ORIGINAL ARTICLE

The precision of longitudinal lung function measurements: monitoring and interpretation

E Hnizdo, L Yu, L Freyder, M Attfield, J Lefante, H 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; ehnizdo@cdc.gov

Accepted
16 February 2005
....

Background: The efficacy of decision making based on longitudinal spirometric measurements depends critically on the precision of the available data, which is determined by the magnitude of the within-person variation.

Aims: Firstly, to describe and investigate two statistical methods—a pairwise estimate of within-person standard deviation s_p and the reliability coefficient G—for use in the monitoring of precision of longitudinal measurements of forced expiratory volume in one second (FEV1). Secondly, to investigate the effect of longitudinal data precision on the detectable excess rate of decline in FEV1.

Methods: The authors "monitored" retrospectively on a yearly basis the magnitude of the within-person variation s_p and the coefficient G in 11 workplace based spirometric monitoring programmes conducted from 1987 to 2001 on 12 729 workers in various industrial plants.

Results: The plant-specific mean values $\bar{s}_p$ (range 122–166 ml) and $\bar{G}$ (range 0.88–0.95), averaged over all years of follow up, correlated well with the plant-specific within-person standard deviation s_r (range 130–177 ml) estimated from all longitudinal data. The correlations were 0.90 for $\bar{s}_p$ and 0.68 for $\bar{G}$. The average precision of the longitudinal FEV1 measurements affected the duration of follow up needed to identify a "true" excess rate of decline in FEV1 in an individual.

Conclusions: The results show that monitoring of longitudinal spirometry data precision (1) allows that data precision can be improved or maintained at levels that allow individuals with a rapid decline to be identified at an earlier age; and (2) attaches a measure of precision to the data on which decision making is based.

C hronic obstructive pulmonary disease (COPD) is generally a slowly progressive airway disease that produces a decline in lung function that is not fully reversible. According to the World Bank/World Health Organization, COPD is expected to rise to the 5th ranked burden of disease by the year 2020.[1] COPD is most frequent in blue collar workers where tobacco smoking, occupational exposure, and socioeconomic status all contribute to the increased risk of the disease.[2] Thus, prevention of the development of COPD is an important public health issue worldwide and workplace screening may help in the prevention. The spirometry tests of forced vital capacity (FVC) and forced expiratory volume in one second (FEV1) are the recommended tests for the diagnosis of COPD.[1][3] Of the spirometry tests, the FEV1 is the most reproducible and best suited for measuring changes in lung function over time.[4] Longitudinal FEV1 data allow us to study the rate of change in lung function, and to identify individuals and groups with an increased decline in lung function for an early intervention.[5–9] Potentially, workplace based spirometry monitoring can provide a valuable tool for an early recognition of excessive rate of lung function decline in an individual, that may reflect development of chronic lung diseases caused by occupational or environmental exposures, including smoking. However, the efficacy of the decision making based on longitudinal spirometry data in workplace monitoring programmes depends critically on the precision of the available longitudinal data—that is, the magnitude of within-person variation.[10]

The American Thoracic Society (ATS) provides guidelines for spirometry quality control that help to decrease measurement errors within single test sessions.[11] There is still a need, however, for statistical monitoring of precision when collecting longitudinal measurements over several years.[12] Such monitoring would enable one to investigate and reduce extraneous sources of random variation shortly after these arise, and, at the same time, attach a measure of precision to the data on which the decision making is being done.

In longitudinal spirometric data, precision is determined primarily by the magnitude of the within-person variation in lung function.[13–15] The sources of the within-person variation can be broadly categorised as those arising from measurement procedures (for example, spirometer, subject, or technician procedural errors) and those arising from the within-person fluctuation in lung function around its "true" value.[14][15] Figure 1 illustrates longitudinal FEV1 values for three individuals in our study who had different levels of within-person standard deviation S_{ri} around their individual linear regression lines. When examining an individual person's longitudinal data it is important to know the magnitude of the average within-person variation for the group. This statistic provides an indication of the overall precision of measurements in a specific monitoring programme and influences how to interpret yearly declines that may be excessive.

The objective of the present study is to describe and investigate two statistical methods—a pairwise estimate of within-person standard deviation s_p and the reliability coefficient G—for use in the monitoring of the magnitude of the average within-person standard deviation σ_w (that is, data precision) in longitudinal FEV1 measurements in a group.[12][16][17] Using data from 11 large spirometry screening programmes,[18] we investigated the ability of the two statistics to predict the magnitude of the within-person standard

....

Abbreviations: ATS, American Thoracic Society; COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in one second; FVC, forced vital capacity; LLNR, lower limit of normal for the regression line
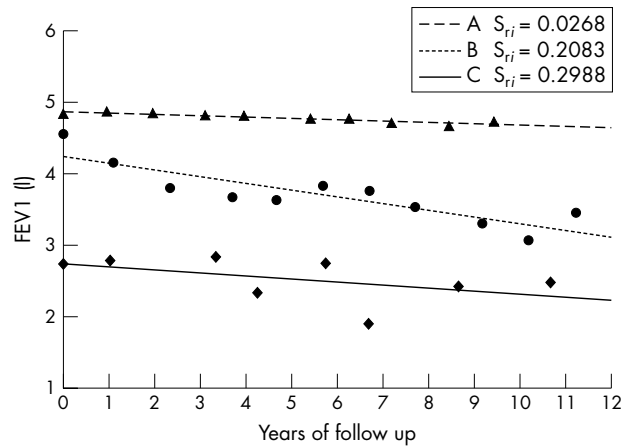
Figure 1 Examples of magnitudes of within-person standard deviation s_{ri} around an individual person i predicted line. For person A the within-person variation around the predicted slope is small (0.0268), for person B it is larger (0.2083), and for person C it is very large (0.2988).

deviation as estimated from all longitudinal measurements. Secondly, we also investigated how the precision of longitudinal FEV1 measurements impacts on the detection of the ''true'' excess rates of decline in an individual.

MATERIALS AND METHODS
Lung function monitoring programmes
In our study we used data from spirometry monitoring programmes implemented in 11 industrial plants during the period 1987–2001. Pulmonary function testing and a medical, smoking, and occupational questionnaire were administered by trained personnel (NIOSH Spirometry Course # 002 presented by one of the investigators, HWG), using a computer system designed to collect such data from remote facilities.[18] This system employs a computer with a resident questionnaire and an online, 8 l, dry rolling seal volumetric spirometer. Validation of the accuracy of this spirometric system[19] has shown that it complied with ATS spirometric test criteria.[20 21] A 3 l calibrating syringe was used for daily calibration. Testing was conducted in the standing position with nose clips, and height was measured without shoes. Spirometric test results were taken from at least three acceptable tests with good initial effort (extrapolated volume less than 5% of the FVC, with distinct superimposable, forced expiratory flow volume curves),[22] good continued effort for at least seven seconds, and repeatable FEV1 and FVC values within 5% or 100 ml. The final database included the largest FEV1 and FVC, and FEV1/FVC computed from the largest values. Quality assurance of the spirometric tests was done by one of the investigators (HWG).

Per cent predicted lung function values were computed for all study subjects using race and sex specific (White and African-American) prediction equations which included height, age, and age[2]. These equations were developed from blue collar never-smokers who denied occupational inhalant exposures, and who were tested on the same type of equipment as in the present study.[23]

The individual worker's participation in the monitoring programme in each plant was voluntary and began either when the monitoring programme started or when a worker became employed at that facility, and stopped either on cessation of monitoring or on cessation of employment. For the purpose of our analysis, time of follow up was represented either by the calendar year of lung function testing (1987–2001) or by the years of follow up. The data from workers who were age 20 years or older were used to

estimate the yearly within-person variation, the coefficient G, and the longitudinal within-person variation estimated by the mixed effects model. The data from workers who were 25 years and older and who had at least three measurements with five or more years of follow up were used to estimate the longitudinal within-person variation by the linear regression analysis.

Statistical methods
Outline
We ''monitored'' the precision of FEV1 measurements on a yearly basis within a period 1987–2001 separately in each of the 11 plants. We used two statistics to monitor the average within-person variation in each plant: (1) the pairwise estimate of within-person standard deviation s_p, and (2) the reliability coefficient G. To evaluate usefulness of these estimates, we correlated the plant specific mean values $\bar{s}_p$ and $\bar{G}$ (calculated as the averages of the yearly values of s_p and G, respectively), to the average plant specific within-person standard deviation estimated from all longitudinal data by the linear regression model (s_r) or the mixed model (s_m). Finally, we showed how the plant specific within-person variation impacts on the ability to detect significant excess rates of decline in FEV1 in an individual. The following four subsections provide further details on the statistical methods.

Estimation of the group within-person variation from pairwise measurements
Consecutive measurements of FEV1 (M_a and M_b), taken within a short duration from each other on a group of workers can be used to assess the group average within-person variability as follows.

1. The pairwise estimate of the within-person standard deviation can be estimated from the difference between the two measurements M_a and M_b as

$$S_p = \frac{1}{\sqrt{2}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} (M_{ai} - M_{bi})^2}$$

where the summation in a specific plant is over n subjects.

2. The coefficient of reliability G can be estimated from the following formula:

$$G = S_b^2 / (S_b^2 + S_w^2)$$

where S_b^2 is the between-subject variance and S_w^2 is the within-person variance. For a given S_b^2, increasing S_w^2 leads to a lower G value—that is, a lower data precision. A simple method of estimation of the coefficient G is to calculate the Pearson correlation coefficient r_{MaMb} on the consecutive measurements M_a and M_b.[12 16 17] To remove variation due to systematic population effects such as age, height, sex, or race, the correlation should be done on the percent predicted values. (Because an additional source of error can be introduced by errors in these covariates, these should be kept constant as much as possible.)

The interval between two measurements M_a and M_b should be sufficiently long to include all potential short term random effects, but short enough to avoid time related systematic changes—for example, those due to age. In the present analysis, year specific values of s_p and G were calculated by using two FEV_1 measurements repeated within 18 months; this period was chosen for practical reasons for yearly monitoring. We have established previously[12] and in our current study, that the value of the coefficient G does not

change much when the period between two measurements is increased from 12 to 18 months. The reason for this is that the variability in the expected decline in FEV1 per year is much smaller than the random variation around the slope.[13] The date of the first test determined the follow up year. This procedure gave us a sufficient number of tests per year while not noticeably affecting the yearly values of s_p and G when compared with a 12 month period. In the few cases where there were more than two repeated tests, only the first pair of tests per subject was used within the follow up year. Note that this strategy is appropriate for workplace monitoring programmes where annual measurements are available on all individuals or on a random sample of individuals.

Estimation of the within-person variation from all longitudinal data

The average plant-specific within-person standard deviation was estimated from all the longitudinal data using the following two methods:[24 25]

1. Estimation by the two-stage method. In the first stage, we fitted the linear regression model, $FEV1_i = \beta_{0i} + \beta_{1i}\cdot time + \epsilon_i$, to each i-person's vector of repeated FEV1 measurements and the time covariate (years of follow up). The individual within-person variance S_{ri}^2 was estimated by the regression MSE_i. In the second stage, we estimated the average plant-specific within-person standard deviation s_r as the square root of the mean S_{ri}^2 calculated across all subjects. (Because of the relatively short follow up time, we assumed that the longitudinal FEV$_1$ values were linearly related to time, with a constant variance.) We estimated the plant-specific s_r using individuals with five or more years of follow up.[26]

2. Estimation by the mixed effects model. The mixed model is represented by the equation $Y_i = X_i\beta + Z_ib_i + \epsilon_i$, where, for the i-person, Y_i represents the vector of repeated FEV1 measurements, X_i is the matrix of fixed (population) covariates, β is the vector of unknown regression coefficients for the population covariates, Z_i is the matrix of random subject-specific time covariates, and b_i is the vector of unknown subject-specific estimates of random effects (that is, the intercept and slope for time). We fitted the above mixed model to each plant-specific set of data to estimate the plant-specific within-person standard deviation s_m from the residual variance.[24 25] Individuals with at least one follow up measurement were included in this analysis (n = 6440). The fixed effects included in our model were age, height, sex, race, wheezing, smoking status, and time variant pack years (that is, pack years cumulated with increasing years of follow up).

Impact of group data precision on identifying excessive decline in FEV1

To illustrate the effect of the group data precision on our ability to identify an excess rate of decline in FEV1 in an individual, we used as an example two people, both having an FEV1 of 4 l at 34 years of age and an expected "normal" rate of decline of 30 ml/year,[27] but one being from a plant where data are collected with high precision (PLANT-HP) and the other from a plant where data are collected with low precision (PLANT-LP).

Firstly, we derived the approximate longitudinal one sided 95% confidence limit for an expected decline, b, over a specific duration, D (years). This limit is given by $D\times[b + 1.645\times SE(b)]$. Subtracting this limit from the initial FEV1 measurement gives the lower limit of normal for the FEV1

measurement (LLNR), as derived using the formula:[28 29]

$$LLNR = \text{ initial FEV1-} D\times[b + 1.645\times SE(b)]$$

where D is duration of follow up in years starting at $D = 1$ at 35 years of age and incrementing by one at each year, and the term $[b + 1.645\times SE(b)]$ determines the maximal regression slope of decline for an expected FEV$_1$ decline (that is, 30 ml/year). The standard error of the slope b, $SE(b)$, is derived from the formula derived by Schlesselman:[30]

$$SE(b) = \sigma_w\sqrt{12(P-1)}/D\sqrt{P(P+1)}$$

where P is the number of tests done during D years of follow up. Schlesselman's method of estimation was also applied previously to estimate sample size for longitudinal spirometry studies.[13 31 32]

The value of σ_w in the SE(b) formula was estimated by the average within-person standard deviation s_r for plants in our study that had the highest precision PLANT-HP and lowest precision PLANT-LP, respectively.[30] Because the estimate of within-person variation σ_w is derived from a large number of subjects, we assume the normal standard deviate for a one sided limit when p value is 5% to be $Z_\alpha = 1.645$.

The estimate of LLNR was based on SE(b) derived for a specific number of tests P done during the D years of follow up. Similarly, $[b + 1.645\times SE(b)]$ provides a limit of normal rate of decline given the parameters σ_w, D, P. Any individuals having observed declines below or rates of decline above these two types of limits would be detected as having an excess decline. Hence in subsequent discussion we refer to these two criteria as the detectable excess decline and the detectable excess rate of decline.

Variability in longitudinal data

To investigate how well the estimated LLNR agree with the observed data, we also show variability in the rate of change in FEV1 observed within one year, summed over all individuals and years, for the plants with the highest and lowest data precision. The rate of change in FEV$_1$ (l/year) was calculated as a difference between two repeated measurements done within 12 months of each other.

RESULTS

Table 1 gives, for the 11 plants, the number of workers who participated in the monitoring by follow up years. Table 1 also shows the number of workers who had at least three measurements over five or more years of follow up after 25 years of age (n = 3130). Figure 2 shows the plant-specific yearly values of within-person standard deviation s_p; the plants with consistently high and low values of s_p are indicated by thicker lines.

Table 2 shows the plant-specific values of $\bar{s}_p$ and $\bar{G}$ calculated as means of the yearly values, and their coefficients of variation (CV). Table 2 also shows plant-specific within-person variations s_r and s_m estimated from longitudinal data using the two-stage method and the mixed model. The correlations between the variables are shown at the bottom of the table.

Table 2 shows that, on the basis of the values of $\bar{s}_p$, plants 11 and 3 have the highest and lowest data precision, respectively. These two plants are used to represent PLANT-HP and PLANT-LP, respectively. For PLANT-HP, $\bar{s}_p = 0.122$ l, $\bar{G} = 0.954$, $s_r = 0.130$ l, and $s_m = 0.124$ l. For PLANT-LP, $\bar{s}_p = 0.166$ l, $\bar{G} = 0.898$, $s_r = 0.177$ l, and $s_m = 0.173$ l.

Figure 3 shows the longitudinal lower limits of normal LLNR for the "normal" decliners from PLANT-HP (LLNR–HP) and PLANT-LP (LLNR-LP), respectively. The LLNR were
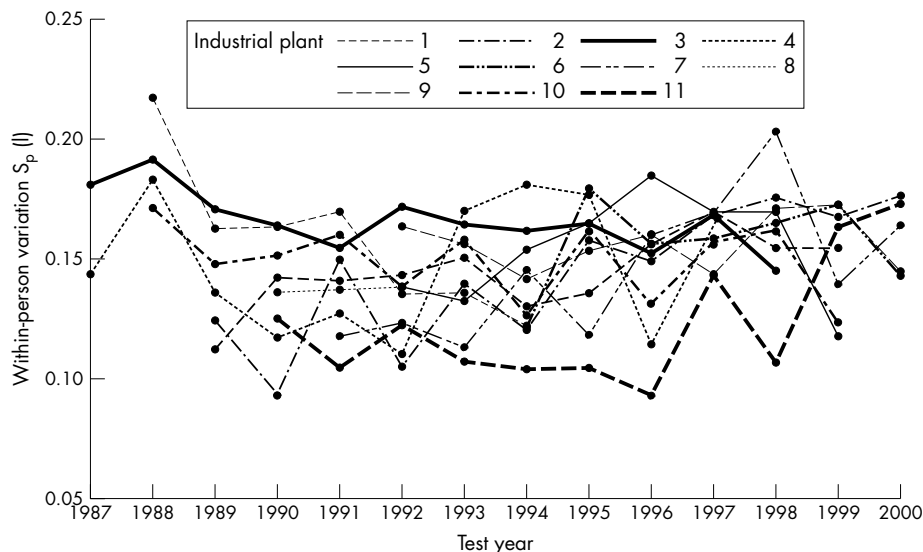
Figure 2 The plant-specific yearly values of pairwise within-person standard deviation s_p. Based on average within-person standard deviation $\bar{s}_p$, plant 3 (thick solid line) has lowest data precision and plant 11 (dashed thick line) has highest data precision.

estimated for $P = 2$, that is, for two tests done at the start of monitoring at age 34 and at various specific ages thereafter. Figure 3 also shows a predicted line for a person with a rapid decline of 60 ml/year and the ages at which a rapid decliner's predicted line crosses the longitudinal LLNR-HP or LLNR-LP.

Figure 4 shows on the left vertical axis the detectable excess decline in FEV1 (ml) and on the right vertical axis the detectable excess rate of decline in FEV1 (ml/year) defined by the limit of normal for an individual person after D = 1, 2, 3, 4, 5, 8, 20 years of follow up, for a known magnitude of the within-person variation σ_w, based on two measurements. The solid line for one year of follow up represents the detectable excess decline and also the detectable excess rate of decline. According to this line, the detectable excess decline is ≈ 260 ml when $\sigma_w = 100$, ≈ 375 ml when $\sigma_w = 150$, and ≈ 500 ml when $\sigma_w = 200$.

The right vertical axis and the solid lines of the figure show how the detectable excess rate of decline becomes smaller as the duration of follow up increases (the procedure becomes more sensitive). For example, with a data precision $\sigma_w = 150$, the detectable excess rate of decline decreases with

an increasing duration of follow up, D = 1, 2, 3, 4, 5, 8, 20, as 375 ml/year, 200 ml/year, 150 ml/year, 120 ml/year, 100 ml/year, 80 ml/year, and 50 ml/year, respectively. Note that one needs data precision of $\sigma_w = 130$ to detect a "true" slope of 90 ml/year after five years of follow up; this slope was recommended by the American College of Occupational and Environmental Medicine (ACOEM) to be considered as significant loss of lung function.[33] [34]

To establish how well the estimated excess decline for one year agrees with observed data, we calculated the percentile statistics for the observed yearly changes in FEV1 (calculated across all subjects and all years of follow up) for plant 11 and plant 3. These yearly changes represent the observed yearly fluctuation in FEV1. For plant 11, 95% of the negative changes (that is, the 5th percentile) were within −342 ml per year. For plant 3, 95% of the declines were within −433 ml per year. When we calculated the observed yearly changes in FEV_1 for groups of plants with $s_r \approx 0.15$ (plants 5, 18, 22) and plants with $s_r \approx 0.16$ (plants 8, 9, 13, 15) (table 2), the 5th percentiles were −0.394 ml and −0.389 ml, respectively. These results agree approximately with our estimates from figure 4, for one year of follow up.

Table 1 Plant-specific data on the number of workers who participated in the monitoring by follow up year, and the number of subjects who participated in the longitudinal estimation by the regression analysis (that is, ≥25 years of age with ≥3 tests and ≥5 years of follow up)

| Plant | Number by follow up years | | | | | | Subjects ≥25 years of age with ≥3 tests and ≥5 years of follow up | | |
	0	1	2	3	4–5	6+	n	Age*	Years†
1	777	422	359	275	124	90	120	36.8	8.51
2	995	583	481	418	343	255	335	38.2	8.83
3	2316	901	783	666	408	222	402	36.4	7.95
4	966	550	477	442	341	259	326	34.6	7.92
5	1192	760	684	602	468	348	458	35.1	8.43
6	1561	801	652	508	264	43	256	36.4	6.04
7	980	631	584	537	446	366	438	37.4	8.81
8	878	510	409	323	223	157	219	38.6	8.21
9	802	407	362	318	174	101	167	39.7	7.77
10	1655	567	487	417	308	245	296	39.9	8.47
11	607	308	276	240	117	60	113	36.1	7.67
Total	12729	6440	5554	4746	3216	2146	3130	37.2	8.06

*Mean age at the start of follow up.
†Mean years of follow up.

Table 2 The plant-specific values of $\bar{s}_p$ and $\bar{G}$ and their respective coefficients of variation (CV) estimated from the yearly analysis, and the within-person standard deviation estimated from the longitudinal analysis by the two-stage method s_r and by the mixed-effects model s_m. Plants are sorted by $\bar{s}_p$

	Pairwise data				Longitudinal data	
Plant	$\bar{G}$	CV	$\bar{s}_p$ (l)	CV	s_r (l) (2-Stage)	s_m (l) (Mixed)
3	0.898	2.7	0.166	7.5	0.177	0.173
1	0.921	2.3	0.158	19.9	0.179	0.163
8	0.913	2.1	0.156	7.3	0.172	0.171
6	0.885	3.5	0.156	12.7	0.159	0.164
5	0.924	3.5	0.155	12.1	0.167	0.169
4	0.905	4.1	0.147	18.5	0.163	0.161
5	0.924	2.6	0.145	19.9	0.163	0.164
9	0.944	2.3	0.144	10.7	0.151	0.154
2	0.946	4.3	0.144	19.3	0.148	0.161
10	0.934	1.8	0.143	10.3	0.153	0.153
11	0.954	3.6	0.122	21.7	0.130	0.124

Correlations: $\bar{G}$ v $\bar{s}_p$ −0.762 (p<0.01); $\bar{G}$ v $s_{r(2\text{-stage})}$ −0.684 (p<0.05); $\bar{G}$ v $s_{m(mixed)}$ −0.662 (p<0.05); $\bar{s}_p$ v $s_{r(2\text{-stage})}$ 0.920 (p<0.0001); $\bar{s}_p$ v $s_{m\ (mixed)}$ 0.922 (p<0.0001).

DISCUSSION

Decision making based on imprecise longitudinal spirometry is likely to be ineffective and can be even counterproductive. A major task in longitudinal screening programmes and studies is to maintain a continued low level of within-person variation. This ensures that an individual's rate of change in lung function is estimated reliably. Due to the increased accuracy of commercially available spirometers, the random measurement error due to an instrument error (calibration procedures, malfunction, and so on) can be minimised, but other sources of the within-person variation still remain an issue.[15][22]

The results from our study show that by monitoring the precision of the longitudinal data using the within-person standard deviation s_p based on two repeated measurements or the G statistic, one can predict the magnitude of the within-person variation estimated from longitudinal data with five or more years of follow up. The plant-specific within-person variation $\bar{s}_p$ correlated more strongly with the longitudinal estimate s_r than the plant-specific coefficient $\bar{G}$, and thus it may be more suitable for monitoring of data precision especially in a smaller sample. However, the values of $\bar{s}_p$ were systematically lower than those of the longitudinal s_r or s_m, which may be because the shorter follow up does not include all potential errors that can occur during the longer follow up and also because of autocorrelation.[30]

The advantage of the coefficient G is its simplicity of estimation. Our data suggest that the value of coefficient G estimated from per cent predicted FEV1 values should be maintained above 0.90 at minimum, but ideally above 0.95. Although the coefficient G is easy to calculate, it has some inherent limitations. Because G is determined by the magnitude of the between-person variance as well as the within-person variance, in smaller samples, significant fluctuations in G may arise from fluctuation in the between-person variation, and it may be better to employ the within-person standard deviation s_p estimate. Based on our observations, the coefficient G based on per cent predicted values and a minimum sample size over 100 reflects changes in the within-person variability almost as well as the s_p statistic.

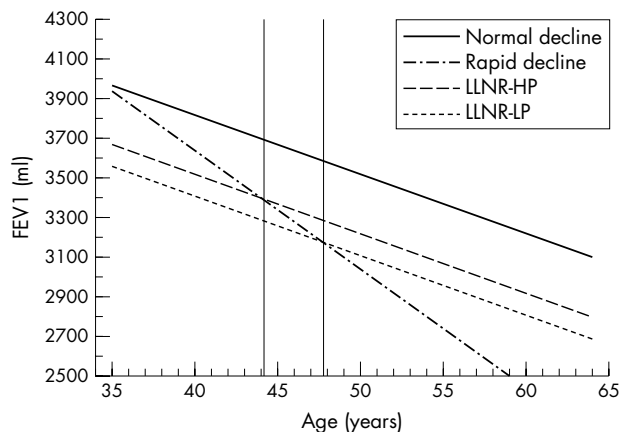The 11 plants that we investigated used standardised spirometry methods based on ATS recommendations. The range of plant-specific s_r based on slopes with at least five years of follow up was 0.130–0.179 l (see table 2, two-stage). These values are comparable to previously published values for large monitoring programmes (0.114–0.160 l).[13] Because the values of the within-person variation did not change substantially after we adjusted for age, height, symptoms of wheezing, and time variant smoking in the mixed model, we suspected that differences in measurement procedures, especially variability in technicians may have been the main sources of the within-person variation. Based on incomplete technician records, the testing in plant 11 was done by two technicians, whereas in plant 3 at least six technicians performed the testing. We also cannot exclude the possibility that occupational exposure increased the within-person variation in some plants.

Monitoring and maintaining data precision is important. Based on theoretical considerations, we show that the degree
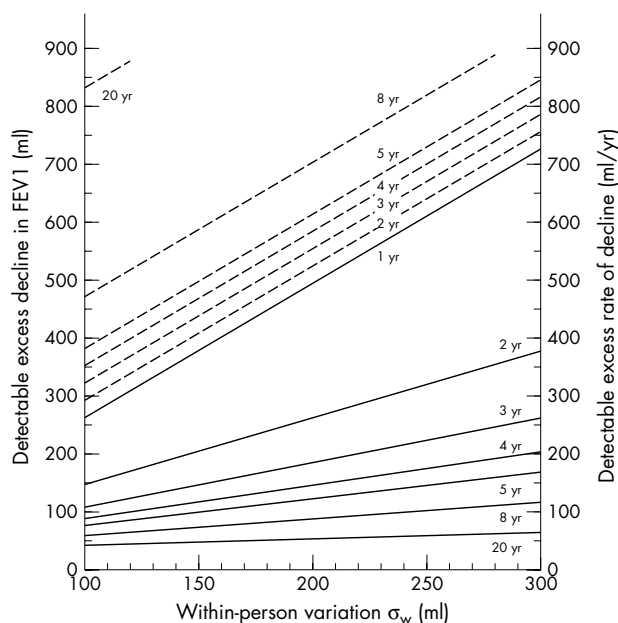

Figure 3 Predicted line for a ''normal'' decliner (30 ml/year) and a rapid decliner (60 ml/year), and approximate longitudinal lower 95% confidence band for the regression line (LLNR) for the ''normal'' decliner from PLANT-HP (LLNR-HP) and PLANT-LP (LLNR-LP). The vertical lines show the age at which the decline given by the slope of 60 ml/year would be considered excessive based on LLNR-HP and LLNR-HP.

Figure 4 The detectable excess decline in FEV1 (ml) on the left vertical axis (dashed line) and the detectable excess rate of decline (ml/year) on the right vertical axis (solid line), by within-person standard deviation σ_w, for increasing years of follow up, when the number of tests $P = 2$.

of precision in longitudinal FEV1 measurements affects the ability to detect abnormal decline in individuals. Figure 3 shows that the LLNR, based on the group average within-person variation, is higher for a person from PLANT-HP than for a person from PLANT-LP. The increased precision in longitudinal measurements affects the ages when the rapid decliner of 60 ml/year is crossing the LLNRs. For the LLNR-HP ($s_r = 0.130$ l), the intersection is at ≈ 44 years of age. For the LLNR-LP ($s_r = 0.177$ l), the intersection is at ≈ 48 years of age. Thus, the precision of the longitudinal data can affect the age at which we can identify a ''true'' rapid decliner. However, if we used the LLNR-HP for a decision-making in PLANT-LP, we could identify ''false'' rapid decliners because the random variation in FEV1 in PLANT-LP is higher than in PLANT-HP.

A recent study suggests that a yearly decline of 8% or 330 ml should not be considered normal in healthy working males tested according to ATS standards.[35] Similarly, in the Lung Health Study, 95% of the yearly differences in FEV1 were within 320 ml for men with early COPD.[36] Figure 4 shows that for the duration of follow up $D = 1$ and number of tests $P = 2$, in 95% of individuals the yearly decline in FEV1 would be within ≈ 330 ml for σ_w of 130 ml.

Figure 4 illustrates how the size of the within-person standard deviation σ_w can impact the detection of excess decline in FEV1 (dashed line) and excess rate of decline (solid line), for a given duration of follow up D and two repeated measurements. It takes longer in an imprecise monitoring programme to identify a ''true'' excessive rate of decline in FEV1. For example, it takes five years to identify an excess rate of decline of 90 ml/year when $\sigma_w = 130$ ml and eight years when $\sigma_w = 210$ ml. Conversely, after five years of follow up the detectable excess rate of decline increases with increasing value of σ_w as follows: for $\sigma_w = 100$ it is 75 ml/year, for $\sigma_w = 130$ it is 90 ml/year, for $\sigma_w = 150$ it is 100 ml/year, for $\sigma_w = 250$ it is 150 ml/year, and for $\sigma_w = 300$ it is 170 ml/year. The 5th percentiles for the observed yearly changes in FEV1 found for our plants 3 (-0.342) and plant 11 (-0.433) agreed with the estimated data for one year of follow up in figure 4.

For example, in figure 1 one may not consider the decline of 400 ml from the first to the second FEV1 observation for person B to be abnormal if the value of $\bar{s}_p$ for the monitoring programme is 200 ml. However, one should try to identify extraneous sources of within-person variation and decrease

s_p. If, on the other hand, $\bar{s}_p$ is ≈ 100 ml, then one should consider the decline of 400 ml excessive and take appropriate action.

These results show that a measure of data precision needs to be attached to the longitudinal data on which decision making is being made even if the subjects are tested according to the ATS recommendations. One can also increase precision of the estimated slopes by increasing the number of observations for an individual person whose measurements fall bellow the longitudinal lower limits of normal. An abnormal decline based on a predicted slope estimated over five or more years of follow up could then trigger more definite intervention measures.

In conclusion, the results demonstrate that there is a need for monitoring of data precision in spirometry monitoring programmes performed by technicians trained in ATS standards. For a little additional cost, the gain in the precision of the estimates on which decision making is made could be invaluable as it would allow identification of rapid decliners at an earlier age and prevent development of airflow obstruction.

. . . .

Authors' affiliations

E Hnizdo, M Attfield, Division of Respiratory Disease Studies, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Morgantown, WV, USA

L Yu, Constella Health Sciences, Morgantown, WV, USA

L Freyder, H W Glindmeyer, Section of Pulmonary, Critical Care, and Environmental Medicine, Department of Medicine, Tulane Medical School, New Orleans, LA, USA

J Lefante, Department of Biostatistics, Tulane School of Public Health and Tropical Medicine, New Orleans, LA, USA

REFERENCES

1 **Pauwels RA**, Buist SA, Calverley PMA, Jenkins CR, Hurd SS, on behalf of the GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2001;**163**:1256–76.
2 **Hnizdo E**, Sullivan PA, Bang KM, *et al*. Airflow obstruction attributable to work in industry and occupation among U.S. race/ethnic groups: A study of NHANES III data *Am J Ind Med* 2004;**46**:126–35.
3 **Celli BR**, MacNee W, and committee members. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004;**23**:932–46.
4 **Lebowitz MD**, Quackenboss J, Camilli AE, *et al*. The epidemiological importance of intraindividual changes in objective pulmonary responses. *Eur J Epidemiol* 1987;**3**:390–8.
5 **Rijcken B**, Schouten JP, Weiss ST, *et al*. ERS/ATS workshop on longitudinal analysis of pulmonary function data, Barcelona, September 1995. *Eur Respir J* 1997;**10**:758–63.
6 **Weiss ST**, Ware JH. Overview of issues in the longitudinal analysis of respiratory data. *Am J Respir Crit Care Med* 1996;**154**:S208–S211.
7 **Schouten JP**, Tager IB. Interpretation of longitudinal studies. An overview. *Am J Respir Crit Care Med* 1996;**154**:S278–S284.
8 **Ware JH**, Weiss S. Statistical issues in longitudinal research on respiratory health. *Am J Respir Crit Care Med* 1996;**154**:S212–S216.
9 **Sherrill D**, Viegi G. On modeling longitudinal pulmonary function data. *Am J Respir Crit Care Med* 1996;**154**:S217–S222.
10 **Rothman KJ**, Greenland S. *Modern Epidemiology*. Lippincott-Raven Publishers, 1998:115–34.
11 **American Thoracic Society**. Standardization of spirometry: 1994 update. *Am J Respir Crit Care Med* 1995;**152**:1107–36.
12 **Hnizdo E**, Churchyard G, Barnes D, *et al*. Assessment of reliability of lung function screening programs or longitudinal studies. *Am J Respir Crit Care Med* 1999;**160**:2006–11.
13 **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.
14 **Glindmeyer HW**, Jones RN, Diem JE, *et al*. Useful and extraneous variability in longitudinal assessment of lung function. *Chest* 1987:877–82.

15 **Becklake MR**, White N. Sources of variation in spirometric measurements. Identifying the signal and dealing with the noise. *Occupational Medicine: State of the Art Reviews* 1993;**8**:241–61.
16 **Gardner MJ**, Heady JA. Some effects of within-person variability in epidemiological studies. *J Chron Dis* 1973;**26**:781–95.
17 **Shepard DS**. Reliability of blood pressure measurements: Implications for designing and evaluating programs for control hypertension. *J Chron Dis* 1981;**34**:191–209.
18 **Glindmeyer HW**, Lissak T. A network approach to national respiratory health care. In: *Computerization and automation in health facilities*. Boca Raton, FL: CRC Press Inc, 1984:1–16.
19 **Nelson SB**, Gardner RM, Crapo RO, *et al*. Performance evaluation of contemporary spirometers. *Chest* 1990;**97**:288–97.
20 **Ferris BG**. Epidemiology standardization project. *Am Rev Respir Dis* 1978;**118**:7–53.
21 **Gardner RM**, Hankinson JL, Clausen JL, *et al*. ATS statement on standardization of spirometry - 1987 update. *Am Rev Respir Dis* 1987;**136**:1285–98.
22 **Glindmeyer HW**, Jones RN, Barkman HW, *et al*. Spirometry: quantitative test criteria and test acceptability. *Am Rev Respir Dis* 1987;**136**:449–52.
23 **Glindmeyer HW**, Lefante JJ, McCollister C, *et al*. Blue-collar normative spirometric values for Caucasian and African-American men and women age 18–65. *Am J Respir Crit Care Med* 1995;**151**:412–22.
24 **Singer JD**, Willett JB. *Applied longitudinal data analysis*. Oxford, UK: Oxford University Press, 2003:45–74.
25 **Verbeke G**, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer, 2000:19–29.
26 **Burrows B**, Lebowitz MD, Camilli AE, *et al*. Longitudinal changes in forced expiratory volume in one second in adults: Methodologic considerations and findings in healthy nonsmokers. *Am Rev Respir Dis* 1986;**133**:974–80.
27 **Sherrill DL**, Lebowitz MD, Knudson RJ, *et al*. Continuous longitudinal regression equations for pulmonary function measures. *Eur Respir J* 1992;**5**:452–62.
28 **Neter J**, Wasserman W. *Applied linear statistical models*. In: Richard D, ed. Homewood, IL: Irwin Inc, 1974.
29 **Hankinson J**, Wagner GR. Medical screening using periodic spirometry for detection of chronic lung disease. *Occup Med State Art Rev* 1993;**8**:353–61.
30 **Schlesselman JJ**. Planning a longitudinal study: II. Frequency of measurement and study duration. *J Chron Dis* 1973;**26**:553–8.
31 **Wang M**, Gunel E, Petsonk EL. Design strategies for longitudinal studies. Study duration and measurement frequencies. *Am J Respir Crit Care Med* 2000;**162**:2134–8.
32 **Lefante J**. The power to detect differences in average rates of change in longitudinal studies. *Stat in Med* 1990;**9**:437–46.
33 **American College of Occupational and Environmental Medicine**. Spirometry in the occupational setting. *J Occup Environ Med* 2000;**42**:228–45.
34 **Buist AS**, Vollmer WM. The use of lung function tests in identifying factors that affect lung function growth and aging. *Stat Med* 1988;**7**:11–18.
35 **Wang M**, Petsonk EL. Repeated measures of FEV1 over six to twelve months: what change is abnormal? *JOEM* 2004;**46**:591–5.
36 **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.