medical; Braintree, MA) and OMI software (Occupational Marketing; Houston, TX) to administer spirometry following the updated ATS recommendations.[9] A technician calibrated equipment once in the morning and once in the afternoon. A minimum of three acceptable forced expiratory maneuvers was recorded for each participant, but for most individuals, five acceptable maneuvers were recorded. PEF, FEV_1, FVC, and FEV_1/FVC measures were used in the analysis. Percentage of predicted values were calculated based on the prediction equations of Hankinson et al[15] for whites, African Americans, and Hispanics.[15] The white prediction equations were used for the Asians and other races in our study.

Symptom Questionnaire

The ATS questionnaire[16] was administered to collect information on demographics, respiratory history, and current symptoms. Race was grouped into Asian, black, Hispanic, white, and other categories. Asthma was defined as a self-reported physician diagnosis of current asthma. Respiratory symptoms were defined individually according to a "yes" answer to the following questions: Do you usually have a cough? Do you usually bring up phlegm at all on getting up, or first thing in the morning? Does your chest ever sound wheezy or whistling occasionally apart from colds? During the past 12 months, has your chest ever felt tight for longer than a minute? During the past 12 months have you had an attack of shortness of breath or coughing that came on shortly after you stopped exercising? *No symptoms* was defined as a "no" answer to all five questions above. Since these symptoms did not vary across the diary, they were treated as fixed characteristics.

Allergies and Atopy

Researchers administered an allergy questionnaire and any person who reported allergies was labeled as having allergies. A licensed nurse administered a skin-prick test for atopy to those individuals without any known allergies. The nurse applied allergens using a Multi-Test kit (Lincoln Diagnostics; Decatur, IL) on the volar aspect of a subject's forearm. The following allergens were obtained from Bayer Pharmaceutical (Spokane, WA): dust mite (*Dermatophagoides farinae* and *Dermatophagoides pteronyssinus*), ragweed mix, red oak, grass mix, cockroach mix, mold mix, a negative control, and a positive histamine control. Fifteen minutes after applying the allergens, the nurse used a transparency to trace the outlines of the wheal and erythema of each reaction. Reactions were scored based on the size of the wheal, following guidelines recommended by Lincoln Diagnostics. A score of 1+ meant that the wheal may or may not be present, and if present the wheal must be as large as the negative control; 2+ meant that the wheal was 5 to 7 mm. Atopy was defined as any reaction 2+ or greater. Using a criterion of any reaction 1+ or greater would not change who was defined as reacting in this study, since no one had exclusively 1+ reactions.

Statistical Methods

The PEF parameters of interest were functions of session, day, or the individual subject. Most of the PEF measurement attributes of interest are characteristics of the session: the trial number of the maximum value, the presence of supervision, within-session reproducibility, and evidence of MIB. Diurnal variation, as measured by daily amplitude, is a feature of the day. Several PEF measurement attributes were also characterized for a subject, such as average PEF level, average daily amplitude, average degree of within-session reproducibility, and percentage of MIB sessions over the entire diary. Differences in mean values between subgroups were tested using either a *t* test or analysis of variance for unbalanced data. For *t* tests, we assumed either equal or unequal variances, as appropriate. After using analysis of variance to test for differences in means across several groups, the Tukey multiple comparison test was used to compare pairs of the groups. For prevalence comparisons, a χ^2 test of proportions was used unless the number in a cell was five or less, in which case the Fisher exact test was applied.

Results

The university cohort consisted of 223 staff and students. The mean age of the population was 39 years (range, 19 to 70 years; median, 38 years). Seventy-five percent of the participants were white, 12% were African American, 6% were Asian, 5% were Hispanic, and 2 participants were in none of these categories. The Asians, Hispanics, and the unspecified group were categorized as "other" in the results by race. Among the 44% of the population who had ever smoked cigarettes, the median age of starting regularly smoking was 17 years old, and the median total pack-years was 11. For the 27% who had quit smoking, the median number of years since quitting was 12 years. The median number of cigarettes smoked per day among the current smokers was 20 cigarettes per day. The 21 self-reported physician-diagnosed asthmatics used inhalers on 16% of their diary days.

On average, participants completed 32 sessions in their diaries. The average number of complete days in a diary—that is, a day with an arising, bedtime, and either lunch, or leaving work sessions—was 5.8. Out of the 223 participants, 220 completed the ATS questionnaire, 219 underwent spirometry, and 209 completed the atopy and allergy assessment. Of the 110 people with allergies or atopy, 56% were based on self report.

The average PEF level, percentage of predicted PEF, and PEF variability as measured across sessions by daily amplitude, are presented for subgroups of the cohort defined by demographic and respiratory characteristics (Table 1) [each participant contributed one grand average PEF to the group means]; p values < 0.05 are reported in the table for each outcome and demographic category. Female subjects had lower mean PEF but higher percentage of predicted PEF. Female subjects also had significantly more variability, as measured by higher amplitude, 10% vs 8%. Twenty-five percent of the cohort was nonwhite, and 12% were African American. There were no significant differences in PEF across racial groups. Older subjects had slightly higher amplitude, 10% vs 8%, but not significantly different percentage of predicted values. Current smokers had substantially lower percentage of predicted PEF and higher amplitude. The 8% of the cohort who reported current physician-diagnosed asthma and had either allergies or atopy had lower percentage of predicted PEF and slightly higher amplitude. Those in the asthma/no allergy category and allergy/no asthma category had higher percentage of predicted PEF than those with neither asthma nor allergies, although the differences were not statistically significant. When allergy categories were grouped into self-reported allergies, allergic by skin test, and nonallergic by self report and skin test, there were no significant differences between groups for mean PEF, percentage of predicted PEF, or amplitude.

We compared PEF levels between the first day of the diary and other days to examine for a learning effect. No differences were found between the first day and other days, among weekdays after stratifying by time of day. In fact, average levels of PEF at all sessions besides arising were slightly higher on the first day than on other days, although the differences were not statistically significant. Twenty-two percent of participants in this learning-effect analysis started their diaries on a Monday, and the rest were spread out throughout the week. Supervision was fairly evenly distributed throughout the diary days, with only 9% of supervised sessions occurring on the first day of the diary. To evaluate whether higher values early in the diary were due to participants losing interest over time, weekday PEF levels from the last 2 diary days were compared to the rest, but no differences were found. Without any evidence of a learning effect, we did not exclude any diary days from the subsequent analysis.

Table 1—*Demographics and PEF Characteristics of a University Sample of 223 Staff and Students*

Characteristics	No. (%)	Average PEF Level, L/min* (SD)	Percentage Predicted PEF, %	Average Daily Amplitude, % (SD)
Overall	223 (100)	495 (101)	105	9.0 (5.5)
Gender				
Male	97 (43)	573 (81)†	103	7.8 (4.0)‡
Female	126 (57)	434 (66)	107	10.0 (6.2)
Race				
White	168 (75)	494 (104)	106	9.1 (5.5)
African American	27 (12)	497 (107)	106	8.7 (6.1)
Other	28 (13)	495 (74)	101	8.8 (4.7)
Age, yr				
≤ 40	121 (54)	507 (90)	106	8.1 (4.7)§
> 40	102 (46)	480 (111)	104	10.1 (6.0)
Smoking status				
Current smoker	37 (17)	445 (108)‖	99¶	12.3 (7.1)‖
Ex-smoker	61 (27)	498 (102)	106	8.5 (5.1)
Never smoker	125 (56)	508 (93)	107	8.3 (4.7)
Asthma/allergy or atopy				
Both asthma/allergy	17 (8)	445 (93)	93#	10.1 (6.1)
Asthma/no allergy	4 (2)	484 (104)	116	9.0 (3.9)
Allergy/no asthma	93 (42)	512 (106)	108	8.7 (4.1)
None	95 (43)	483 (95)	104	9.5 (6.7)

*Mean of the distribution of each person's diary average level using all five trials of each session.
†Statistically different from female subjects ($p < 0.05$).
‡Statistically different from female subjects ($p < 0.05$).
§Statistically different from age > 40 ($p < 0.05$).
‖Statistically different from ex-smokers and never smokers ($p < 0.05$).
¶Statistically different from never smokers ($p < 0.05$).
#Statistically different from allergy/no asthma ($p < 0.05$).

Number of Trials

A third of the maximums occurred on the first effort and almost 70% occurred on the first, second, or third trial (Table 2). Although the maximum in 32% of the sessions did not occur until the fourth or fifth trial, the percentage gain in the maximum derived by including the fourth and fifth trials was trivial: $\leq 2\%$ in 90% of all sessions. For nonzero gains, the average percentage gain based on including the fourth and fifth trials was 3.6%.

Mean reproducibility, or within-session variability, from all sessions is also reported in Table 2 by the number of trials used in the calculation. As the number of trials used increased, the calculated within-session variability decreased. This can be explained by noting that variability is determined only by the top two values; therefore, variability would decrease with more trials since it is more likely that the maximum is reproduced.

In practice, patients could be asked to record three maneuvers, determine if the two largest values match within 30 L/min or 40 L/min, quit if they do, but do two or three additional maneuvers if they do not. We determined how maximum PEF and reproducibility changes from three to five trials when following this protocol. Among those sessions that had an absolute difference > 40 L/min at three trials, the 90th percentile of the increase in max PEF from three to five trials was 3.4%, and the 75th percentile increase was 0. Among sessions with an absolute difference > 30 L/min at three trials, the 90th percentile of increase in maximum PEF from three to five trials was 2.9%, and the 75th percentile increase was 0. Reproducibility improved with more trials, even among those sessions with absolute difference > 30 L/min or > 40 L/min. Based on the findings that the fourth and fifth trials do not provide significant additional information, in most of the subsequent analyses the fourth and fifth trials were not included.

Table 2—*Characteristics of Five PEF Trials*

Trial No.	When Maximum of Five Trials Occurs, %	% Increase* in Maximum With Additional Trial		Mean Reproducibility, 90th Percentile
		75th Percentile	90th Percentile	
1	32			
2	20	2.4	6.1	9.2
3	17	1.1	3.3	6.2
4	17	0	2.3	5.2
5	15	0	1.9	4.8

*(maximum from [n + 1] trials − maximum from n trials)/maximum from n trials.

Supervision

PEF level and reproducibility were compared between supervised and unsupervised sessions. The comparisons were paired within person and stratified by time of day. A total of 79 individuals had both supervised and unsupervised sessions at start of work, 110 individuals had both types of sessions at the lunch measurement, and 46 individuals in the leave work session. Each individual contributed an average maximum PEF of three trials from all supervised sessions, as well as an average maximum from all unsupervised sessions to this analysis. Using paired *t* tests, PEF levels were slightly higher, approximately 10 L/min higher ($p < 0.01$), in the supervised sessions at all three times of day (Table 3). Supervision was also associated with more reproducible PEF performance, although not significantly so, at all three times of day; the mean percentage difference between the two best tests was smaller in the supervised sessions.

The effect of supervision is not as clear when examining the frequency of nonreproducible or MIB sessions (Table 3). This is probably due to the small number of supervised sessions per person, so percentages are not very stable. In general, there are fewer nonreproducible sessions during supervision, but there were no patterns observed for supervision and frequency of MIB. None of these differences were statistically significant.

Table 3—*Impact of Supervision on Mean PEF Level and Within-Session Reproducibility*

Variables	Session		
	Start Work	Lunch	Leave Work
Participants, No.*	79	110	46
PEF level, L/min			
Supervised	511	522	531
Unsupervised	501	514	518
p value	< 0.01	< 0.01	< 0.01
Reproducibility, %			
Supervised	2.0	2.3	2.0
Unsupervised	2.5	2.5	2.5
p value	0.1	0.6	0.3
High within-session variability (% of sessions with difference > 5%), %			
Supervised	8.9	12.7	8.7
Unsupervised	12.6	11.5	14.2
p value	0.2	0.7	0.3
High frequency of MIB (% of sessions with MIB), %			
Supervised	8.9	9.1	1.1
Unsupervised	6.2	8.8	4.7
p value	0.4	0.9	0.1

*No. of people with both supervised and unsupervised sessions within a particular time period.

Variability Within a PEF Session

Poor reproducibility was defined as a session with a percentage difference $> 5\%$, by analogy with previous ATS guidelines for FEV_1 reproducibility.[17] Fifteen percent of all sessions had a percentage difference $> 5\%$, and 4% of sessions had a percentage difference $> 10\%$. When examining absolute difference cutoffs, 13% of sessions had an absolute difference > 20 L/min, 7% of sessions were > 30 L/min, and 4% of sessions were > 40 L/min.

The relationship between absolute difference and percentage difference was investigated in terms of sensitivity and specificity calculations of 30 L/min and 40 L/min absolute difference cutoffs using a 5% difference cutoff as a "gold standard." Among the nonreproducible sessions with percentage difference $> 5\%$, 46% of sessions had an absolute difference > 30 L/min and 26% had an absolute difference > 40 L/min, which are the measures of sensitivity for each absolute difference cutoff. All reproducible sessions with a percentage difference $\leq 5\%$ had an absolute difference < 30 L/min, which means that the specificity for both 30 L/min and 40 L/min cutoffs is 100%. Between these two cutoffs for absolute difference, the 30 L/min cutoff has the highest sensitivity and specificity for a percentage difference cutoff of 5%. Correlations between absolute difference, percentage difference, and height and maximum PEF were calculated to determine how the reproducibility measures related to the size of the person. We found that absolute difference was not more correlated with height or maximum PEF than percentage difference was, so we found no evidence that absolute difference is a more biased measure.

The relationship between reproducibility and MIB was examined at the session level in Table 4. Within the percentage difference categories in Table 4, the 90th percentile of the distribution of absolute differences is included to illustrate the overlap between the two measures. Twenty-six percent of the sessions with MIB were nonreproducible (percentage difference $> 5\%$). By contrast, MIB was observed in only 8% of all test sessions and in 13% of sessions with poor reproducibility. The odds ratio for MIB and poor reproducibility is 2.2 (95% confidence interval, 1.8 to 2.7), suggesting a moderate association between the two characteristics. Nonreproducibility also calculated as a percentage difference $> 10\%$ is included in Table 4. The odds ratio between reproducibility with a 10% difference cutoff and MIB is 2.4 (95% confidence interval, 2.2 to 2.5), which shows that the relationship between reproducibility and MIB is not greatly dependent on the cutoff chosen for nonreproducibility.

MIB was also examined in relation to inhaler use, which was recorded daily and not at each session. Only 3% of all sessions were measured on a day when an inhaler was used. There was no difference between the proportion of MIB sessions on days when an inhaler was used and on days when an inhaler was not used. The odds ratio between inhaler use and MIB was 1.0.

Session Reproducibility and PEF Level: To examine whether poorly reproducible tests had lower peak flow values, we compared PEF levels between reproducible and nonreproducible sessions. Calculations were paired so that each individual contributed one summary PEF value from the maximum of three blows, separately for reproducible and nonreproducible sessions at each of the five daily measurement times. This pairing within subject eliminated the potential effects of confounding by smoking or other demographic characteristics, and stratification by time of day removed diurnal effects. When comparisons were calculated using a subject's average PEF level, PEF values in nonreproducible sessions were significantly higher than in reproducible sessions (data not shown). Comparisons were recalculated using the maximum PEF value of a subject's sessions (Table 5), and PEF levels in nonreproducible sessions were equal to reproducible sessions or still slightly higher at the arising and bedtime time points. As discussed below, overall these findings

Table 4—*Reproducibility and MIB Characteristics by Session*

Variables	90th Percentile of Absolute Difference, L/min	No Session MIB	Session MIB	Column Total (%)
% difference $\leq 5\%$	20	5674	395	6,069 (85)
% difference $> 5\%$	60	894	139	1,033 (15)
Row total (%)*		6,568 (92)	534 (8)	7,102 (100)
% difference $\leq 10\%$	20	6,344	493	6,837 (96)
% difference $> 10\%$	120	224	41	265 (4)
Row total (%)*		6,568 (92)	534 (8)	7,102 (100)

*p value < 0.01 from a χ^2 test of association.

Table 5—*Maximum PEF Level Compared Between Reproducible and Nonreproducible Sessions (With Percentage Difference > 5%) and Compared Between Non-MIB and MIB Sessions*

Maximum of a Subject's Sessions	Reproducibility				MIB			
	No.*	Reproducible	Nonreproducible	p Value From Paired *t* Tests	No.*	Non-MIB	MIB	p Value From Paired *t* Tests
Arising	117	486	496	0.02	72	524	510	< 0.01
Start work	103	500	502	0.4	90	524	515	< 0.01
Lunch	100	489	487	0.5	86	537	516	< 0.01
Leave work	102	489	490	0.8	78	522	503	< 0.01
Bedtime	95	504	512	0.09	81	525	512	0.03

*No. of people with both types of sessions within a particular time period.

suggest that it is the higher PEF values that cannot be reproduced.

Contrary to expectation, average PEF levels were higher in MIB sessions than non-MIB sessions (data not shown). However, this pattern was not observed using a maximum PEF level, and maximum PEF values were significantly lower in a MIB session than the maximum values from a non-MIB session (Table 5).

Variability by Person

To this point, airways variability has been described as an attribute of a particular PEF test session consisting of a set of maneuvers performed over several minutes. We were also interested in characterizing general airway variability for an individual with a summary measure of short-term airway lability. To this end, an average percentage difference and percentage MIB were calculated over all sessions for each subject. We then compared the prevalence of respiratory symptoms, spirometry outcomes, and other demographic characteristics between subjects with more and less short-term airways variability.

Nine percent of the 223 participants had an average percentage difference $> 5\%$, and only two individuals had an average percentage difference $> 10\%$. Sixty-eight subjects (30%) had evidence of MIB in at least 10% of their sessions, 29 people (13%) had a total of $\geq 15\%$ of MIB sessions, and 10 people (4%) had a total of $\geq 20\%$ of MIB sessions.

The proportions of individuals with high within-session variability and a high frequency of MIB were examined within symptomatic subgroups of the population (Table 6). The cutoffs for poor reproducibility (average within-session percentage difference $> 5\%$) and high frequency of MIB (percentage sessions with MIB $> 15\%$) were chosen because they separated approximately the top 10% of individuals in each group. Results indicate that these two variability parameters are measuring quite different physiologic attributes. MIB appeared to be more common among asthmatics with allergy or atopy and subjects with wheeze. By contrast poor reproducibility was more common among smokers and subjects with cough and phlegm. Both characteristics are associated with depressed percentage of predicted FEV_1, FVC, and PEF, and slightly differentiated by depressed FEV_1/FVC (results not shown because

Table 6—*Proportion of Participants Characterized as Having Poor Reproducibility (Average Within-Session % Difference > 5%) and Having a High Frequency of MIB (% of Sessions With MIB > 15%)*

Variables	No.	High Within-Session Variability, %	High Frequency of MIB, %
Overall	223	9	13
Symptoms			
Usual cough	23	22*	13
Morning phlegm	25	20	12
Wheeze	40	10	20
Chest tightness	43	7	9
Shortness of breath	44	9	14
No symptoms†	126	9	14
Smoking			
Current smoker	37	22*	8
Ex-smoker	61	8	8
Never smoker	125	6	17
Asthma/allergy or atopy			
Both asthma/allergy	17	0	29*
Asthma/no allergy	4	0	25
Allergy/no asthma	93	9	13
None	95	13	9
Spirometry			
FEV_1			
< 85% % predicted	34	21*	32*
≥ 85% % predicted	185	7	10
FVC			
< 85% % predicted	24	29*	42*
≥ 85% % predicted	195	7	10
PEF			
< 85% % predicted	25	24*	24*
≥ 85% % predicted	193	7	12

*p value < 0.10 from a two-sample test of proportions, using the "no symptoms" category as the reference group. Fisher exact test was used if a cell count was ≤ 5.

†None of the five selected symptoms.

only eight individuals had a percentage of predicted FEV_1/FVC value $< 85\%$). There is not much overlap between those with percentage of predicted PEF $< 85\%$ and those with both asthma and allergy. Out of the 25 people with percentage of predicted PEF $< 85\%$, 16% have both asthma/allergy, 0% have asthma/no allergy, 32% have allergy/ no asthma, and 48% have neither asthma nor allergy.

Examining the overlap between asthma, symptoms, and poor reproducibility, none of the 21 asthmatics had high within-session variability. Six asthmatics had a high frequency of MIB: one patient had no respiratory symptoms, and the remaining had a combination of wheeze, chest tightness, and shortness of breath. When individuals with a high frequency of MIB are stratified by asthma status, the numbers are very small but show a higher prevalence of MIB among asthmatics with wheeze, chest tightness, and shortness of breath compared to nonasthmatics with these symptoms.

Discussion

The results of this study support the current practice of recording three PEF trials at each session. Although the maximum of a session occurred on the fourth or fifth trials one third of the time, the incremental gain in the maximum PEF value was not large enough to warrant collecting the additional trials; however, single trials only slightly underestimate the maximum value of three trials (on average the first PEF value was 97% of the maximum based on the first three trials). With a single trial, however, we have no assessment of within-session reproducibility that provides a measure of validity as well as precision and is therefore useful in interpreting results. The findings of this study illustrate the value of recording three trials in a session in order to identify distinct patterns of variability that are associated with individual characteristics.

Increased amplitude was observed for current smokers and asthmatics, consistent with previous findings.[4,12,18–23] No differences were found between racial groups.

Our finding that PEF level or reproducibility was not associated with atopy or allergies is consistent with previous literature. In his review, Burrows[24] concluded that there was no convincing evidence that atopy, in the absence of asthma, was an important determinant of airway responsiveness. In arriving at this conclusion he noted that "[the] only nonasthmatic subjects who showed significantly increased responsiveness for their level of pretest lung function were atopics and/or nonsmokers who complained of respiratory symptoms."[24] He did, however, indicate that atopic status was useful in identifying individuals who might have a preclinical asthmatic-like disorder, without regard to whether atopy was a precursor of the disorder or just a feature of the preclinical state.[24]

Unlike spirometry, which is always performed under supervision, measurement of PEF is rarely supervised. Our findings showed that supervision increased the maximum PEF level by approximately 2%, and slightly improved within-session reproducibility. This increase in PEF level due to supervision is unlikely to be important clinically; however, in an epidemiologic study the difference may be considered significant. Care should be taken, therefore, when planning and managing an epidemiologic study of PEF.

Using a reproducibility cutoff of 5% difference at a single session, 15% of all sessions were nonreproducible. When the percentage difference was averaged across an entire diary for each person, 9% of the cohort had an average percentage difference $> 5\%$. Using these criteria, nonreproducibility was shown to be associated with individuals with symptoms in the large airways: current smokers and those with usual cough and morning phlegm.

MIB as defined by Enright and colleagues[10] was also useful in characterizing the diary data. The MIB pattern was successful in identifying a PEF characteristic that appeared distinct from the reproducibility criterion (percentage difference). As the results in Table 4 show, only 26% of the sessions with MIB were also nonreproducible. The MIB pattern appears to identify asthmatics and individuals with wheeze who are not identified by the nonreproducibility criterion. This finding among individuals with a high prevalence of sessions with MIB was also distinct from findings in individuals with a high average within-session variability.

We found results for reproducibility similar to those of Enright and colleagues[10]: the best two PEF values were within 10 L/min in 69% of all sessions, and within 30 L/min in 93% of all sessions (compared to 73% and 95% reported by Enright et al[10]). They found that an MIB occurred in $< 5\%$ of sessions, while we found an MIB in 8% of sessions. In the data of Enright et al,[10] the maximum value occurred on the third trial 40% of the time, while in our data, using the first three trials alone, we found that the maximum occurred on the first trial 44% of the time and 26% on the third trial. More frequent maximums on the first trial are consistent with more frequent MIBs, and perhaps our population was more symptomatic than the population of Enright et al.[10]

Considering the 40 L/min cutoff recommended by the European Respiratory Society, our data showed

96% of sessions were within 40 L/min, consistent with previous findings by Quanjer et al.[12] Any particular individual can have a nonreproducible session, and 96% of individuals had < 20% of sessions with variability > 40 L/min. Fifty-seven percent of individuals had no sessions in their diary with variability > 40 L/min.

The finding that average PEF level is higher in nonreproducible sessions compared to reproducible sessions is an interesting result, especially since the difference is consistent across all time sessions. Initially we suspected that outlying high PEF values, or "super-blows,"[25,26] were driving the difference. Only five of the values were the absolute maximum value on the mini-Wright flowmeter, 800 L/min, and excluding these values did not change the results. Strayhorn et al[25] found an average increase of 12% in mini-Wright flowmeter measurements when individuals used an incorrect technique. The increases we found were not as high, and many of the values were not outliers in a person's distribution of PEF values, so we could not identify which values may have been artificially high.

It is possible that higher nonreproducible levels were due to a bronchodilation effect due to deep inhalation that has been documented.[27–29] This bronchodilation effect occurs in healthy subjects as well as asymptomatic asthmatics, and is not commonly observed in severe asthmatics. Since the PEF maneuvers are separated by seconds, the brief bronchodilation that would explain our results is consistent with the changes in specific airway conductance measured by Parham and colleagues,[29] in which the greatest increase in airway conductance was measured at the earliest possible measure, 7 s after deep inhalation, and dropped nonlinearly within 60 s.

There is a documented compliance problem with PEF self reporting and invented values.[30–32] This problem can be eliminated by the use of computerized instruments like the VMX mini-log peak flowmeter (Clement Clarke; London, UK) that automatically stores data, and ideally all peak flow monitoring would be done with these instruments. In previous validation studies of PEF monitoring, the best compliance was observed in the first week of data collection, among individuals who were not being evaluated for occupational asthma. Verschelden et al[31] found that when there were stored values, written values were in agreement 90% of the time. In our data, we would not expect a bias in reported values with respect to level or reproducibility. Assuming that not all reported values are accurate, the effect is the addition of random noise to the data that weakens associations between variables. Given that we were able to find some associations, this suggests that we had sufficient information in our data.

As a recommendation for future work involving PEF measurements, at least three PEF trials should be recorded at each time session. The results from this study suggest that no data should be excluded on the basis of a reproducibility criterion; rather, reproducibility and MIB should be characterized for study subjects. Our finding that nonreproducibility is associated with cough and phlegm, while MIB is associated with wheeze should be examined and confirmed in future studies. It is also possible that nonreproducibility captures a bronchodilation effect on deep inhalation that occurs in healthy subjects, which could be further investigated.

This PEF monitoring was done in a university setting, which is one example of an adult population in an environment without many respiratory hazards. These findings would most likely be applicable to working populations in office environments, and they would be a good reference for working populations in industrial environments containing respiratory hazards.

ACKNOWLEDGMENT: The authors thank the participants for their cooperation and the reviewers for their contributions.

References

1 Boezen HM, Postma DS, Schouten JP, et al. PEF variability, bronchial responsiveness and their relation to allergy markers in a random population (20–70 yr). Am J Respir Crit Care Med 1996; 154:30–35
2 Higgins BG, Britton JR, Chinn S, et al. Comparison of bronchial reactivity and peak expiratory flow variability measurements for epidemiologic studies. Am Rev Respir Dis 1992; 145:588–593
3 Neukirch F, Liard R, Segala C, et al. Peak expiratory flow variability and bronchial responsiveness to methacholine. Am Rev Respir Dis 1992; 146:71–75
4 Quackenboss JJ, Lebowitz MD, Krzyzanowski M. The normal range of diurnal changes in peak expiratory flow rates: relationship to symptoms and respiratory disease. Am Rev Respir Dis 1991; 143:323–330
5 Becklake MR. Epidemiology of spirometric test failure. Br J Ind Med 1990; 47:73–74
6 Eisen EA, Oliver LC, Christiani DC, et al. Effects of spirometry standards in two occupational cohorts. Am Rev Respir Dis 1985; 132:120–124
7 Kellie SE, Attfield MD, Hankinson JL, et al. Spirometry variability criteria: association with respiratory morbidity and mortality in a cohort of coal miners. Am J Epidemiol 1987; 125:437–444
8 Ng'Ang'a LW, Ernst P, Jaakkola MS, et al. Spirometric lung function: distribution and determinants of test failure in a young adult population. Am Rev Respir Dis 1992; 145:48–52
9 American Thoracic Society: Standardization of spirometry: 1994 update. Am J Respir Crit Care Med 1995; 152:1107–1136
10 Enright PL, Sherrill DL, Lebowitz MD. Ambulatory monitoring of peak expiratory flow: reproducibility and quality control. Chest 1995; 107:657–661
11 National Asthma Education and Prevention Program. Expert panel report 2: guidelines for the diagnosis and management

of asthma. Bethesda, MD: National Institutes of Health, 1997; publication No. 98–4051
12 Quanjer PH, Lebowitz MD, Gregg I, et al. Peak expiratory flow: conclusions and recommendations of a Working Party of the European Respiratory Society. Eur Respir J 1997; 10(suppl 24):2s–8s
13 Britton J. Measurement of peak flow variability in community populations: methodology. Eur Respir J 1997; 10(suppl 24): 42s–44s
14 Gannon PFG, Newton DT, Pantin CFA, et al. Effect of the number of peak expiratory flow readings per day on the estimation of diurnal variation. Thorax 1998; 53:790–792
15 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999; 159:179–187
16 Epidemiology Standardization Project, Executive Committee, American Thoracic Society. Recommended respiratory disease questionnaires for use with adults and children in epidemiological research. Am Rev Respir Dis 1978; 118:7–52
17 American Thoracic Society: Standardization of spirometry: 1987 update. Am Rev Respir Dis 1987; 136:1285–1298
18 Thiadens HA, De Bock GH, Dekker FW, et al. Value of measuring diurnal peak flow variability in the recognition of asthma: a study in general practice. Eur Respir J 1998; 12:842–847
19 Enright PL, Burchette RJ, Peters JA, et al. Peak flow lability: association with asthma and spirometry in an older cohort. Chest 1997; 112:895–901
20 Lebowitz MD, Krzyzanowski M, Quackenboss JJ, et al. Diurnal variation of PEF and its use in epidemiological studies. Eur Respir J 1997; 10(suppl 24):49s–56s
21 Timonen KL, Nielsen J, Schwartz J, et al. Chronic respiratory symptoms, skin test results, and lung function as predictors of peak flow variability. Am J Respir Crit Care Med 1997; 156:776–782
22 Siersted HC, Hansen HS, Hansen NCG, et al. Evaluation of peak expiratory flow variability in an adolescent population sample: the Odense Schoolchild Study. Am J Respir Crit Care Med 1994; 149:598–603
23 Higgins BG, Britton JR, Chinn S, et al. The distribution of peak expiratory flow variability in a population sample. Am Rev Respir Dis 1989; 140:1368–1372
24 Burrows B. Allergy and its relationship to asthma, airway responsiveness, and chronic airway obstruction: an epidemiologic perspective. In: Weiss ST, Sparrow D, eds. Airway responsiveness and atopy in the development of chronic lung disease. New York, NY: Raven Press, 1989; 258–259
25 Strayhorn V, Leeper K, Tolley E, et al. Elevation of peak expiratory flow by a "spitting" maneuver: measured with five peak flow meters. Chest 1998; 113:1134–1136
26 Hankinson JL, Das MK. Frequency response of portable PEF meters. Am J Respir Crit Care Med 1995; 152:702–706
27 Pellegrino R, Sterk PJ, Sont JK, et al. Assessing the effect of deep inhalation on airway calibre: a novel approach to lung function in bronchial asthma and COPD. Eur Respir J 1998; 12:1219–1227
28 Malmberg P, Larsson K, Sundblad BM, et al. Importance of the time interval between FEV_1 measurements in a methacholine provocation test. Eur Respir J 1993; 6:680–686
29 Parham WM, Shepard RH, Norman PS, et al. Analysis of time course and magnitude of lung inflation effects on airway tone: relation to airway reactivity. Am Rev Respir Dis 1983; 128:240–245
30 Malo J-L, Trudeau C, Ghezzo H, et al. Do subjects investigated for occupational asthma through serial peak expiratory flow measurements falsify their results? J Allergy Clin Immunol 1995; 96:601–607
31 Verschelden P, Cartier A, L'Archeveque J, et al. Compliance with and accuracy of daily self-assessment of peak expiratory flows (PEF) in asthmatic subjects over a three month period. Eur Respir J 1996; 9:880–885
32 Quirce S, Contreras G, Dybuncio A, et al. Peak expiratory flow monitoring is not a reliable method for establishing the diagnosis of occupational asthma. Am J Respir Crit Care Med 1995; 152:1100–1102