ORIGINAL ARTICLE

Limits of longitudinal decline for the interpretation of annual changes in FEV_1 in individuals

Eva Hnizdo, Kanta Sircar, Tieliang Yan, Philip Harber, James Fleming, Henry W Glindmeyer

....

See end of article for authors' affiliations
....

Correspondence to:
Dr E Hnizdo, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, 1095 Willowdale Road, Morgantown, WV 26505, USA; ehnizdo@cdc.gov

Objective: Spirometry-based screening programmes often conduct annual assessment of longitudinal changes in forced expiratory volume in 1 second (FEV_1) to identify individuals with excessive rates of decline. Both the American Thoracic Society (ATS) and the American College of Occupational and Environmental Medicine (ACOEM) recommend a reference limit value of $\geqslant 15\%$ for excessive annual decline. Neither the ATS nor the ACOEM adjust this limit for the precision of the existing spirometry data. The authors propose an improved method of defining the reference limit of longitudinal annual FEV_1 decline (LLD) based on the precision of the spirometry data.
Method: The authors used data from four monitoring programmes and measured their data precision using a pair-wise within-person variation statistic. They then derived programme- and gender-specific absolute and relative LLD values and validated these against the 95th percentiles for observed yearly changes in FEV_1.
Results: The relative limit for annual decline was more practical than the absolute limit as it adjusted for gender differences in the magnitude of FEV_1. The programme-specific relative limit values were in good agreement with 95th percentiles for year-to-year FEV_1 changes and ranged from 6.6% to 15.8%. For individuals with COPD and bronchial hyperreactivity the 95th percentiles for year-to-year changes were about 15% and higher.
Conclusions: The relative longitudinal limit for annual FEV_1 decline based upon precision of measurements is valid and can be generalised to different gender and population groups. A relative limit of approximately 10% appears appropriate for good quality workplace monitoring programmes, whereas a limit of about 15% appears appropriate for clinical evaluation of individuals with an obstructive airway disease. Computer software based on the method described is available from the corresponding author.

The interpretation of an individual's longitudinal spirometry data in workplace spirometry monitoring programmes should include assessment of both the level of lung function and rate of lung function decline. The level of lung function is usually interpreted against an expected value for an asymptomatic non-smoker of the same age, height, race/ethnicity, and gender. An abnormal lung function level is usually identified using the lower limit of normal (LLN), which approximates the one-sided 95% confidence limit for the expected value and identifies approximately 5% of healthy never-smokers as abnormal.

Interpretation of the rate of lung function decline is less standardised. There are sophisticated statistical methods designed for the analysis of historically collected longitudinal data comparing mean rates of decline among groups in research studies. However, there has been less effort in developing and validating two important practical applications: (1) evaluating an individual's rate of decline prospectively for early identification of those with excessive lung function decline; and (2) monitoring the quality (that is, data precision) of workplace spirometry monitoring programmes. Currently, there are two recommended methods for prospective evaluation of the rate of decline in an individual. The American Thoracic Society (ATS) recommends a reference limit of annual longitudinal decline for forced expiratory volume in 1 second (FEV_1) of 15% as clinically important decline.[1] Alternatively, the American College of Occupational and Environmental Medicine (ACOEM) has proposed a longitudinal reference limit based on a 15% decline FEV_1 for working populations.[2] Neither of these recommended longitudinal limits takes into account the precision (that is, the within-person variability) of the existing spirometry data, thus prohibiting their use for evaluating the reliability of the predicted decline and for evaluating and enhancing the quality of the collected data.

Measurement errors can have a substantial effect on the amount of the within-person FEV_1 variability observed in longitudinal spirometry data and on the uncertainty associated with the estimated rate of FEV_1 decline.[3–6] For example, as a result of differences in standardised testing procedures and/or varying adherence to those procedures over time, there can be substantial systematic differences in FEV_1 data precision among monitoring programmes, even if they conform to ATS or European Respiratory Society (ERS) standards.[7 8] As most monitoring programmes conduct annual or less frequent testing, the number of measurements is usually insufficient to obtain a reliable estimate of an individual's within-person variation over a typical follow-up period. On the other hand, monitoring overall longitudinal data precision, using a pair-wise estimate of within-person variation estimated on a yearly basis on a group of workers, enables establishing overall data precision and detecting temporal variability in data precision in a spirometry testing programme.[8] Furthermore, the pair-wise estimate of within-person variation allows for the development of a limit of longitudinal decline (LLD) that will provide a relatively simple and practical method for quality control of an individual's longitudinal data, and for identification of excessive declines in FEV_1.[9]

....

Abbreviations: ACOEM, American College of Occupational and Environmental Medicine; ATS, American Thoracic Society; BHR, bronchial hyper-reactivity; COPD, chronic obstructive pulmonary disease; ERS, European Respiratory Society; FEV_1, forced expiratory volume in 1 second; LLD, limit of longitudinal decline; LLN, lower limit of normal; NIOSH, National Institute for Occupational Safety and Health

This paper further investigates the LLD method for facilitating interpretation of annual longitudinal changes in FEV_1. In particular, we further evaluate the statistical validity of the absolute limit method,[9] and propose and evaluate a relative LLD method. These methods have not been previously evaluated in the literature, and establishing their statistical validity across programmes with varying data precision and in groups with different demographic and respiratory health characteristics is needed. We address the following questions:

1. How valid is the ATS limit of 15% for the assessment of annual declines in workplace male and female populations, and in patients with COPD or asthma?

2. Is an absolute or per cent change criterion better for the identification of excessive declines?

3. What relative limit reference value would be appropriate for workplace monitoring programmes where testing adheres to ATS/ERS spirometry standards?

MATERIALS AND METHODS
Monitoring programmes studied
We provide examples of within-person variability in FEV_1 from three distinct sources: (1) two workplace monitoring programmes conducted in manufacturing plants, which we have previously studied for longitudinal data precision (Programme 1 and Programme 2);[8] (2) a spirometry monitoring programme done on more than 1600 fire-fighters (Programme 3); (3) a longitudinal epidemiological study of 5887 cigarette smokers with early COPD[10] (Programme 4).

All the programmes were conducted in the US on predominantly white males and females. The programmes used equipment and applied procedures and computational methods consistent with ATS spirometric criteria.[12] Workplace testing was administered by personnel who had successfully completed a National Institute for Occupational Safety and Health (NIOSH)-approved course in spirometric testing.[13] The two manufacturing plants and Programmes 4 used the same make of dry-rolling seal spirometer throughout their follow-up (Infodyne Systems 8L and Spirotech S500, respectively). Programme 3 used a dry-rolling seal spirometer (Spirotech S400) up to July 2001, and then switched to a flow-based spirometer (Renaissance II). The longitudinal spirometry data for all programmes included the largest back-extrapolated FEV_1 and FVC from the best three curves, and the ratio FEV_1/FVC computed from the largest values.

For Programmes 1 and 2, central quality assurance of the spirometric tests was done by one of the authors (HWG) in both flow-volume and volume-time format.[13 14] The spirometric data and responses to a standardised questionnaire on respiratory symptoms and disease was automatically computerised. Asthma was defined as a positive response to a question: ''Do you have asthma?'' Programme 3 was conducted by trained technicians[12] and no special central quality control was done; only demographic and spirometry data were available. Programme 4 was a randomised clinical trial designed to determine the effect of a smoking cessation intervention and bronchodilator use on the rate of decline in FEV_1 in cigarette smokers with early COPD.[10 15] Spirometry measurements were obtained over six annual visits in several centres using stringent standardised methods of testing.[10] The presence of bronchial hyper-reactivity (BHR) was established by a methacholine challenge conducted at the onset of the study. BHR was defined by the slope of methacholine dose-FEV_1 relation (Methacholine Challenge Slope, SMCT) as (difference between baseline FEV_1 and final FEV_1)/(highest dose of methacholine used).[15] Individuals were categorised into quartiles for SMCT; those in the lowest quartile were categorised as having BHR.

The present study was approved by the NIOSH Human Subject Review Board.

Statistical methods
We first evaluated the temporal changes in data precision over all years of follow-up for each monitoring programme, using the pair-wise measure of within-person variation s_p (see Appendix). Next, we calculated programme- and gender-specific average pair-wise estimates of within-person variation $\bar{s}_p$ and pair-wise estimates of relative within-person variation $\bar{s}_r$ (Appendix). Using these estimates of within-person variation, we then calculated programme- and gender-specific absolute and relative limits for an annual decline, and evaluated their agreement with the 95th percentiles for year-to-year changes in FEV_1. We also estimated these limits for individuals with asthma and COPD.

Limits of longitudinal decline
The absolute and relative limits of annual longitudinal decline for FEV_1 is defined as follows:

1. *Absolute limit of longitudinal decline* (LLD_a) (ml) is the approximate one-sided 95% confidence limit for longitudinal decline:[9]

$$LLD_a = t\,(b + 1.645 \times SE(b))$$

where b is the referent slope, for which we used the value of 30 ml/year,[16] and $t = 1$ represents one year of follow-up. The standard error of the slope b is given by a formula derived by Schlesselman:[17]

$$SE(b) = \sigma_w \sqrt{12(p-1)} \,/\, t\sqrt{p(p+1)} \qquad (1)$$

Here, $P = 2$ is the number of repeated measurements done during the follow-up time t of one year, and σ_w is the within-person standard deviation. Note that, in comparison to the duration of follow-up t, P has only a small effect on $SE(b)$ and on LLD_a.[9] By substituting the programme- and gender-specific values of $\bar{s}_p$ for the within-person variation σ_w in equation (2), we predicted the programme- and gender-specific LLD_a.

2. *Relative limit of longitudinal decline* (LLD_r) standardises for the size of FEV_1, and is defined as

$$LLD_r = t(b/\overline{FEV}_{1b} + 1.645 \times SE_r(b)) \qquad (2)$$

where b is standardised by the programme- and gender-specific mean baseline FEV_1, and $SE_r(b)$ is the approximate standard error of

$$b/\overline{FEV}_{1b}$$

calculated by substituting the programme- and gender-specific values of $\bar{s}_r$ for the within-person variation σ_w in equation (2).

For an individual, the reference limit for an annual FEV_1 decline may be calculated in terms of the level of FEV_1. Individuals whose FEV_1 values fall below this reference limit should raise concern. The limit can be calculated in terms of the individual's baseline (previous year's FEV_1) FEV_{1b} value and LLD_a, and LLD_r as follows:

$$FEV_1 = FEV_{1b} - LLD_a$$
$$\text{or } FEV_1 = FEV_{1b} - FEV_{1b} \times LLD_r.$$

Agreement between the limits of longitudinal decline and 95th percentiles
We evaluated the agreement between programme- and gender-specific LLD_a values and the 95th percentiles for all absolute

Table 1 Demographic characteristics (mean (standard deviation)) and follow-up characteristics for the four monitoring programmes

Programme	n†	Age (years) at baseline	FEV$_1$ (L) at baseline	Height (cm) at baseline	Mean duration of follow-up (years)
P1- Male	159	37 (10.0)	4.2 (0.8)	176 (5.9)	5.2
Female	37	34 (7.7)	3.1 (0.6)	162 (5.3)	5.1
P2- Male	714	37 (11.1)	3.9 (0.7)	179 (6.6)	5.4
Female	111	35 (9.5)	2.8 (0.6)	167 (6.8)	3.8
P3A-Male	1291	38 (9.9)	4.2 (0.7)	179 (10.3)	8.4
Female	49	32 (4.2)	3.6 (0.4)	170 (4.9)	7.8
P3B-Male	1213	40 (10.7)	4.1 (0.7)	178 (11.3)	3.5
Female	53	35 (6.4)	3.4 (0.5)	170 (5.1)	3.2
P4-Male	3510	48 (7.0)	2.9 (0.5)	177 (6.6)	4.9
Female	2087	49 (6.6)	2.1 (0.4)	164 (5.9)	4.9

†Number of individuals who had at least one follow-up measurement within 14 months.

year-to-year changes ($\Delta FEV_1 = FEV_{1b} - FEV_1$), and similarly we evaluated agreement between the LLD$_r$ values and the 95th percentiles for all relative year-to-year changes (%$\Delta FEV_1 = \Delta FEV_1/(FEV_{1b} + FEV_1)/2 \times 100$). To examine the relation between the two sets of measures, we first plotted the limits against the 95th percentiles and then tested the differences using the Bland-Altman method.[18]

Agreement between LLD values and 95th percentiles in individuals with respiratory disease

We also calculated the appropriate LLD$_r$ values for one year of follow-up for individuals reporting asthma (using data from Programme 2), and for individuals with early COPD, according to the severity of bronchial BHR (using data from Programme 4 where BHR was measured).

RESULTS

Table 1 provides programme- and gender-specific baseline demographic data and duration of follow-up for individuals who had at least one follow-up period within 14 months. Because Programme 3 had a marked change in data precision when a new spirometer started to be used (fig 1), we conducted the evaluations for two periods: years 1990–9 (Programme 3A), and years 2000–4 (Programme 3B). The three workplace programmes (Programmes 1 to 3) had similar demographic characteristics, but individuals in Programme 4, who had early COPD, were older and had lower mean FEV$_1$.

Figure 1 shows the time-related differences in data precision among the programmes, as measured by the yearly values of
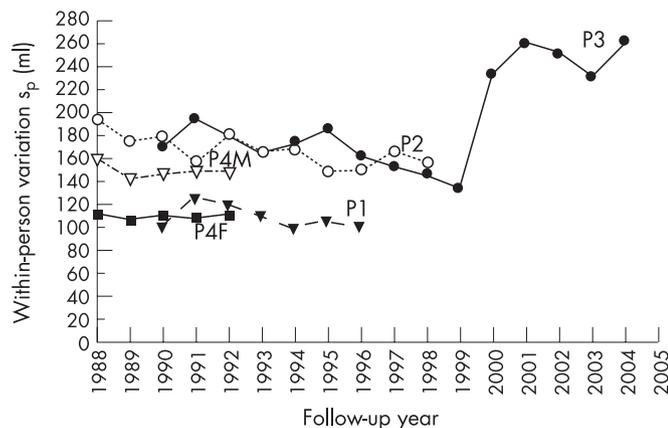
Figure 1 Yearly values of the pair-wise within-person variation s_p are plotted for the three work place monitoring programmes (P1, P2, P3) and for the COPD study population (subdivided by gender: P4M = men, P4F = women).

the pair-wise within-person variation s_p. For Programme 4 males and females are plotted separately; in the other programmes the effect of females on the overall s_p values was small and these were combined with males.

Table 2 shows the number of year-to-year intervals on which the calculation of the limits of annual decline and the 95th percentiles were based, and the mean duration of the intervals. For the absolute and relative limits, the table shows the programme- and gender-specific pair-wise within-person standard deviations $\bar{s}_p$ and $\bar{s}_r$, the calculated limits LLD$_a$ and LLD$_r$, and the 95th percentiles for ΔFEV_1 and %ΔFEV_1. Figure 2 shows the predicted LLD$_a$ values for $\bar{s}_p$ ranging from 70 ml to 270 ml (line), and the observed programme- and gender-specific 95th percentiles (points) plotted against $\bar{s}_p$. Notably, males have consistently higher within-person variability $\bar{s}_p$ and higher 95th percentiles by almost 100 ml, and males and females from Programme 4 have second lowest 95th percentiles and estimated LLD$_a$ values (table 2). The mean difference between LLD$_a$ and the 95th percentile for ΔFEV_1 was 15.6 (SD 21.7) ml/year. The LLD$_a$ method overestimated the 95th percentiles by about 16 ml/year, and all the individual differences were within the limits of agreement of -26.9 to 58.1 ml/year, calculated as 15.6 (SD 1.96).[18] The average difference of 15.6 ml/year is acceptable in view of the current recommended limit for annual decline of 15% (about 630 ml for FEV$_1$ of 4.2 l—the mean value for males from Programme 1).

There was also good agreement between the predicted LLD$_r$ and the 95th percentiles. Figure 3 shows the predicted LLD$_r$ for within-person variation $\bar{s}_r$ ranging from 2% to 7% (line) and the programme- and gender-specific 95th percentiles for %ΔFEV_1 (points) plotted against $\bar{s}_r$. The mean difference between LLD$_r$ and the 95th percentile for %ΔFEV_1 was 0.45% (SD 0.80%) (fig 4). The LLD$_r$ method overestimated the 95th percentiles by about 0.5% on average and all the differences ranged within the agreement limits of -1.1% and 2.0%. This magnitude of agreement is acceptable.

The practical advantage of LLD$_r$ over LLD$_a$ is that the adjustment for FEV$_1$ size had effectively adjusted for the gender differences seen in figure 2. Also, the adjustment for the smaller FEV$_1$ size for those with early COPD from Programme 4 resulted in the LLD$_r$ values being higher than those for the workplace monitoring programmes. Since Programme 4 had excellent quality control, measurement error is an unlikely reason for the higher relative variability in FEV$_1$, unlike in Programme 3B where measurement error was likely the main reason for the increased variability.

Effect of asthma and BHR on year-to-year variability in FEV$_1$

Table 3 gives the relative limit statistics for individuals who reported asthma (from Programme 2), and for individuals with

Table 2 Programme- and gender-specific pair-wise within-person standard deviations $\bar{s}_p$ and $\bar{s}_r$, the corresponding absolute LLD_a and relative LLD_r limits, and the 95th percentiles for observed year-to-year ΔFEV_1 and $\%\Delta FEV_1$

Programme	Gender	n†	Mean interval (months)	Absolute limit			Relative limit		
				$\bar{s}_p$ (ml)	LLD_a (ml/y)	95th percentile ΔFEV_1	$\bar{s}_r$ (%)	LLD_r (%)	95th percentile $\%\Delta FEV_1$
P1 (1990–6)	Male	281	10.5	114.6	297	310	2.9	7.4	7.6
	Female	63	10.4	74.6	204	190	2.5	6.6	7.5
P2 (1988–98)	Male	1930	12.2	174.0	435	440	5.1	12.6	12.2
	Female	225	12.1	137.2	349	346	5.1	12.5	12.9
P3A (1990–9)	Male	4304	11.9	161.7	406	380	4.2	10.6	9.9
	Female	147	11.6	133.0	339	320	4.1	10.2	9.2
P3B (2000–4)	Male	3062	11.5	246.9	604	560	6.3	15.3	13.8
	Female	127	11.3	222.2	547	490	6.5	15.8	14.3
P4	Male	15872	12.0	149.5	378	370	5.7	14.2	14.3
	Female	9570	12.0	109.3	284	280	6.0	15.0	15.8

†Number of year-to-year intervals on which statistics are based.

early COPD by quartiles of BHR (from Programme 4). Males and females were combined for this analysis since their results for the relative within-person variation did not practically differ. For individuals with asthma, the estimated LLD_r was 12.8%. For individuals with early COPD, the LLD_r was about 18% for those with greatest severity of BHR (BHRq1), and around 10% for individuals with least BHR (BHRq4). Thus in Programme 4, BHR associated with early COPD appears to be the main determinant of the increased within-person variability. In the quartile with least severe BHR, the within-person variability was comparable to that in Programme 3A.

DISCUSSION
A major issue in the interpretation of longitudinal spirometry data in worker monitoring programmes is the healthcare providers' uncertainty about spirometry data quality. Monitoring of longitudinal data precision over time using the pair-wise estimate of within-person variation helps to improve: (1) the quality of data collected by the programme and (2) the precision with which individual workers with excessive rate of decline are identified.

Identifying temporal changes in FEV_1 variation can lead to timely corrective actions on a programme's level. For example,

in Programme 3, a change of a spirometer in 2001 was associated with a substantial increase in data variability (fig 1) due to systematic procedural and equipment errors. The increased data variability was identified only in 2004, when the group within-person variability monitoring began. Assessment of spirometry quality identified major problems with the testing instrument itself (incorrect sensors functioning as a result of moisture accumulation) and incorrect use of the spirometer due to lack of technician training in using the instrument. Earlier recognition of this increased variation could have facilitated a more timely intervention.

Our study shows that using limits of annual decline that reflect within-person variability in the FEV_1 measurements facilitates improved precision of interpretation of annual declines in FEV_1 in an individual, especially during the early stages of follow-up. The longitudinal limit for an annual decline predicts the 95th percentile cut-off point for observed annual changes and thereby identifies 5% of individuals with excessive declines. This approach facilitates quality control on an individual basis, as it helps to identify individuals for whom spirometry quality control and/or respiratory conditions may need further investigation, or those who should have more frequent testing. Figure 2 provides guidance for selecting
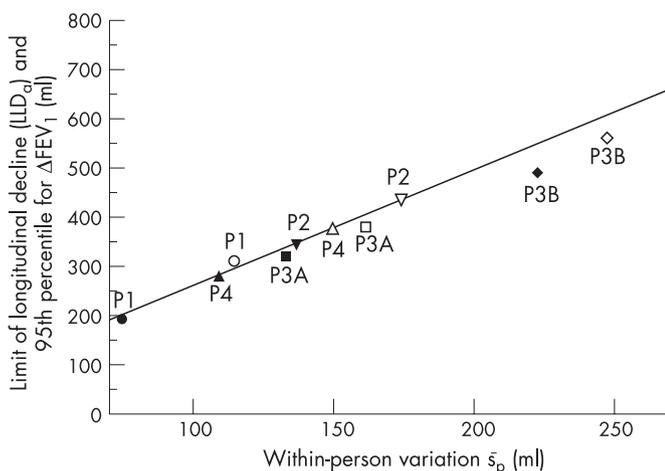
Figure 2 Predicted *absolute* longitudinal limit of decline LLD_a for FEV_1 (solid line) and programme- and gender-specific 95th percentiles for year-to-year declines (solid points for females, empty points for males) plotted against the mean absolute coefficient of precision $\bar{s}_p$, for Programme 1 (P1: ● and ○), Programme 2 (P2: ▼ and □), Programme 3A (P3A: ■ and □), Programme 3B (P3B: ♦ and ◇), and Programme 4 (P4: ▲ and △).
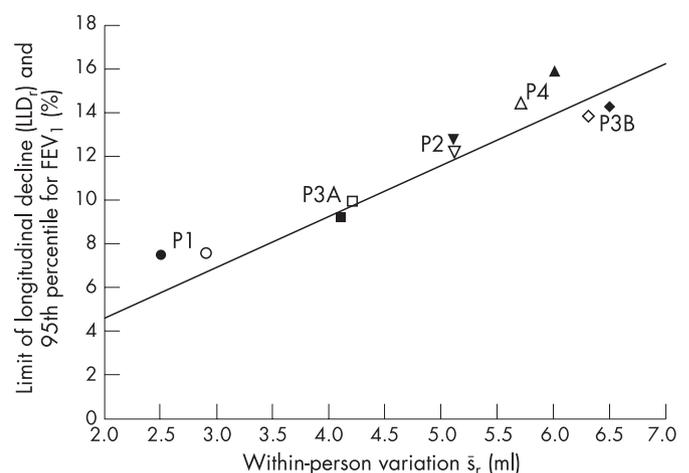
Figure 3 Predicted *relative* longitudinal limit of decline LLD_r (line) and programme- and gender-specific 95th percentiles for year-to-year declines (solid points for females and empty points for males) plotted against the mean relative coefficient of precision $\bar{s}_r$, for Programme 1 (P1: ● and ○), Programme 2 (P2: ▼ and □), Programme 3A (P3A: ■ and □), Programme 3B (P3B: ♦ and ◇), and Programme 4 (P4: ▲ and △).
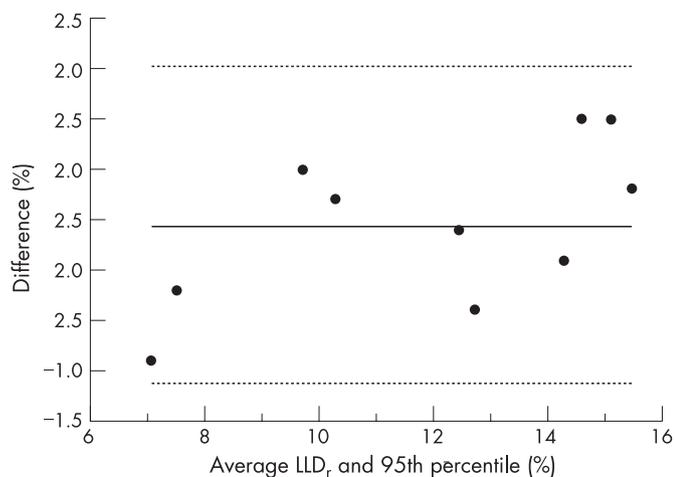
Figure 4 Bland-Altman plot of the programme- and gender-specific differences between LLD$_r$ and 95th percentiles for %ΔFEV₁. The mean difference was 0.45% and all the differences were within the limits of agreement of −1.1% and 2.0%, defined by the mean (SD 1.96).

absolute annual declines that should be considered excessive depending on the existing data within-person variation. (The absolute values are mostly restricted to Caucasian populations.) For example, in Programme 3A, an annual decline greater than 380 ml/year and 320 ml/year can be considered excessive in males and females, respectively. However, after 1999 only respective declines greater than 560 ml/year and 490 ml/year can be considered excessive in Programme 3B (table 2). This example demonstrates the impact of data precision on what decline can be detected as excessive. More precise early warning is especially important in workplace situations where occupational exposure has been shown to be associated with rapid excessive loss of lung function that can lead to disabling respiratory disease, as, for example, in a study of popcorn workers.[19]

Comparing the absolute and relative limits, our study shows, however, that the relative limit LLD$_r$ has more general validity and is more practical, as it adjusts for FEV₁ size. Table 3 and figure 3 demonstrate that the relative limit LLD$_r$ effectively adjusts for gender differences, which turned out to be mainly due to FEV₁ size; male and female groups thus could use the same longitudinal limit. The LLD$_r$ also adjusts for differences in mean FEV₁ size among different population groups as indicated by increased relative within-person variability in Programme 4, which was found to be mainly due to increased BHR.

The study also shows that it is possible for workplace monitoring programmes conducted by trained technicians using ATS/ERS standards[7] to achieve good data precision corresponding to LLD$_r$ of about 10%. This corresponds to average relative within-person variation $\bar{s}_r$ of about 4.0% (fig 3,

programme P3A). This degree of precision allows identification of a ''true'' rapid decliner of 90 ml/year after about five years of follow-up.[9] A recent study of spirometry monitoring data from a chemical plant also reports the 95th percentile for annual decline to be 10.4% for males and 10.6% for females.[20] For the cases with early COPD who did not have BHR, the limit of annual decline was also around 10% (table 3). A previous report on the Lung Health Study (Programme 4), reported 95% of differences between FEV₁ measurements taken 21 days apart to be within 240 ml for females and within 320 ml for males.[21] We estimated 95% of the annual declines for Programme 4 to be within 280 ml for females and 370 ml for males (table 2). Similarly, results from a cohort of 389 blue-collar male workers with good quality spirometry data reported a yearly decline in FEV₁ of 8% or 330 ml for healthy workers and good quality spirometry data.[22] Our study provides a simple general statistical framework through which these published results could be interpreted.

However, unlike our study, a study investigating short-term (less than 3 months) changes in FEV₁ in patients with COPD recommended using an absolute limit of 225 ml, irrespective of the baseline level of FEV₁.[23] The authors noted that the absolute difference in FEV₁ between two spirometry sessions did not vary with the baseline FEV₁. We found that FEV₁ variability and the absolute longitudinal limit LLD$_a$ were related to gender, data precision, and to the lower mean FEV₁ size in older individuals with early COPD, and that the relative limit adjusts for these differences. The authors[23] suggested as an alternative approach to accept a per cent difference greater than 10% with an absolute change of at least 150 ml as a short-term limit of decline. This result is in agreement with findings from our study. Nevertheless, we show that this limit may not be appropriate for individuals with BHR or asthma, or in monitoring programmes with poor spirometry quality (table 4).

Our study shows that clinical conditions such as airway hyper-responsiveness and asthma increase FEV₁ variability. Another study found that week-to-week variability in per cent change was greater in adult patients with asthma and COPD than in normal adults, and recommended that a significant change was 12% in normal individuals and 23% in adults with obstructive disease.[24] These results support using the ATS criterion for annual FEV₁ decline of 15% for individuals with asthma or COPD with increased BHR in clinical practice. However, in workplace monitoring programmes, where the number of individuals with asthma and COPD is usually small, those individuals should have the same stringent LLD$_r$ criteria as that applied to all the workers in the programme.

A limitation of the proposed method is that in situations where occupational exposure is associated with increased excessive decline in lung function and/or increased within-person variability, the derived limit may increase as a result of these two effects. As the excessive decline is unlikely to affect a large proportion of the workforce, its impact on within-person

Table 3 The relative pair-wise within-person variation $\bar{s}_r$, and the relative limit of longitudinal decline LLD$_r$ for individuals with asthma and with COPD by quartiles of bronchial hyperreactivity (BHR)

Program	Condition	%ΔFEV₁ (n)	$\bar{s}_r$ (%)	LLD$_r$ (%)	95th percentile %ΔFEV₁
P1	Asthma	83	5.2	12.8	11.6
P4	COPD & BHRq1†	7112	7.3	17.9	18.9
P4	COPD & BHRq2	6575	5.8	14.6	15.3
P4	COPD & BHRq3	8830	4.9	12.3	12.2
P4	COPD & BHRq4	2925	4.3	10.9	10.5

†BHRq1-q4 represent quartiles of BHR; q1 is highest and q4 is lowest degree of BHR.

variation is likely to be relatively small in comparison to the effect of measurement errors (Appendix). However, excessive within-person variability can only be identified if it is monitored, and the causes of increased within-person variability investigated. This study provides data on acceptable within-person variability values, and the proposed method will be more sensitive than the limit based on the 15% only if the relative pair-wise within-person variation $\bar{s}_r$ is less than 6.5% (fig 3). Some systematic error resulting from aging may have inflated our estimates of the within-person standard deviation, but in a preliminary analysis we found the effect from this factor can be ignored for practical purposes. Although the number of measurements repeated on each individual within 14 months varied by the programmes, this effect was unlikely to significantly affect the precision of the estimated pair-wise within-person variance since the yearly values shown in figure 1 were consistent.

In conclusion, we show the following. (1) The limits of longitudinal decline (LLD_a and LLD_r) method is valid and practical for evaluating annual changes in FEV_1 for efficient detection of persons with excessive decline or for detection of measurement errors. (2) Assessment of longitudinal data precision using the pair-wise estimate of within-person variation, s_p or s_r, can help to identify extraneous sources of variability in FEV_1 soon after they arise in a group, and implement interventions. (3) The LLD_r for annual decline should be about 10% or less for good quality workplace monitoring programmes; the ATS recommended 15% criterion appears excessive. (4) For individuals with airways disease associated with BHR (that is, asthma or COPD), the ATS-recommended 15% criterion for annual decline appears appropriate for clinical practice. Use of the longitudinal limit methods described here enables: (1) to take into consideration data quality when interpreting the longitudinal data, and (2) to identify longitudinal data that are of substandard quality so that intervention on an individual and programme level can be taken to improve data quality. Based on the described method, we have developed computer software for the analysis and interpretation of individual subjects' longitudinal spirometry data; this software is available on request from the first author.

Authors' affiliations
E Hnizdo, K Sircar, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Morgantown, WV, USA
T Yan, Constella Group, Morgantown, WV, USA
P Harber, Division of Occupational and Environmental Medicine, Department of Family Medicine, University of California Los Angeles, Los Angeles, CA, USA
J Fleming, Phoenix Fire Department, Phoenix, AZ, USA
H W Glindmeyer, Section of Pulmonary, Critical Care, and Environmental Medicine, Department of Medicine, Tulane Medical School, New Orleans, LA, USA

REFERENCES
1 **American Thoracic Society**. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Resp Dis* 1991;**144**:1202–18.
2 **ACOEM**. Position statement: Evaluating pulmonary function change over time in the occupational setting. *J Occup Environ Med* 2005;**47**:1307–16, Available at http://www.acoem.org/guidelines.aspx?id=756 (accessed 2 August 2007).
3 **Becklake MR**, White N. Sources of variation in spirometric measurements. Identifying the signal and dealing with the noise. *Occup Med* 1993;**8**:241–61.
4 **Glindmeyer HW**, Jones RN, Diem JE, et al. Useful and extraneous variability in longitudinal assessment of lung function. *Chest* 1987;**92**:877–82.
5 **Berry G**. Longitudinal observations. Their usefulness and limitations with special reference to the forced expiratory volume. *Bull de Physio-Path Respir* 1974;**10**:643–55.
6 **Hankinson JL**, Wagner GR. Medical screening using periodic spirometry for detection of chronic lung disease. *Occup Med* 1993;**8**:353–61.
7 **Miller MR**, Hankinson J, Brusasco V, et al. Standardization of spirometry. *Eur Respir J* 2005;**26**:319–38.
8 **Hnizdo E**, Yu L, Freyder L, et al. Precision of longitudinal lung function measurements—monitoring and interpretation. *Occup Environ Med* 2005;**62**:695–701.
9 **Hnizdo E**, Sircar K, Glindmeyer HW, et al. Longitudinal limits of normal decline in lung function in an individual. *JOEM* 2006;**48**:625–34.
10 **Enright PL**, Johnson LR, Connett JE, et al. Spirometry in the Lung Health Study 1. Methods and quality control. *Am Rev Respir Dis* 1991;**143**:1215–23.
11 **American Thoracic Society**. Standardization of spirometry - 1987 update. *Am Rev Respir Dis* 1987;**136**:1285–98.
12 **CDC NIOSH**. The NIOSH-spirometry training course. Available at http://www.cdc.gov/niosh/topics/spirometry (accessed 28 March 2006).
13 **Glindmeyer H**, Lissack T. A network approach to national respiratory health care. In: Rubin M, eds. *Computerization and automation in health facilities*. Boca Raton, FL, USA: CRC Press, Inc, 1984:1–16.
14 **Glindmeyer HW**, Jones RN, Barkman HW, et al. Spirometry: quantitative test criteria and test acceptability. *Am Rev Respir Dis* 1987;**136**:449–52.
15 **Tashkin DP**, Altose MD, Bleecker ER, et al. The lung health study: airway responsiveness to inhaled methacholine in smokers with mild to moderate airflow obstruction. The Lung Health Study Research Group. *Am Rev Respir Dis* 1992;**145**:301–10.
16 **Sherrill DL**, Lebowitz MD, Knudson RJ, et al. Continuous longitudinal regression equations for pulmonary function measures. *Eur Respir J* 1992;**5**:452–62.
17 **Schlesselman JJ**. Planning a longitudinal study: II. Frequency of measurement and study duration. *J Chron Dis* 1973;**26**:553–8.
18 **Bland JM**, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurements. *Lancet* 1986;**1**:307–10.
19 **Akpinar-Elci M**, Travis WD, Lynch DA, et al. Bronchiolitis obliterans syndrome in popcorn production plant workers. *European Resp J* 2004;**24**:298–302.
20 **Wang ML**, Vashia BH, Petsonk LE. Interpreting periodic lung function tests in individuals. The relationship between 1- to 5-year and long-term FEV1 changes. *Chest* 2006;**130**:493–9.
21 **Enright PL**, Connett JE, Kanner RE, for the Lung Health Study Group, et al. Spirometry in the Lung Health Study: II Determinants of short-term intraindividual variability. *Am J Respir Crit Care Med* 1995;**151**:406–11.
22 **Wang M**, Petsonk EL. Repeated measures of FEV1 over six to twelve months: what change is abnormal? *J Occup Environ Med* 2004;**46**:591–5.
23 **Herpel LB**, Kanner RE, Lee SM, for the Lung Health Study Research Group and the National Emphysema Treatment Trial Research Group, et al. Variability of spirometry in chronic obstructive pulmonary disease. Results from two clinical trials. *Am J Respir Crit Care Med* 2006;**173**:1106–13.
24 **Pennock BE**, Rogers RM, McCaffree DR. Changes in measured spirometric indices; what is significant? *Chest* 1981;**80**:97.

APPENDIX

DATA PRECISION OF THE MONITORING PROGRAMMES
The pair-wise estimate of within-person standard deviation was calculated as

$$S_p = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}\left(FEV_{1_{1i}} - FEV_{1_{2i}}\right)^2}$$

for each year of follow-up on the n individuals who had two consecutive measurements $FEV1_{1i}$ and $FEV1_{2i}$ done within about 12 months (maximum range 14 months). The date of the first test determined the assigned year of follow-up, and only one comparison per person per year was used. To avoid systematic effect of age on FEV_1 change, the interval should be relatively short, but still long enough to reflect most potential sources of within-person variability. We used the interval of 12 months for practical reasons. For practical purposes we also neglected the effect that the average decline within the 12 months has on the s_p estimate. Theoretically, s_p will overestimate the "true" within-person variation by the

average rate of decline b in a group by approximately $\sqrt{(b^2/2)}$. Thus, for average slopes of 30 ml/year (healthy never-smokers), 60 ml/year (current smokers) and 90 ml/year (an extreme situation), s_p will be inflated by about 21, 42 and 64 ml, respectively. The differences from the 30 ml/year slope are 21 and 43 ml for the 60 and 90 ml/year average slopes, respectively. In most workplace monitoring situations, the fraction of workers with a "true" slope greater than 90 ml/year is usually relatively small. We excluded also extreme year-to-year outliers $|\Delta FEV_1| > |1.7|$ l, which constituted less than 0.1% of the ΔFEV_1 values overall; the few extreme outliers can cause large deviations in s_p values and misrepresent the average within-person variation.

The programme- and gender-specific average within-person variation $\bar{s}_p$ was then calculated by calculating the s_p values over all years of follow-up. These average $\bar{s}_p$ values agreed well with the within-person standard deviation estimated by the mixed-effects model.

The pair-wise estimate of the relative within-person standard deviation, which adjusts for the individual's FEV$_1$ size, was defined as[21]

$$S_r = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} \left(\frac{FEV_{1_{1i}} - FEV_{1_{2i}}}{(FEV_{1_{1i}} + FEV_{1_{2i}})/2} \right)^2}$$

and was calculated together with its programme- and gender-specific averages, $\bar{s}_r$, using the same method as that for the calculation of $\bar{s}_p$.