American Industrial Hygiene Association Journal

ISSN: 0002-8894 (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/aiha20

PERFORMANCE OF LABORATORIES MEASURING SILICA IN THE PROFICIENCY ANALYTICAL TESTING PROGRAM

Stanley A. Shulman , Jensen H. Groff & Martin T. Abell

To cite this article: Stanley A. Shulman , Jensen H. Groff & Martin T. Abell (1992) PERFORMANCE OF LABORATORIES MEASURING SILICA IN THE PROFICIENCY ANALYTICAL TESTING PROGRAM, American Industrial Hygiene Association Journal, 53:1, 49-56, DOI: 10.1080/15298669291359294

To link to this article: https://doi.org/10.1080/15298669291359294

AM. IND. HYG. ASSOC. J. 53(1):49–56 (1992)

PERFORMANCE OF LABORATORIES MEASURING SILICA IN THE PROFICIENCY ANALYTICAL TESTING PROGRAM*

Stanley A. Shulman
Jensen H. Groff
Martin T. Abell

National Institute for Occupational Safety and Health, Division of Physical Sciences and Engineering, 4676 Columbia Parkway, Cincinnati, OH 45226

A statistical study was performed on the results reported by laboratories analyzing silica samples in the first 101 rounds of the Proficiency Analytical Testing (PAT) program. Five laboratories participated in the first round of the PAT program in 1972, and participation grew to 130 laboratories before falling to 105 in Round 101. The laboratories use all three of the major methods of analysis: colorimetry, x-ray diffractometry, and infrared spectroscopy. The objectives of the study were to determine bias between methods, the variability associated with the methods, and any changes in bias or variability caused by a number of factors. The colorimetric method has consistently given the lowest results, particularly at higher loadings. X-ray diffractometry results were biased higher than infrared spectroscopy results during one period but not in the following period. Between the two periods, the procedures and materials used to prepare PAT samples changed in a number of ways, but the switch to quartz dust with a smaller particle size is a likely explanation for the bias difference. Generally, silica analyses have improved in precision over time, and this improvement has taken place for all three of the methods. The colorimetric method has shown the poorest precision of the three methods, but, unlike the differences in bias, the differences in precision have diminished considerably over time. Precision estimates from other studies were compared to those from this study to learn more about sources of variability. The largest source of variability, the differences between laboratories, was large even when laboratories used the same method, as they did in a collaborative study of silica methods.

Proficiency testing and laboratory accreditation are used to improve the accuracy of test results in various fields.(1) In 1972, researchers at the National Institute for Occupational Safety and Health (NIOSH) started the Proficiency Analytical Testing (PAT) program to evaluate and improve the agreement of analyses obtained from several laboratories performing work for the Occupational Safety and Health Administration (OSHA). In 1973, this quality assurance effort was expanded by allowing other laboratories to participate in the PAT program and by funding the development of criteria for a national laboratory accreditation program operated by the American Industrial Hygiene Association (AIHA).(2) Today, the PAT program is a cooperative effort of AIHA and NIOSH with over 1400 participating laboratories from the United States and other countries. PAT performance, along with a number of other requirements and a laboratory site visit, has been used by AIHA to accredit approximately 350 of these laboratories.

The PAT program was the only program that regularly evaluated laboratories on their ability to measure small quantities of silica dust on filters until the British WASP program(3) began in 1988. Over 100 laboratories currently analyze and report on PAT silica samples. Over the years, the silica samples sent out in the PAT program have changed and so have the methods used to analyze them. A description of the samples, and how they have changed over the years, is presented in the section entitled PAT Program Silica Samples and Analytical Methods.

The primary goal of this study was to find trends in the reported silica results for the first 101 rounds, particularly trends in the bias and variability of the results reported during the last 70 rounds. Biases between methods, as determined from the PAT results, are discussed in the Bias section. The variability of the methods, differences in variability between the methods, and changes in variability over different periods are discussed in the Variability section. Finally, the variability found in this study is compared to that found in other studies in the section entitled Comparison to Other Studies. Additional discussion of the data analysis procedures are in the Appendix.

PAT PROGRAM SILICA SAMPLES AND ANALYTICAL METHODS

The silica samples sent to PAT program laboratories each round consist of five polyvinyl chloride filters (FWSB, Mine Safety Appliances, Pittsburgh, Pa.). One of the filter samples is a blank and four filters contain varying amounts of silica dust (Min-U-Sil 5, Pennsylvania Glass Sand Corp., Murfreesboro, W.Va.) and one or more of the following contaminant dusts: magnesium silicate

---

*Mention of company and product names does not constitute endorsement by the National Institute for Occupational Safety and Health.

(Talc 399, Whittaker, Clark & Daniels, South Plainfield, N.J.), coal dust (Pittsburgh Seam 200 mesh, Bureau of Mines, Pittsburgh, Pa.), or calcite (Micro-White 25, Sylacauga Calcium Products, Sylacauga, Ala.). All samples in a round have the same matrix. The samples are collected from aerosols of the separate dusts in a chamber that can accommodate 320 filters per run. The PAT silica samples have changed over the years, as shown in Table I. Changes in the sample producers, matrices, loadings, filter type, particle size of the silica, and the method of generation are all important considerations when analyzing PAT results. The loading levels have decreased over time. Before Round 13, silica loadings for samples in a given round were frequently above 200 µg. After Round 20, this is rarely the case, as can be seen in Figure 1.

TABLE I. Production of PAT Silica Samples

Period	*Round*	*Date (Month/ Year)*	*Silica*	*Matrix*	*Filter*	*Method*	*Producer*
1	1–9	05/72–08/72	Min-U-Sil	$CaCO_3$	FWSB	air	NIOSH[A]
2	10–15	10/72–01/73	Min-U-Sil 5	none	FWSD	liquid	NIOSH
3	16–18	02/73–04/73	Min-U-Sil 15	none	FWSD	liquid	NIOSH
4	19–29	05/73–06/74	Min-U-Sil 15	2 mg $Na_2SiO_39H_2O$	FWSD	liquid	NIOSH
5	30–45	10/74–11/77	Min-U-Sil 15	2 mg $Na_2SiO_39H_2O$	FWSD	liquid	Contractor 1
6	46–46	01/78–01/78	Min-U-Sil 10	0.5 mg $Na_2SiO_39H_2O$	DM5000	liquid	Contractor 2
7	47–62	03/78–09/80	Min-U-Sil 10	0.5 mg $Na_2SiO_39H_2O$	FWSD	liquid	Contractor 2
8	63–101	11/80–04/90	Min-U-Sil 5	$MgSiO_4$(Talc)[B]	FWSB	air	Contractor 2

[A] Cincinnati, Ohio, location. All others Morgantown, W.Va.

[B] Since Round 77, the matrix has alternated between talc, calcite, and coal dust, or combinations thereof.

The analytical methods used to quantitate samples of silica on filters may differ in numerous details but are usually categorized and referred to by the instrumental technique used for the final measurement. There is a space on the PAT reporting form for recording the method. The three most common are visible spectrometry or colorimetry (COL), x-ray diffractometry (XRD), and infrared (IR) spectrometry.(4) Some laboratories used atomic absorption spectroscopy or another measurement method, but they were so few that they were ignored in the data analysis. The number of laboratories that used each of the methods for PAT program silica samples is shown in Figure 2. From Round 1 to Round 91, the number of laboratories analyzing silica samples grew from 5 to over 130, then dropped to 105 by Round 101. Initially, most laboratories used COL, but after Round 25, the number that used COL stabilized at about 30, then dropped to 16 by Round 101. Meanwhile, the number of laboratories that use IR and XRD have grown to 36 and 53, respectively. Although there is evidence that laboratories have switched methods, overall, the number of transitions from method to method was not very high.

BIAS

An objective of this study of PAT data was to determine if biases exist among methods. For example, the results obtained by XRD may have been different from those obtained by IR. There were a variety of factors that may have caused bias, and it is possible that biases among the methods varied over the levels of these other factors. Some of these factors are indicated by studying Table I. The table lists eight periods, and factors affecting bias between methods may change between periods or within periods. Factors that change between periods are obvious from the column headings, such as silica type, sample matrix, or sample producer. Factors that change within the periods are less obvious and are discussed below.

Biases among methods may have varied because of the differences between the specific laboratories that used the method, the differences between rounds (including the differences in average loading between rounds), or the differences in sample matrices in different rounds. The first two factors, laboratories and rounds, can be studied separately within each period

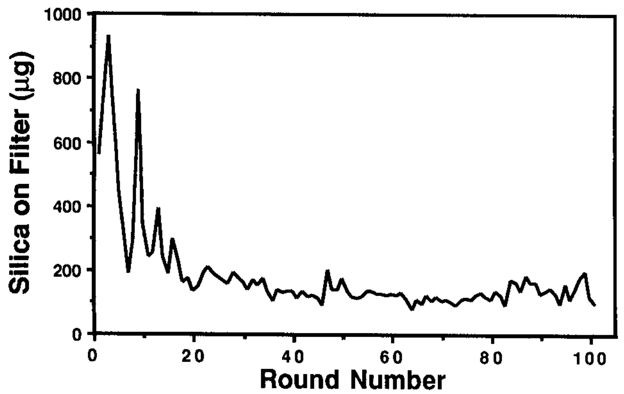

FIGURE 1. Median result reported by PAT laboratories for the most heavily loaded filter of the four sent each round

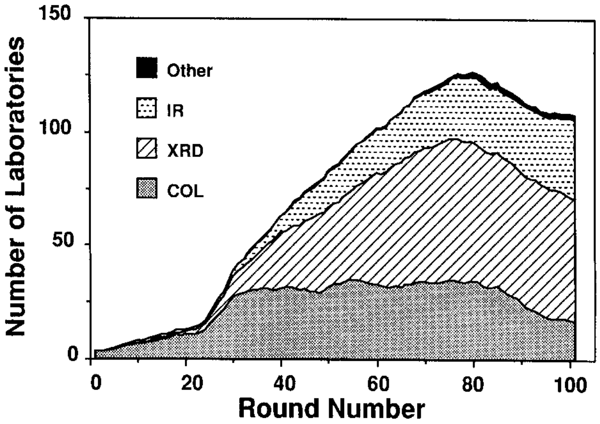

FIGURE 2. Number of laboratories that used a particular method of silica analysis in the PAT program

in Table I. For the third factor, matrix, only the data from Period 8, in which four different matrices were used and no other factors were changed, are relevant.

The bias of each silica method relative to each of the other methods was calculated for PAT results within Periods 5, 7, and 8. Earlier periods were not used because only after Round 29 are there enough laboratories that use methods other than COL that bias estimation becomes reasonable. Of the four periods that included Round 30 and later, the study was limited to Periods 5,7, and 8 because Period 6 consisted of just one round. Examination of the data indicates that method biases vary considerably with loading, and the analyses described below are aimed toward estimation of the bias-loading interaction.

The procedure for estimating bias between methods was chosen from several alternatives and consisted of the following five steps. First, for each of the three methods, a median of the results reported for each of the four samples in a given round was computed. Second, for each sample in each round, the middle median (the median of the three method medians) was selected. Third, for each method within each round, the four method medians were regressed on the four middle medians, thereby obtaining a straight line relationship between the method median and the middle medians for each round and each method. Fourth, these estimated straight lines were used to get predicted values for each method in each round at four loadings: 60, 90, 120, and 150 µg. Fifth, these predicted values were then used as new data for a model in which differences among the methods could be estimated.

The median was used rather than the mean because the median is less sensitive to outliers. More complicated models are described in the Appendix. In general, the results of the fuller models are similar to the results presented in this paper.

In large samples from normal populations, the variance of the median is approximately the variance of a single measurement divided by the number of measurements. It was assumed that between-method differences in the variances of a single measurement were not large compared to differences in the numbers of laboratories that used the methods in any given round. Thus, weighted least squares was used in the analyses that follow where each middle median was weighted by the number of laboratories that used the method in the particular round.

The results of the estimates based on the additive models are given in Table II. Note that the statistically significant differences presented in the table are significant at the 2.5% level. For each period or period-matrix combination, the overall significance level for all four loadings is 10%. That is, it was considered acceptable to indicate a bias where none existed 10% of the time. Significance at the 10% level is a reasonable choice because the PAT data were not specifically collected for this type of study.

The most consistent trend shown in the table is that the COL measurements were lower than the measurements by the other methods. This was true for all three periods. Note, however, that for Period 8 samples with talc matrix, COL yielded slightly higher results than XRD at the 60-µg loading. Because the different periods made use of different particle sizes and different matrices for the silica, the results suggest that COL gave somewhat lower measurements regardless of particle size or matrix. There is an indication that the negative bias increased at higher loadings.

Between-period comparisons are also possible. Because Table II is arranged to control for Type I error within periods, further statistical analyses were performed to ensure control of the Type I error for the between-period comparisons. For these analyses, the predicted values for the four loadings given in Table II for each round in each period were used. An exception was made for Period 8, for which only the talc data were used. The talc rounds constituted most of the rounds in Period 8; allowing for matrix differences in Period 8 would have added considerable complexity. The predicted values in each period by each method were regarded as a random sample, because round-to-round differences within a period were assumed to be insignificant after the rounds were adjusted to the same loading. From each analysis,

TABLE II. Bias of Silica Medians by Method, within Periods, at Several Loadings

Period (Rounds)	*Loading (µg)*	*Methods Differing at 2.5% Level*	*Confidence Intervals for Biases (µg)*
5 (30–45)	60	none	
	90	COL < IR	COL – IR = (–27.8, –0.05)
	120	COL < XRD	COL – XRD = (–22.0, –3.56)
	150	COL < XRD	COL – XRD = (–29.9, –8.44)
7 (47–62)	60	none	
	90	COL < XRD	COL – XRD = (–14.8, –4.68)
		IR < XRD	IR – XRD = (–19.0, –6.65)
	120	COL < XRD	COL – XRD = (–24.3, –11.2)
		IR < XRD	IR – XRD = (–26.7, –10.7)
	150	COL < XRD	COL – XRD = (–35.7, –15.8)
		IR < XRD	IR – XRD = (–36.7, –12.4)
8 Talc[A]	60	COL > XRD	COL – XRD = (2.38, 8.25)
	90	COL < IR	COL – IR = (–9.31, –2.30)
		IR > XRD	IR – XRD = (0.33, 6.53)
	120	COL < IR	COL – IR = (–19.8, –9.06)
		COL < XRD	COL – XRD = (–14.5, –5.62)
	150	COL < IR	COL – IR = (–30.9, –15.2)
		COL < XRD	COL – XRD = (–24.3, –11.3)
8 Calcite (86, 89, 93, 97, 101)	60	IR > XRD	IR – XRD = (0.54, 19.0)
	90	IR > XRD	IR – XRD = (1.64, 21.1)
	120	COL < IR	COL – IR = (–32.6, –3.85)
		IR > XRD	IR – XRD = (1.61, 24.4)
	150	COL < IR	COL – IR = (–42.6, –7.92)
		IR > XRD	IR – XRD = (0.87, 28.3)
8 Talc and coal (81, 85, 95, 99)	60	none	
	90	none	
	120	none	
	150	none	
8 Coal (77, 84, 88, 92, 96, 100)	60	none	
	90	COL < XRD	COL – XRD = (–14.8, –0.35)
	120	COL < IR	COL – IR = (–27.3, –4.32)
		COL < XRD	COL – XRD = (–23.6, –3.65)
	150	COL < IR	COL – IR = (–41.1, –7.07)
		COL < XRD	COL – XRD = (–36.7, –7.30)

[A] Period 8 rounds that had talc matrix: 63–76, 78–80, 82, 83, 87, 90, 91, 94, 98.

based on a linear model including a factor for time period, method, and their interaction, statistically significant interaction at the 1% level was obtained. The main reason for this interaction appears to be that IR values are lower than XRD values in Period 7 by about 15 μg, on average. The corresponding differences in Periods 5 and 8 are much smaller.

It may be that the differences cited above are attributable to particle size effects. Period 7 data are based on Min-U-Sil 10; Period 8 data are based on the smaller Min-U-Sil 5. On the basis of the literature, the expected effect of decreased particle size on each silica method is as follows: (1) infrared absorption should increase(5); (2) XRD intensity should decrease(6,7); and (3) COL results should decrease because of dissolution of silica in the phosphoric acid wash.(8) This corresponds with the findings. Table II indicates that it never happens in Period 8, regardless of matrix, and that the XRD measurements are positively biased relative to IR (whereas XRD is positively biased relative to IR for the higher loadings in Period 7). Also, whereas COL is negatively biased relative to IR in many instances in Period 8, COL is never negatively biased relative to IR in Period 7.

However, the Min-U-Sil 15 used in Period 5 has a larger particle size than the silicas in either of the two other periods and differences between XRD and IR are not statistically significant. Possible problems with the Period 5 data are that there are fewer than five laboratories that use IR before Round 42, and the silica samples for that period may have had higher variability than those for later periods. The high variability associated with the Period 5 data is discussed in the Variability section.

As stated above, alternative analyses were carried out for the bias data, and these included an analysis of the individual measurements. Normal probability plots of each sample's determinations from Round 92 to Round 101 indicate that the measurements are normally distributed on the original scale. Thus, for the routine analysis of PAT silica results, which includes the detection of excessive bias, it is not necessary to transform the data from the original scale as has been done in the past.

VARIABILITY

Evaluating the variability of silica results was the other objective of this work. Total variability and within-laboratory variability were investigated for each method. These are discussed in subsections below.

Variability is discussed in terms of relative standard deviation, also known as coefficient of variation. The estimated relative standard deviation is symbolized by s_r and the true relative standard deviation by s_r(true). Thus, $s_{r,t}$ is an estimate of the total standard deviation and $s_{r,t}^2 = s_{r,b}^2 + s_{r,w}^2$ where $s_{r,b}$ and $s_{r,w}$ are the estimates of between-laboratory and within-laboratory relative standard deviation, respectively. Analogous relations apply to the true values of relative standard deviation. As in the bias study, the factors that influence total and within-laboratory s_r can be classified as within period or between period. The discussion of these factors in the first two paragraphs of the Bias section is applicable.

Total Variability

The total s_r ($s_{r,t}$) value for each sample in each round was calculated by dividing the standard deviation of the results for that sample by their mean. The $s_{r,t}$ values were used to make within-period comparisons based on selected samples. The procedure for selecting the samples was as follows. First, for each of the three methods, a median of the results reported for each of the four samples in a given round was computed. Second, for each sample in each round, the middle median (the median of the three method medians) was selected. Finally, for each round in Periods 5, 7, and 8, the sample with the smallest middle median was chosen, and the three method $s_{r,t}$ values for the chosen sample were used in the comparison of methods within the period. A similar comparison was made by choosing from each round the sample that had the largest median.

These comparisons were much simpler than the analogous comparisons of medians. There was little evidence for interaction between method and loading in their combined effect on $s_{r,t}$(true) within the minimum and maximum analyses. Thus, the difference in $s_{r,t}$(true) values from different methods was almost constant across rounds within each period. The analysis was done on the natural log scale, on which the residuals were more nearly normally distributed and there was less evidence of interaction than on the original scale. Weighted least squares was used, in which the observations were weighted by the degrees of freedom. The degrees of freedom is the number of laboratories that made measurements on the particular sample in the specific round minus one.

The results for the analyses for the three periods, also carried out by maximum and minimum medians in each round, are given in Table III. Statistically significant differences were set at the 5% significance level. Because two loadings were chosen for Periods 5, 7, and 8, overall significance level, by period, was 10%. That is, it was considered acceptable to indicate a difference in variability where none existed 10% of the time. Note that COL estimates for Periods 1–4 are also included in Table III. There are not enough laboratories that use the two other methods to make within-period comparisons. However, the COL data from these periods are compared to each other and are compared to later periods in the discussion of between-period comparisons below.

The within-period comparisons of Periods 5, 7, and 8 indicate that COL $s_{r,t}$(true) values tend to be higher than those for the other methods. For the minimum loadings in Period 5, for instance, the COL $s_{r,t}$(true) values may be twice as high as for the other methods. For the maximum, the ratio of COL to XRD is probably smaller than 2. Because the COL measurements are biased low relative to the two other methods, the COL $s_{r,t}$ values have smaller denominators and might be larger. Note that in Period 5 the $s_{r,t}$ values at the higher loadings are considerably lower than those at lower loadings. For the most recent data (Period 8) the loading effect is relatively small. In general, the Period 8 COL $s_{r,t}$ values are not much more than one-third larger than the IR and XRD values. Also, Period 8 analyses indicate that method differences are consistent across silica matrices.

Two pairs of between-period models were constructed. The first pair was used primarily for a comparison of the early period COL data. The first model of the first pair included all the COL $s_{r,t}$ values from all periods for the minimum median sample in each round. The other model in the first pair was developed for the maximum median samples in each round. The model that used the minimum medians yields no statistically significant differences for tests at the 10% level in pairwise comparisons among the first four periods. Thus, the increase from 0.483 to 0.771 from Periods 2 to 4 at minimum loadings is not statistically significant. Because the number of laboratories reporting results was relatively low, it may be that the comparison lacks power. During Period 2, the COL method was used by 6 to 9 laboratories, but in Period 4, the number grew from 9 to 22. This large increase in the number of laboratories that used the method in the PAT program may have been responsible for a change in $s_{r,t}$ values.

The COL $s_{r,t}$ values for Period 5 differed little from those for Period 4, even though the number of laboratories that used the COL method continued to increase to as high as 35 laboratories. However, Period 5 samples came from a different producer than those for Periods 2 to 4; those for Periods 7 to 8 had still a different producer. It could be that differences between Periods 5, 7, and 8, which are discussed below, had to do with producer differences.

The second pair of models was primarily used for comparisons among methods in Periods 5, 7, and 8. These models were again constructed separately for $s_{r,t}$ values from minimum and maximum loadings in each round but for all three methods in Periods 5, 7, and 8. The data (Table III) suggest improvements in that there are fewer differences in later rounds among the methods, and the between-period analyses support this viewpoint. For the minimum median data, the ratio of COL $s_{r,t}$ to XRD $s_{r,t}$ in Period 5 ($0.875 \div 0.456 = 1.92$) is approximately 50% higher than the corresponding ratio in Periods 7 (1.26) or 8 (1.22); these differences are significant at the 5% level, according to the model. In other words, at low loadings, the $s_{r,t}$ values of COL relative to XRD are much closer in Periods 7 and 8 than in Period 5. At the maximum loading, only the corresponding ratios for Period 5 relative to Period 8 are statistically significant at the 5% level. The estimate of this ratio of ratios is 1.40. Thus, overall, with a Type I error of 10%, COL improved relative to XRD. The IR $s_{r,t}$ values fell in between the two other methods and cannot be distinguished statistically from them.

TABLE III. Total Variability of Silica Methods in Different Periods

Period (Rounds)	Loading	Overall $s_{r,t}^A$	Method	Method $s_{r,t}$	$s_{r,t}$ (True) Comparison	95% Confidence Limits on Ratio
1 (1–9)	min	0.683	COL	0.683		
	max	0.570	COL	0.570		
2 (10–15)	min	0.483	COL	0.483		
	max	0.325	COL	0.392		
3 (16–18)	min	0.687	COL	0.687		
	max	0.369	COL	0.532		
4 (19–29)	min	0.771	COL	0.771		
	max	0.538	COL	0.538		
5 (30–45)	min	0.660	COL	0.875	COL > IR	COL/IR = (1.099, 2.054)
			IR	0.578	COL > XRD	COL/XRD = (1.538, 2.292)
			XRD	0.456		
	max	0.454	COL	0.577	COL > IR	COL/IR = (1.166, 2.097)
			IR	0.403	COL > XRD	COL/XRD = (1.369, 1.962)
			XRD	0.348		
6 (46)	min	0.666	COL	0.575		
			IR	0.791		
			XRD	0.611		
	max	0.599	COL	0.412		
			IR	0.734		
			XRD	0.605		
7 (47–62)	min	0.388	COL	0.451	COL > IR	COL/IR = (1.055, 1.527)
			IR	0.347	COL > XRD	COL/XRD = (1.091, 1.457)
			XRD	0.358		
	max	0.346	COL	0.384	COL > XRD	COL/XRD = (1.036, 1.444)
			IR	0.344		
			XRD	0.307		
8 (63–101) all matrices	min	0.319	COL	0.359	COL > IR	COL/IR = (1.093, 1.437)
			IR	0.300	COL > XRD	COL/XRD = (1.150, 1.449)
			XRD	0.292		
	max	0.276	COL	0.303	COL > XRD	COL/XRD = (1.058, 1.323)
			IR	0.266		
			XRD	0.255		

[A] Overall $s_{r,t} = \sqrt{s^2_{r,t,COL} + s^2_{r,t,IR} + s^2_{r,t,XRD}}$ for Periods 5–8.

Within-Laboratory Variability

Within-laboratory s_r ($s_{r,w}$) values were obtained as follows. Because each PAT round consisted of four different samples plus a blank, there were no results based on exact duplicates with which to estimate the within-laboratory variability. However, in many of the rounds there were two samples with silica loadings similar enough to permit a calculation of within-laboratory variability. The assumption was that the relative standard deviation of silica analytical results was the same for any two results larger than 40 µg and within 20 µg of each other.

When the results for two samples in a PAT round met these conditions, they were used as a data set for this study. A discussion of the rationale for this matching scheme is found in the Appendix. An analysis of variance was carried out on the log scale, adjusting for each sample mean and for each laboratory mean in the round. The square root of the residual mean square was used as an estimate of $s_{r,w}$ because, for data that are approximately normal on the log scale, the standard deviation of those data approximates the relative standard deviation on the original scale. Because of the matching scheme, trends in the $s_{r,w}$ values may be of more value than the absolute values of the estimates.

A comparison of $s_{r,w}$ values was carried out for the data from Periods 5, 7, and 8. The results are presented in Table IV. Statistically significant differences are cited at the 10% significance level. The matching scheme described above was used in this case. As for $s_{r,t}$, the analysis was carried out on the log scale. Six pairs were chosen for the Period 5 comparison, 13 sample pairs for the Period 7 comparison, and 26 sample pairs for the Period 8 comparison. The $s_{r,w}$ decreased somewhat with loading, but this did not hamper the comparison of methods because there appeared to be little evidence of interaction between method differences and loading. The analysis of the data for each period indicated that within each period, the COL $s_{r,w}$ values were (statistically) significantly higher than either or both of the $s_{r,w}$ values for the other methods. Tests were carried out at the 10% significance level.

A between-period model was also constructed for the $s_{r,w}$ values, analogous to that for $s_{r,t}$, but both loadings were combined. The analysis indicates that at the 10% significance level, Period 8 differs from Periods 5 and 7 in the following way. The ratio of COL to the average of IR and XRD in Periods 5 and 7 is 1.64; in Period 8 the corresponding ratio is 1.13. The ratio of the two ratios is 1.45, which is significantly different from a ratio of 1 at the 10% significance level. The conclusion, then, is that the $s_{r,w}$ values agree better in Period 8 than in either Period 5 or Period 7. This differs from the comparison of $s_{r,t}$ values, for which Period 7 was more like Period 8 than Period 5. What matters most is that the trend indicates improved $s_{r,w}$(true) values in more recent rounds.

COMPARISON TO OTHER STUDIES

Table V summarizes variability estimates available to the authors from other sources. They are (1) analytical precision for the three methods, (2) precision data from the quality control checks performed by the sample producer on each sample set provided in the various rounds, and (3) $s_{r,w}$ and $s_{r,t}$ estimates from a collaborative study that involved both the IR and XRD methods. In addition, $s_{r,t}$ values for PAT data in Period 8, minimum loading, and $s_{r,w}$ data for Period 8, as determined in this study, are given in Table V. Note that the s_r estimates for the collaborative study IR results are given as a range in Table V because s_r increased when kaolinite or large amounts of coal dust were added to the samples. The XRD results were less affected by these interferences.

The samples for the collaborative test were generated in a slightly modified chamber of the same generator used to make PAT program samples. Thus, the sample-to-sample variability for the collaborative test should be similar to that for the PAT program samples. That variability is also fairly small, approximately 0.06, and is based on the PAT quality control $s_{r,w}$ with the analytical method $s_{r,w}$ component removed. The $s_{r,w}$ values found in the collaborative test for IR and XRD are in the range of $s_{r,w}$ for the PAT QC results.

For both the present study and the collaborative test, there was substantial between-laboratory variability. Recall that $s_{r,b}^2 = s_{r,t}^2 - s_{r,w}^2$. In other words, in both of the studies, there was substantial variability between laboratories compared to that within laboratories. In the case of the collaborative test, all laboratories were required to use the same written method, so it appears that there are still procedures or materials not completely specified in the method.

The estimates of s_r found for the PAT program in this study for recent rounds, both within laboratory and total, are higher than those obtained in the collaborative test. That is not surprising because PAT participation does not require that a specific method be used. The only thing that matters is that laboratories not produce results too far away from the mean of the reference laboratories.

TABLE IV. Within-Laboratory Variability of Silica Methods In Different Periods

Period (Rounds)[A]	*Overall* $s_{r,w}$	*Method*	*Method* $s_{r,w}$	*Method* $s_{r,w}$ *Comparison*	*90% Confidence Limits*
5 (30–45)	0.315	COL	0.407	COL > XRD	COL/XRD = (1.14, 2.39)
		IR	0.244	COL > IR	COL/IR = (1.00, 2.79)
		XRD	0.247		
7 (47–62)	0.271	COL	0.360	COL > IR	COL/IR = (1.43, 2.48)
		IR	0.191	IR < XRD	IR/XRD = (0.58, 0.99)
		XRD	0.251	COL > XRD	COL/XRD = (1.18, 1.78)
8 (63–101)	0.183	COL	0.198	COL > IR	COL/IR = (1.03, 1.40)
		IR	0.165		
		XRD	0.186		

[A] Selected rounds from each period, as indicated in the text.

TABLE V. Variability Estimates from Various Sources

	Analytical Method	*PAT QC*(9) *(Average)*	*PAT Labs*		*Collaborative Test*(10) *(Median)*	
			(Period 8)	*(Period 8, Min)*		
Method	$s_{r,w}$	$s_{r,w}$	$s_{r,w}$	$s_{r,t}$	$s_{r,w}$	$s_{r,t}$
COL	0.09(4)	0.10	0.198	0.359		
IR	0.05(11)	0.101	0.165	0.300	0.07–0.10	0.13–0.22
XRD	0.08(4)	0.101	0.186	0.292	0.08	0.17

CONCLUSIONS

Examination of the PAT data has indicated that with regard to bias, the COL method has tended to give somewhat lower results than the two other methods. This is true over a long period of time and at a variety of loadings. This difference becomes bigger at higher loadings where the COL method gives results more than 20 µg lower than the two other methods. Some possible biases appear to be associated with different particle sizes and matrices. In Period 7 (Rounds 47 to 62), the silica material used was Min-U-Sil 10, and the IR measurements were lower than the XRD measurements by 7–25 µg. The difference increased with increasing loading. For Period 8 (Rounds 63 to 101), Min-U-Sil 5 (with a smaller particle size) was used. The XRD data were never consistently statistically higher than the IR data and were lower for the calcite matrix by approximately 10 µg. Attributing these differences to particle size assumes that other differences between Periods 7 and 8 data, such as filter type, do not matter.

The study of total s_r ($s_{r,t}$) values indicates that these $s_{r,t}$ values have been declining over time and are less loading-dependent than they once were. For the time period from Round 63 to 101, the difference in $s_{r,t}$ values betweeen low and high loadings was less than 4% and the average was about 0.27. There still appear to be differences among the methods, in that COL $s_{r,t}$ values are still statistically higher than XRD $s_{r,t}$ values by approximately 15%. However, this is a much smaller difference than in Period 5 and is evidence of improvement in COL relative to the other methods. There is no evidence that differences among method $s_{r,t}$ values vary across silica matrices. The within-laboratory s_r ($s_{r,w}$) values from the last time period indicate that the COL $s_{r,w}$ values are about 20% higher than the IR. However, because the average $s_{r,w}$ value is about 0.18, this average difference amounts to only about 0.03. As for $s_{r,t}$ values, the improvement of COL relative to the other methods is statistically significant.

Finally, examination of the most recent rounds indicates that it is no longer necessary to transform the silica data for purposes of routine PAT analyses.

ACKNOWLEDGMENT

The authors thank the following NIOSH reviewers whose comments helped considerably in the writing of this article: R. Hornung, K. Mulligan, M. Keane, P. Bierbaum, and L. Doemeny. R. Hornung's comments were particularly useful.

REFERENCES

1. **International Laboratory Accreditation Conference:** *International Directory of Laboratory Accreditation Systems and Other Schemes for Assessment of Testing Laboratories.* London: IMS-WORLD Publications Ltd., 1981.
2. **American Industrial Hygiene Association:** *Accreditation, A Plan for Accreditation of Industrial Hygiene Laboratories.* Akron, Ohio: American Industrial Hygiene Association, 1984.
3. **Health and Safety Executive:** *WASP: Workplace Analysis Scheme for Proficiency, Information for Participants.* London: Health and Safety Executive, 1988.
4. **National Institute for Occupational Safety and Health:** *NIOSH Manual of Analytical Methods,* edited by P. Eller (DHHS/NIOSH Publication No. 84-111). Cincinnati, Ohio: National Institute for Occupational Safety and Health, 1984. pp. 7500, 7601, 7602.
5. **Foster, R.D. and R.F. Walker:** Quantitative Determination of Crystalline Silica in Respirable-Size Dust Samples by Infrared Spectrophotometry. *Analyst 109*:1117–1127 (1984).
6. **Gordon, R.L. and G.W. Harris:** Effect of Particle-Size on the Quantitative Determination of Quartz by X-Ray Diffraction. *Nature 175(4469)*:1135 (1955).
7. **Edmonds, J.W., W.W. Henslee, and R.E. Guerra:** Particle Size Effects in the Determination of Respirable α-Quartz by X-Ray Diffraction. *Anal. Chem. 49(14)*:2196–2203 (1977).
8. **Jephcott, C.M. and H.F.V. Wall:** Determination of Quartz of Various Particle Sizes in Quartz-Silicate Mixtures. *Arch. Ind. Health 11*:425 (1955).
9. **Parish, H. and J. Teichman:** *Quarterly Report, Proficiency Analytical Testing Program.* NIOSH Contract 200-83-2063. 1980–1988. [Unpublished].
10. **National Institute for Occupational Safety and Health:** *Collaborative Tests of Two Methods for Determining Free Silica in Airborne Dust,* by C.C. Anderson. DHHS/NIOSH Publication No. 83-124. Cincinnati, Ohio: National Institute for Occupational Safety and Health, 1983.
11. **Freedman, R.W., S.Z. Toma, and H.W. Lang:** On-Filter Analysis of Quartz in Respirable Coal Dust by Infrared Absorption and X-Ray Diffraction. *Am. Ind. Hyg. Assoc. J. 35(7)*:411–418 (1974).

APPENDIX

Other Bias Models

The results from the bias model that are presented in the Bias section of the main text make use of a relatively simple model. Method medians were used in that model, and each method's medians for each round were adjusted to the same loadings by means of linear regression. An alternative approach is to use the individual measurements, rather than medians. To make the loadings under consideration more homogeneous, only one sample was used from each round in any given period. The sample chosen was always the sample with the lowest median loading. Models fitted to this data for Periods 5, 7, and 8 included the factors mentioned at the beginning of the Bias section. Thus, factors included in the resulting analysis of variance were round, method, laboratory (within method), and interactions involving these factors. In Period 8, separate models were fitted for each of the matrices. Having fitted all of these models, means of the predicted values were computed by round and method. Method biases for different loadings were then compared with the biases given in Table II of the main text. In general, the estimated biases from this fuller model agreed well with those presented in Table II of the main text. Analogous models were developed that used the sample from each round that had the maximum median of all samples within the round. As for the models that used the minimum medians, the results from these models agree well with those from the corresponding loadings shown in Table II of the main text.

Matching Scheme Used for Estimation of Within-Laboratory Variability

The scheme given in the Variability section of the main text for pairing results may appear somewhat arbitrary. The sole use is the estimation of the within-laboratory variability, $s_{r,w}$, because no such pairing is required in order to estimate the total variability—combined within- and between-laboratory variability. The sample standard deviation computed from the reported values for one sample in a round is an estimate of the total standard deviation of the PAT laboratories for that sample.

All three silica methods use calibration curves; determinations from the same part of a single curve would be expected to be similarly biased. If the loadings on two samples in a single round are sufficiently close, then the residuals remaining after adjustment for sample mean and laboratory mean are less likely to be inflated by interaction. Thus, relatively little interaction is present to inflate the residual.

The problem is, what constitutes "approximately the same loading"? In order to quantify the effect of loading, the following study was carried out. For each of the Rounds 68 to 101, for XRD, six s_r's were computed, one for each of the six sample pairs: (1,2), (1,3), (1,4), (2,3), (2,4), (3,4). These $s_{r,w}$ values were estimated by the square root of the residual mean square from an additive model (in sample number and laboratory) fitted for the specified sample pairs. This additive model was fitted on the natural log scale, in order that the square root of the mean square residual should be approximately equal to the $s_{r,w}$ on the original scale. Similar analyses were carried out for IR and COL, although not for all rounds from 68 to 101. For IR, 19 rounds of data were studied, and for COL, 25. The results are in Table VI. In the table, the column labeled $s_{r\Delta 60}$–$s_{r\Delta 10}$ is the estimated amount by which $s_{r,w}$(true) values from sample pairs whose medians *differ* by 60 µg exceed $s_{r,w}$(true) values of sample pairs whose medians differ by 10 µg. The column labeled s_{r60}–s_{r120} is the estimated amount by which $s_{r,w}$(true) values from sample pairs whose medians *are* 60 µg exceed $s_{r,w}$(true) values of sample pairs whose medians are 120 µg.

The differences cited in this table are associated with inclusion of statistically significant parameters in the models. In practical terms, the effects of differences among medians and loadings are important. The smallest differences are for XRD, for which the $s_{r,w}$(true) values for sample pairs whose medians differ by 60 µg are estimated to be 15% higher than for those differing by 10 µg. Likewise, the $s_{r,w}$(true) values for sample pairs whose average medians (for the paired samples) are 60 µg are estimated to be 15% higher than for pairs with average medians of 120 µg.

How, then, should samples within a given round be paired, so that the estimate of the within-round $s_{r,w}$ values is a good estimate? As was suggested above, both statistically and from the nature of calibration curve determinations, it makes sense to think that when the sample loadings of the samples within the matched pair are close together, there may be less interaction than when the loadings are far apart. The statistical analysis seems to bear this out. If this conclusion seems reasonable, then it makes sense to use $s_{r,w}$ estimates based on samples that are close together—because there are no perfect matches. The choice of a 20-µg limit for pairing samples was made because enough samples were available to meet this criterion. Requiring that the medians be closer together than this would reduce the number of samples available for pairing. Because exact duplicates are not available, the $s_{r,w}$ estimates obtained from this pairing scheme may tend to be overestimates of the true values. Also, most sample pairs obtained by this 20-µg criterion had median sample loadings of at least 40 µg.

TABLE VI. Effect of Loading on Within-Laboratory Variability

Method	*Average* $s_{r,w}$	$s_{r\Delta 60}$–$s_{r\Delta 10}$[A]	s_{r60}–s_{r120}[A]
COL	0.23	0.0591	0.0950
IR	0.21	0.0566	0.0521
XRD	0.20	0.0305	0.0288

[A] See text.