Regulatory Toxicology and Pharmacology 81 (2016) 268–274

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

How accurately and consistently do laboratories measure workplace concentrations of respirable crystalline silica?

Louis Anthony (Tony) Cox Jr.

Cox Associates, 503 Franklin St., Denver, CO 80218, United States

ARTICLE INFO

Article history:
Received 11 July 2016
Received in revised form
25 August 2016
Accepted 6 September 2016
Available online 13 September 2016

Keywords:
Respirable crystalline silica (RCS)
Accuracy
Reliability
Laboratory error
Estimation error
Performance statistics

ABSTRACT

Permissible exposure limits (PELs) for respirable crystalline silica (RCS) have recently been reduced from 0.10 to 0.05 mg/m^3. This raises an important question: do current laboratory practices and standards for assessing RCS concentrations permit reliable discrimination between workplaces that are in compliance and workplaces that are not? To find out, this paper examines recent laboratory performance in quantifying RCS amounts on filters sent to them to assess their proficiency. A key finding is that accredited laboratories do not reliably (e.g., with 95% confidence) estimate RCS quantities to within a factor of 2. Thus, laboratory findings indicating that RCS levels are above or below a PEL provide little confidence that this is true. The current accreditation standard only requires laboratories to achieve estimates within three standard deviations of the correct (reference) value at least two thirds of the time, rather than a more usual standard such as within 25% of the correct value at least 95% of the time. Laboratory practices may improve as the new PEL is implemented, but they are presently essentially powerless to discriminate among RCS levels over most of the range of values that have been tested, leaving employers and regulators without a reliable means to ascertain when workplace RCS levels are above or below the PEL.

1. Introduction

Respirable crystalline silica (RCS), which consists of minute quartz particles, increases risk of silicosis in people or animals exposed to sufficiently high concentrations for sufficiently long durations. To reduce or eliminate risk of silicosis in occupationally exposed workers, regulators, employers, employees, and labor organizations have worked together to reduce greatly exposure concentrations of RCS in the air of many workplaces since the 1960s. Fig. 1 shows that, as hoped, silicosis mortalities in the U.S. have declined dramatically, by over 90% since the 0.10 mg/m^3 Permissible Exposure Limit (PEL) was established decades ago to protect worker health. Evidence from clinical, toxicological, epidemiological, and industrial hygiene studies (Cox, 2011), as well as the historical record in Fig. 1, suggest that it has been effective in doing so.

Despite this dramatic progress, silica exposures in some workplaces in recent decades have remained far above the current (and former) PEL levels. As noted by the Centers for Disease Control and Prevention (CDC) (64 MMWR 23, June 19, 2015):

"Results indicate that despite substantial progress in eliminating silicosis, silicosis deaths continue to occur. Of particular concern are silicosis deaths in young adults (aged 15–44 years). These young deaths likely reflect higher exposures than those causing chronic silicosis mortality in older persons, some of sufficient magnitude to cause severe disease and death after relatively short periods of exposure. A total of 12 such deaths occurred during 2011–2013, with nine that had silicosis listed as the underlying cause of death."

From a risk management perspective, it is natural to wonder whether it is possible to further reduce silicosis mortalities and morbidities through improved compliance monitoring and enforcement and/or by reducing the PEL to its current value (0.05 mg/m^3). Would the lower PEL prevent more deaths, or, to the contrary, would rigorously complying with and enforcing the former 0.10 mg/m^3 limit achieve all the human health benefits available? To find out, it is important to consider how accurately workplace exposures are currently monitored and enforced and the possibilities for improving compliance with PELs. A regulation that tells employers "Do not exceed concentration C" can only be effective if it is possible to determine, with useful reliability, when a workplace is in compliance. As permitted concentrations become

E-mail address: tcoxdenver@aol.com.

http://dx.doi.org/10.1016/j.yrtph.2016.09.008
0273-2300/

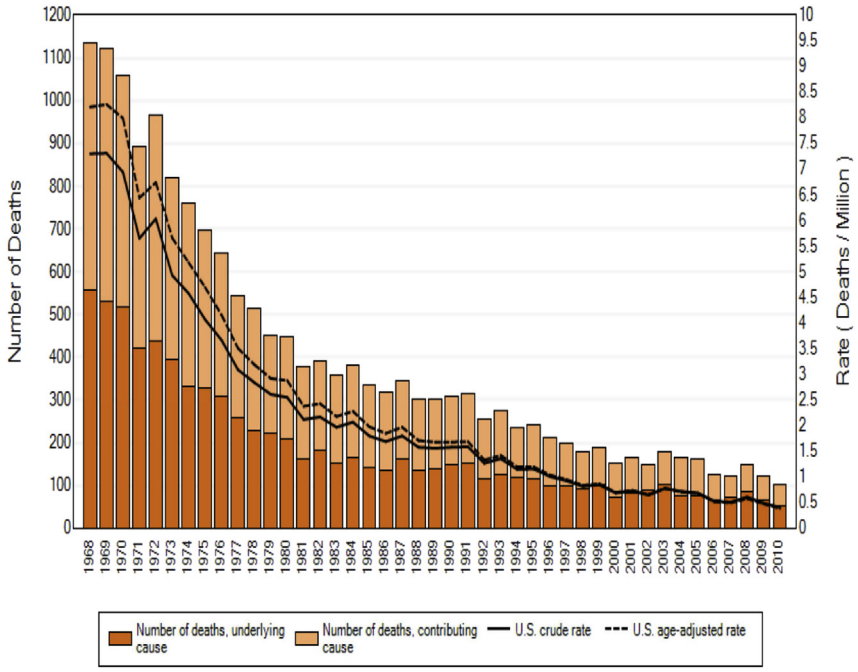

Fig. 1. Annual silicosis-associated mortalities have decreased by over 90% since the 0.10 mg/m^3 PEL was established. (*Source*: U.S. Centers for Disease Control and Prevention, 2015).

lower, determining compliance can become more difficult, requiring increasingly accurate and precise laboratory measurements. Enjoining employers to comply with standards for which compliance cannot be determined reliably risks ineffective action, diverting limited resources to false positives (acting to reduce workplace exposures based on false findings that they are too high) or failing to act due to false negatives (i.e., laboratory findings mistakenly indicating that little or no action is needed). Such errors and potential for useless activity arise whenever occupational exposures are reduced until they are comparable to the "noise," or random measurement error and variability, in laboratory results — or, conversely, whenever the random variability in laboratory results is large enough to obscure the effects that they seek to detect. The science-policy question of how best to set PEL concentrations to protect worker health then becomes complicated by the practical reality that compliance with target concentration levels cannot easily be determined.

The remainder of this paper examines how well today's workplace RCS concentrations can be determined from the laboratory results that are currently used in determining and enforcing compliance. A previous paper described an experiment in which filters with known loads of RCS in different matrices (and also as pure crystalline silica samples) were sent to five different commercial laboratories to determine how accurately and reliably they measured these known loads (Cox et al., 2015). The overall results were striking: the laboratories did not consistently discriminate between concentrations that differed by a factor of 2, and even filters with no crystalline silica load were sometimes misidentified as containing substantial RCS concentrations. However, these findings came from an artificial experiment, leaving open the possibility that laboratories perform better in practice on real samples. This paper tests that possibility by quantifying the variability in results from commercial laboratories, including laboratories accredited by the American Industrial Hygiene Association (AIHA) Industrial Hygiene Proficiency Analytical Testing (IH-PAT) program (www.aihapat.org). These include laboratories used by numerous employers and other organizations to determine whether workplace exposures comply with PELs.

1.1. Data and methods

Employers and other organizations often send workplace air sampling filters to accredited laboratories to determine current workplace air concentrations of RCS. The laboratories, in turn, may use several different analytical methods, such as X-ray diffraction, infrared spectroscopy, and colorimetry to estimate the quantity of RCS on a received filter; of these, X-ray diffraction is widely considered one of the most accurate and reliable analytic methods. All data discussed in this paper were derived by X-ray diffraction. The laboratory returns to the submitting entity the estimated quantity of RCS (mg) on the filter from a given volume of air sampled. If the values ("lab results") returned are sufficiently high, interventions to reduce workplace air concentrations of RCS should be triggered.

To maintain IH-PAT accreditation, laboratories must meet AIHA-specified criteria for proficiency in estimating quantities of RCS (and other substances) sent to them for analysis. This is done as follows. Four times per year (each being called a "round"), AIHA sends to each participating lab four spiked sample filters, prepared from continuously agitated homogeneous suspensions with four different known concentrations; thus the filters received by different laboratories contain approximately the same known loadings of RCS, called samples. Fig. 2 shows a photograph of such samples (front row, clearly identifying the interfering substances) as well as typical real-world samples (back row) which display only unique sample numbers. As explained in AIHA methods documentation, the RCS analytes measured in these accreditation tests consist of "Free silica (quartz) on four 5.0 μm [pore] 37-mm [diameter] PVC (polyvinyl chloride) filter samples containing differing silica concentrations and include a background matrix, on a rotating basis of coal mine dust, talc, calcite, or a combination" (https://www.a2la.org/scopepdf/3300-01.pdfAIHA, 2016).

Each laboratory analyzes the IH-PAT-prepared samples and reports the estimated quantities of RCS back to IH-PAT. The reported

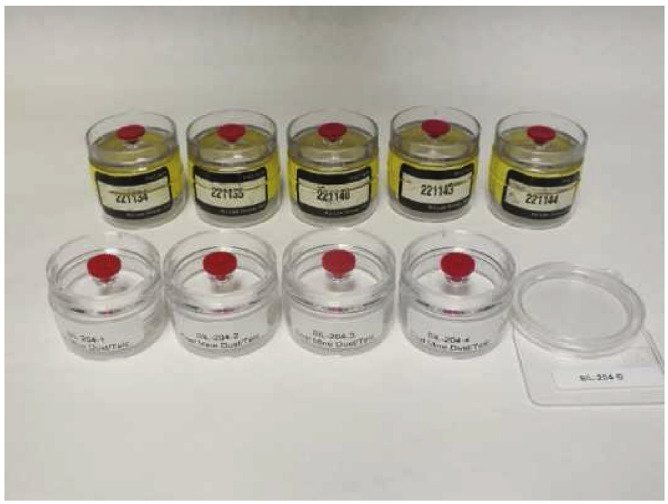

Fig. 2. Real-world (back row) and AIHA-prepared (front row) filters sent to laboratories for analysis.

results from different laboratories for corresponding samples in the same round should be the same if there are no errors or variability in the process. In practice, as discussed further in the Results section, there is considerable variability in the estimated quantities of RCS for these matched samples (although less than in earlier decades) consistent with previous literature (Harper et al., 2014; Stacey, 2007; Shulman et al., 1992; Abell and Doemeny, 1991). The key statistical criterion for determining a laboratory's proficiency is that its results must fall within three standard deviations of the IH-PAT "assigned value" (or "reference mean"). A laboratory is rated proficient if it has passing scores (i.e., within three standard deviations of the reference value for at least 75% of the analyte samples) in at least 2 of the last 3 consecutive rounds (http://www.aihapat.org/Programs/IHPAT/Documents/IHPAT Scheme Plan R2.pdfAIHA, 2013, p. 16). (Conversely, a laboratory whose estimates never fell within a six standard deviation range around the reference value for any of the samples in every third round, and that also failed to land within this range on up to 25% of samples in the remaining rounds, would nonetheless be classified as proficient. Thus, even if such a laboratory could never provide reasonable (e.g., 95%) confidence that any individual result met any specified accuracy criterion (e.g., being no more that 10 times higher or lower than the reference value), it could still meet the IH-PAT criteria for proficiency.) To determine the assigned value for a given sample in a given round, IH-PAT first identifies a subgroup of the participating labs, termed reference labs, based on prior analytical performance. Using the reference labs' reported results, IH-PAT Winsorizes any outliers (replacing them with less extreme values, see https://cran.r-project.org/web/packages/robustHD/robustHD.pdf) and calculates arithmetic means and standard deviations of the reference lab results.

As described by OSHA (https://www.osha.gov/dsg/etools/silica/faq/faq.html, accessed 6/30/2016):

The PAT program is designed to help consumers select laboratories that are proficient. In the PAT program analyses of quartz, the "true" values against which a laboratory's results are compared are based on results from reference laboratories that are a subset of the participating laboratories. Assuming that the PAT samples were made from accurately delivered consensus reference material and that the participants all used the same techniques, instrumentation and methodology, and that the samples are not otherwise flawed so as to introduce bias, the best accuracy that can be achieved by consensus analyses is limited by the standard error of the precision of that analysis [SD/(n)½, where SD is the standard deviation in the results among the n reference labs]. … The current method of PAT quartz sample generation is by aerosol generation using "5 μm" Min-U-Sil 5 without cyclones. In addition to any errors in the generation process, this "total dust" approach introduces a sampling error that may not duplicate the sampling error associated with the use of a cyclone.

In the PAT program, these generation and sampling errors are recognized as significant and are evaluated in statistical tests conducted on sub-batches and batches of PAT samples by the contract laboratory that prepares them. … The results obtained by participants in the PAT program therefore include both the analytical error the participating laboratories introduce and an unknown but potentially large amount of error introduced in the generation and sampling of the aerosol. These latter errors may vary batch to batch."

Multiple years of data on the estimated quantities of RCS returned by different laboratories in response to the spiked samples sent out via the AIHA-PAT program are available on-line in pdf format at www.regulations.gov/#!documentDetail;D=OSHA-2010-0034-4188 from an AIHA-PAT program submission to OSHA; they are also available as Excel files from the present author. Table 1 shows the layout of the data used in all subsequent analyses. The "Lab #" code is a numerical code that uniquely identifies each laboratory (based on our re-coding of the original much longer codes). Data from 26 AIHA-PAT accredited labs (one per row) and for the most recent 2 rounds for which data are available (one per column) are shown, since the most recent data are assumed to be most representative and relevant for current testing conditions. Round 194 took place in July of 2013. (Including more years of data reinforces our findings, but risks losing relevance; Harper et el., 2005, discuss key changes over time in the IH-PAT program.) The numbers in Table 1 represent estimated quantities (mg) of RCS returned by each lab (row) for each sample and round (column); empty cells indicate missing values, e.g., because a lab dropped out of the AIHA-PAT program.

In analyzing these data, we emphasized exploratory and descriptive data analysis and non-parametric methods to avoid introducing potentially erroneous and biased modeling assumptions. The following sections present results of statistical plots of conditional empirical cumulative frequency distributions, nonparametric (smooth) regression, and Spearman's rank correlations to compare and visualize laboratory-specific results vs. approximate true values, i.e., the reference values.

2. Results

Fig. 3 shows individual laboratory estimates of RCS amounts for each reference (estimated true) value. Each small circle represents one result returned by a laboratory. Reference values are on the x axis and the laboratory's corresponding estimated values are on the y axis. If all estimates were perfectly correct and there were no errors in the process, all data points would fall on the line shown (the line of perfect calibration), which equates estimated (y axis) and reference (x axis) values. In practice, individual laboratories return estimated RCS levels that vary widely around the reference values, as indicated by the vertical spread of the individual laboratory results around the line. For example, the spiked samples with a reference value (i.e., estimated true value) of just under 0.14 mg elicited individual laboratory estimates ranging from about 0.06 to over 0.18 mg. Since there are only 26 participating accredited labs in this study, this wide range implies that an employer or other entity who receives a laboratory report of, say, 0.06 or 0.09 mg (the two lowest individual laboratory estimated

Table 1
Data layout for AIHA-PAT data.

Lab #	Sample value 193 01	Sample value 193 02	Sample value 193 03	Sample value 193 04	Sample value 194 01	Sample value 194 02	Sample value 194 03	Sample value 194 04
1	0.08	0.06	0.21	0.09	0.08	0.09	0.08	0.14
4	0.08	0.06	0.22	0.12	0.12	0.15	0.08	0.20
5					0.11	0.15	0.08	0.18
7	0.09	0.07	0.21	0.11	0.12	0.12	0.07	0.17
8	0.10	0.07	0.21	0.11	0.10	0.13	0.06	0.18
9					0.11	0.12	0.07	0.18
12	0.10	0.08	0.24	0.13				
14	0.07	0.06	0.18	0.11	0.08	0.06	0.18	0.11
15	0.06	0.13	0.13	0.18	0.10	0.14	0.07	0.15
16	0.09	0.07	0.21	0.11	0.05	0.14	0.07	0.17
24	0.07	0.06	0.20	0.09	0.11	0.12	0.10	0.18
25	0.09	0.07	0.24	0.13	0.14	0.18	0.08	0.21
26	0.13	0.05	0.20	0.07	0.11	0.16	0.07	0.20
31	0.09	0.07	0.22	0.11	0.13	0.17	0.07	0.24
32	0.08	0.05	0.16	0.07	0.11	0.13	0.06	0.17
34	0.09	0.07	0.21	0.10	0.13	0.16	0.07	0.20
35	0.08	0.04	0.15	0.11	0.11	0.15	0.09	0.20
36	0.08	0.06	0.17	0.11	0.10	0.15	0.07	0.17
38	0.07	0.05	0.21	0.12	0.10	0.14	0.06	0.18
39	0.07	0.05	0.19	0.08	0.10	0.13	0.06	0.18
40	0.09	0.08	0.17	0.14	0.12	0.14	0.06	0.11
45	0.09	0.07	0.24	0.07	0.11	0.15	0.07	0.19
46					0.13	0.16	0.08	0.21
47	0.09	0.06	0.21	0.12	0.12	0.13	0.07	0.20
48	0.09	0.07	0.22	0.14	0.13	0.13	0.08	0.19
49					0.12	0.15	0.08	0.20

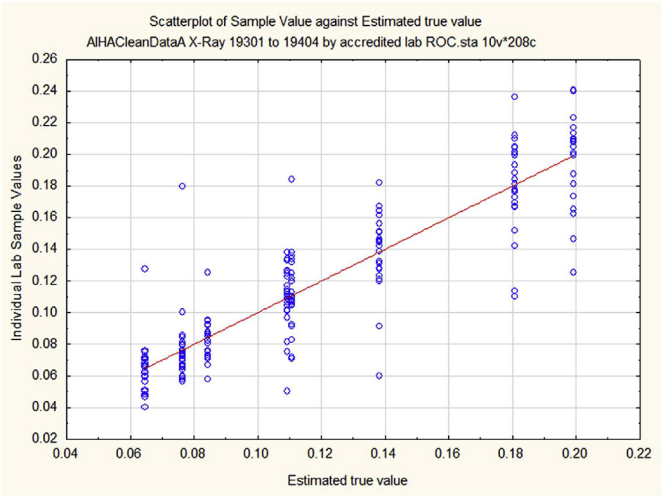

Fig. 3. Individual laboratory values (circles) are widely spread around reference mean values (line). All values are in mg.

values for this reference value) cannot be reasonably sure (e.g., 95% confident) that the true value is less than 0.10 mg. Conversely, the wide vertical scatter of estimates around the reference values that are below 0.10 mg implies that receiving a lab result of 0.12, or even 0.18 mg, does not imply that the true value exceeds 0.10 mg.

Fig. 4 provides a different perspective on this variability in laboratory results by showing laboratory-estimated values (rounded to the nearest mg) on the x axis and corresponding reference values on the y axis. A nonlinear regression (nonparametric smoother) curve is fit to this scatter plot. Very high sample values returned by laboratories tend to over-estimate the reference values (e.g., a sample value of 0.24 mg returned by a laboratory corresponds to an average reference value of about 0.20 mg, as estimated by the regression curve); conversely, very low individual laboratory values tend to under-estimate corresponding reference values. There is very substantial variability in the reference values corresponding to a single estimated sample value, as shown by the vertical range of results (small circles) for specific sample values on the x axis.

The small circles in Figs. 3 and 4 show that, collectively, laboratory results are quite variable, despite the care exercised by the AIHA-PAT program in preparing spiked samples that should all yield closely similar values in the absence of laboratory error. It is natural to wonder whether this might be due to a few individual laboratories that are consistently higher or lower than the rest. To find out, we used Spearman's rank correlations to test whether laboratories that gave higher (or lower) estimates than most others in one quarter also tended to do so over time. The test was carried out by computing the 28 ($= (8 \times 7)/2$) pairs of ordinal correlations between rankings of labs based on sizes of RCS estimates (for the same reference value) in each of the 8 columns in Table 1. Six of the 28 pairs of Spearman's rank correlations differed significantly from zero at the conventional $p = 0.05$ significance level, and all six were positive (with numerical values between 0.39 and 0.73). This provides significant evidence that laboratories that give relatively high or low results compared to others in one quarter are more likely to do so again in another quarter. However, the effect is relatively small, and the wide range of variability in sample RCS estimates shown in Figs. 3 and 4 reflects variability that is more pervasive than one or a few laboratories.

Finally, Table 2 shows the fraction (column 2) and number (column 4) of estimated true values (based on means of results from multiple laboratories) that exceed 0.1 for each sample value (rounded to the nearest mg), and the number of times each sample value was returned. This table allows estimation of the rates of false positives and false negatives that would be generated by different decision rules and frequency distributions of sample values if the desired goal is to take action whenever the true value exceeds 0.10 mg. For example, suppose that the loads on filters are such that each sample value from 0.05 to 0.15 is equally likely, and that no action is taken unless the sample value returned by a laboratory is

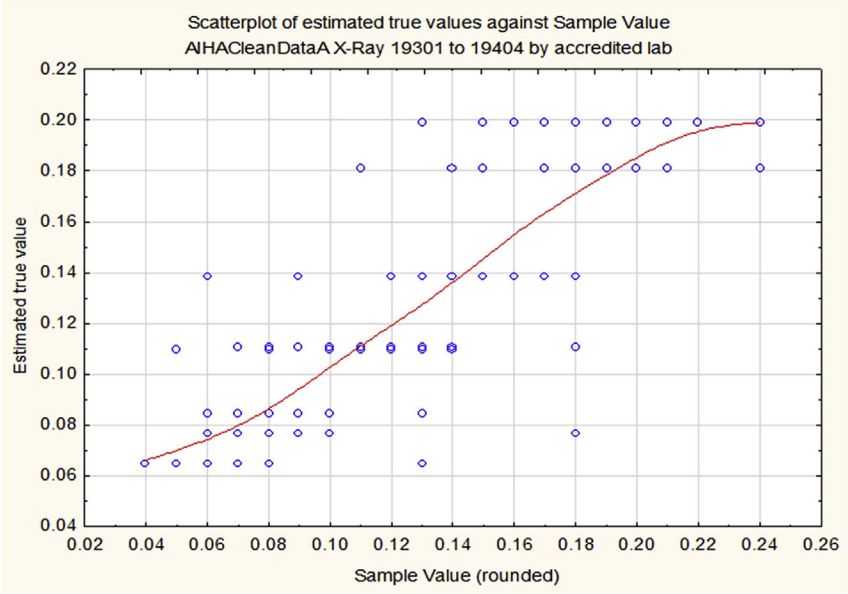

Fig. 4. Estimated true values (y axis) are widely spread around sample values (x axis). All values are in mg.

at least 0.10 mg. If an employer receives a sample value in the range from 0 to 0.09, then the probability that it is a false negative is the average of the false negative rates for sample values from 0.05 to 0.09; this is 0.13. If the sample value is in the range from 0.10 to 0.15, triggering action, then the probability that it is a false positive is $(0.33 + 0.14)/6 = 0.078$, reflecting the false positives at sample values of 0.10 and 0.13. The false negative rate could be reduced to zero by always taking action, but then the false-positive rate would increase to 42% (calculated by averaging the false positive rates for each sample value from 0.05, 0.06, …, 0.15, namely, 0.80 = 1–0.20 for sample value 0.05, and 0.92, 0.88, 0.82, 0.77, 0.33, 0.00, 0.00, 0.14, 0.00, 0.00 for the rest). An employer who wants to be reasonably confident (e.g., 95%) that a returned sample value will be neither a false positive nor a false negative cannot find any decision rule that will deliver such assurance, given the high variability in sample values around true values. The only way to achieve it is to reduce the variability of laboratory results.

3. Discussion

The practical implications of the variability in laboratory estimates of RCS quantities are potentially important to employers, employees, and regulators who rely on such results to determine compliance and need for interventions. If a laboratory returns an estimated value of 0.06 mg for a submitted air sample filter, for example, then Fig. 4 shows that the corresponding reference values (and hence the true values that they approximate) range from a low of about slightly above 0.06 mg to a high of about 0.14 mg (based on the empirically observed range of reference values that generated estimated values of 0.06 mg when sent to the 26 accredited laboratories). One of the four distinct reference values shown for an estimated value of 0.06 mg is 0.14 mg, and all are higher than the estimated value of 0.06 mg. Similarly, a laboratory estimate of 0.13 could correspond to a true value anywhere between about 0.06 mg and 0.20 mg, while an estimate of 0.18 mg could correspond to a

Table 2
Sample values (rounded to nearest mg) vs. estimated true values.

Sample value (rounded)	Estimated fraction of true values > 0.1	N = number of cases	Estimated number of cases wth true values > 0.1
0.04	0.00	1	0
0.05	0.20	5	1
0.06	0.08	13	1
0.07	0.12	25	3
0.08	0.18	17	3
0.09	0.23	13	3
0.1	0.67	9	6
0.11	1.00	16	16
0.12	1.00	11	11
0.13	0.86	14	12
0.14	1.00	8	8
0.15	1.00	8	8
0.16	1.00	4	4
0.17	1.00	7	7
0.18	0.90	10	9
0.19	1.00	3	3
0.2	1.00	8	8
0.21	1.00	9	9
0.22	1.00	3	3
0.24	1.00	4	4

true value anywhere between about 0.08 and 0.20 mg. Thus, no laboratory result between 0.06 and 0.18 mg can be relied on by the employer, employee, or regulator to confidently (e.g., with 95% confidence) discriminate between true (or reference) values above and below 0.10 mg. To the contrary, returned values between 0.06 mg and 0.18 mg (spanning most of the design range for spiked sample values, which run from 0.05 to 0.20 mg) convey very little information about the probable true value, since any returned value between 0.06 mg and 0.18 mg is compatible with a wide range of true values.

That laboratories provide such relatively uninformative estimates for individual samples in no way contradicts or undermines the fact that, on average, higher sample values do indeed correspond to higher reference values, as shown by the nonparametric regression curve in Fig. 4. However, individual employers and employees cannot get the benefits of this useful aggregate relation, since they receive only the individual results of submitted sample filters, and these individual results are too variable to provide trustworthy indications of whether the sampled workplaces are above or below a given limit. Acting to reduce RCS exposures (e.g., by increasing dust controls or administrative controls and use of respirators), or failing to take such measures, on the basis of laboratory-measured values does not provide a reliable approach to taking action when, and only when, appropriate.

4. Conclusions

Various authoritative agencies attribute high value to the information provided by laboratory analyses of RCS. For example, the US Occupational Health and Safety Administration states that "Analytical results on the quartz content of the air samples are necessary to evaluate whether the OSHA PEL is exceeded" (www.osha.gov/dsg/etools/silica/faq/faq.html). AIHA states that "The purpose of proficiency testing is to provide interested parties with objective evidence of a laboratory's capability to produce data that is both accurate and repeatable for the activities listed in its scope of accreditation. A laboratory's competence can be demonstrated through favorable proficiency testing data. This is important to clients, potential customers, accreditation bodies, and other external entities." Likewise, the International Standards Organization (ISO), for which the IH-PAT program provides conformity assessment, states that "The need for ongoing confidence in laboratory performance is not only essential for laboratories and their customers but also for other interested parties, such as regulators, laboratory accreditation bodies and other organizations that specify requirements for laboratories" (ISO/IEC, 2010). These statements about the importance of trustworthy laboratory performance, however true, stop short of addressing the fundamental challenge revealed by the data in this study: the variability in laboratory results is large enough compared with a 0.10 mg/m^3 limit so that results returned by many laboratories do not reliably indicate whether workplace RCS concentrations are above or below that limit. The current (as of October, 2013) AIHA-PAT program considers laboratories proficient as long as their results fall within an interval of six standard deviations (three in each direction) of the mean for reference laboratories for at least 75% of the silica samples at least 2/3 of the time (i.e., in at least 2 of each 3 consecutive rounds): "IHPAT participant results are rated acceptable or unacceptable for each unique analyte sample number. A passing score is 75% or more acceptable results for an analyte group. A participant is rated proficient for the applicable IHPAT analyte group if the participant has a passing score for the applicable IHPAT analyte group in two (2) of the last three (3) consecutive PT rounds." (www.aihapat.org/Programs/IHPAT/Documents/IHPAT%20Scheme%20Plan%20R2.pdf, p.16) For the IH-PAT program data in Table 1, Figs. 3 and 4 show that this is not a sufficiently demanding criterion to assure usefully accurate and reliable results.

In considering how to improve the current system, it is useful to distinguish between how laboratories *can* perform, how they *do* perform, and how they *should* be required to perform. Our analysis has focused on the second of these three questions. The first question – what levels of performance are technically feasible – has been addressed at length by OSHA (2016) as follows:

> "[P]art of OSHA's technological feasibility assessment of a new or revised health standard includes examining whether available methods for measuring worker exposures have sufficient sensitivity and precision to ensure that employers can evaluate compliance with the standard and that workers have accurate information regarding their exposure to hazardous substances. Consistent with the Supreme Court's definition of 'feasibility', OSHA finds that it is feasible to measure worker exposures to a hazardous substance if achieving a reasonable degree of sensitivity and precision with sampling and analytical methods is 'capable of being done' … NIOSH concluded in its post-hearing comment that 'current methods of sampling and analysis for respirable crystalline silica have variability that is acceptable to demonstrate compliance with the proposed PEL and action level.' … Based on the record evidence reviewed in this section, OSHA finds that current methods to sample respirable dust and analyze samples for respirable crystalline silica by XRD and IR methods are capable of reliably measuring silica concentrations in the range of the final rule's PEL and action level. This finding is based on the following considerations: (1) Several sampling devices are available that conform to the ISO/CEN specification for particle-size selective samplers with a level of bias and accuracy deemed acceptable by international convention, and moving to the ISO/CEN convention will maintain continuity with past practice, (2) both the XRD and IR methods can measure respirable crystalline silica with acceptable precision at amounts that would be collected by samplers when airborne concentrations are at or around the PEL and action level, and (3) laboratory proficiency data demonstrate that there is reasonable agreement between laboratories analyzing comparable samples most of the time."

Thus, in OSHA's judgment, technology is sufficient to make reliable measurement of silica concentrations technically feasible. However, the results in Figs. 3 and 4 indicate that this potential for reliable measurement is often not yet achieved in practice, suggesting the possibility that more stringent requirements on laboratory performance might help to close the gap between what is possible and what is done today.

Previous research has noted that large differences among laboratories contribute to variability in PAT program estimates of the RCS content in samples (Shulman et al., 1992), although variability has declined since the introduction of the PAT program in 1972, in part due to changes in RCS samples and in laboratory procedures (*ibid* and Harper et al., 2014). Maciejewska (2006) found that regular use of quality control methods for free silica determination was positively associated with proficiency of laboratories, suggesting that the variability of PAT estimates for RCS can potentially be reduced by such methods. Thus, although Figs. 3 and 4 show that current (2013) variability in laboratory estimates is too great to discriminate reliably among reference concentrations in the range of approximately 0.06 mg–0.18 mg (and hence to assess compliance with PELs for workplaces with concentrations between about half and about double a PEL corresponding to 0.10 mg), it is plausible that this variability could be reduced by stricter quality control for laboratories making RCS determinations, as OSHA's 2016 final

rule reducing the PEL from 0.10 to 0.05 mg/m^3 requires, but demonstrating that it has been successfully accomplished appears to be a prerequisite for obtaining reliable results (Lee et al., 2016).

More stringent statistical quality requirements may also be essential for obtaining more useful results. For example, instead of requiring only that laboratories come within three standard deviations of the reference value on 75% or more of samples in at least 2 of 3 consecutive rounds, AIHA might adopt the NIOSH criterion that "the method must provide results that are within ±25% of the expected ("true") values at least 95 times out of 100" (Ashley, 2015). Such accuracy would require greatly reducing the variability shown in Figs. 3 and 4, where far more than 5% of results lie further than 25% away from the expected values (given by the regression curves). This might be accomplished either by having individual laboratories make greater use of quality control measures for RCS determinations or by modifying compliance determination rules to make greater use of *averages* of RCS values assessed by multiple laboratories, to adjust for the fact that average values are relatively reliable, but individual sample values are currently too variable to meet accuracy criteria such as NIOSH's.

Our findings also have potentially important implications for monitoring and enforcement of the new (2016) OSHA limit of 0.05 mg/m^3 and the new action level of 0.025 mg/m^3. In general, as quantified in earlier studies (Cox et al., 2015), current laboratory procedures do not reliably quantify crystalline silica levels that differ by a factor of 2 (e.g., 0.10 vs. 0.05 mg/m^3). Performance was not improved at lower exposure levels, suggesting that the information obtained from laboratories today is inadequate for monitoring and enforcing compliance with the 0.10 mg/m^3 PEL (Lee et al., 2016), and *a fortiori* for the new, lower PEL (Cox et al., 2015; Lee et al., 2016). In the current data set, relative variability of sample values around estimated true values increases at lower RCS concentrations, as indicated by the fact that the coefficient of variation decreases significantly with RCS concentration level (Pearson's correlation coefficient is −0.7; Spearman's rank correlation coefficient is −0.8, both being statistically significantly different from zero at the conventional 0.05 significance level).

This paper has focused on exposure estimation, rather than on health effects. However, the observed high variability of exposure estimates raises the question for risk assessment of how substantial errors in exposure estimates affect estimated exposure-response relations used in determining standards to protect human health, especially for lung diseases with threshold-like nonlinearities in their exposure concentration-response curves (Cox, 2011). Rhomberg et al. (2011) study the effects of exposure measurement errors on such estimated exposure-response curves and note that "the degree of bias known to apply to actual studies is sufficient to produce a false linear result … The consequences of this could be great, as it could lead to a misallocation of resources towards regulations that do not offer any benefit to public health." Their concern is that findings of elevated risks at relatively low exposure concentrations (e.g., at levels mandated in current or recent standards) could arise from elevated exposures that were misclassified or mis-estimated as being much lower than their true values, leading to the appearance of substantial health risks at exposure levels that do not actually cause them. Whether this is the case for RCS remains to be determined, but the substantial rates of false negatives (i.e., exposures in excess of desired standards that are mis-estimated as being below them) in this study suggest that exposure measurement error is highly relevant for determining to what extent (if any) RCS concentrations that are truly at or below current and recent standards cause increased risks of disease.

Until laboratory practices and/or statistical protocols for determining compliance are radically improved to give demonstrably more accurate results, the current high prevalence of false-positive and false-negative conclusions about compliance implied by Fig. 4 of this study can be expected to inhibit effective allocation of resources to improve and protect worker health. Compliance with OSHA's recent 0.10 mg/m^3 or the current 0.05 mg/m^3 PEL cannot be determined reliably by all laboratories without substantial improvement in current analytical and statistical practices. Until these improvements are made and credibly demonstrated, enforcement activities that are based largely on random noise or error in laboratory results will provide neither incentives nor capability to continue reducing out-of-compliance levels of exposure that could threaten worker health.

Acknowledgment

This research was supported in part by the The National Stone, Sand, and Gravel Association (NSSGA). All data analyses performed and all conclusions drawn are solely those of the author. The author thanks Dale Drysdale and Kelly Bailey for insightful discussions of current employer and laboratory procedures.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.yrtph.2016.09.008.

References

Abell, M.T., Doemeny, L.J., 1991 Aug. Monitoring the performance of occupational health laboratories. Am. IndHygAssoc J. 52 (8), 336–339.

AIHA, 2013. Industrial Hygiene Proficiency Analytical Testing (IHPAT) Program Scheme Plan. Revision 2. www.aihapat.org/Programs/IHPAT/Documents/IHPAT%20Scheme%20Plan%20R2.pdf.

AIHA, 2016. Scope of Accreditation to ISO/IEC 17043:2010. AIHA Proficiency Analytical Testing Programs. www.a2la.org/scopepdf/3300-01.pdf.

Ashley, K.N.I.O.S.H., 2015. Manual of analytical methods 5th edition and harmonization of occupational exposure monitoring. GefahrstReinhaltLuft 2015 (1–2), 7–16.

CDC (Centers for Disease Control and Prevention) (2015). Update: Silicosis Mortality – United States, 1999–2013. Morbidity and Mortality Weekly Report 64(23): 653-654. https://stacks.cdc.gov/view/cdc/31626 (accessed 09.15.16)

Cox Jr., L.A., 2011 Oct. An exposure-response threshold for lung diseases and lung cancer caused by crystalline silica. Risk Anal. 31 (10), 1543–1560.

Cox Jr., L.A., Van Orden, D.R., Lee, R.J., Arlauckas, S.M., Kautz, R.A., Warzel, A.L., Bailey, K.F., Ranpuria, A.K., 2015 Oct. How reliable are crystalline silica dust concentration measurements? Regul. Toxicol. Pharmacol. 73 (1), 126–136.

Harper, M., Sarkisian, K., Andrew, M., 2014. Assessment of respirable crystalline silica analysis using Proficiency Analytical Testing results from 2003-2013. J. Occup. Environ. Hyg. 11 (10), D157–D163.

ISO/IEC, 2010. ISO/IEC 17043:2010(en) Conformity assessment — General requirements for proficiency testing. www.iso.org/obp/ui/#iso:std:iso-iec:17043:ed-1:v1:en.

Lee, R.J., Van Orden, D.R., Cox, L.A., Arlauckas, S., Kautz, R.J., 2016 Oct. Impact of muffle furnace preparation on the results of crystalline silica analysis. Regul. Toxicol. Pharmacol. 80, 164–172.

Maciejewska, A., 2006. [Analysis of the competences of workplace inspecting laboratories for the determination of free crystalline silica (FCS), based on proficiency testing results]. Med. P. R. 57 (2), 115–122. Polish.

OSHA (Occupational Safety and Health Administration) (2016). Silica Advisor: Frequently Asked Questions. https://www.osha.gov/dsg/etools/silica/faq/faq.html (accessed 09.15.16)

Rhomberg, L.R., Chandalia, J.K., Long, C.M., Goodman, J.E., 2011 Sep. Measurement error in environmental epidemiology and the shape of exposure-response curves. Crit. Rev. Toxicol. 41 (8), 651–671.

Shulman, S.A., Groff, J.H., Abell, M.T., 1992 Jan. Performance of laboratories measuring silica in the Proficiency Analytical Testing program. Am. IndHygAssoc J. 53 (1), 49–56.

Stacey, P., 2007 Jan. Measurements of silica in air: reliability at new and proposed occupational exposure limits. J. Occup. Environ. Hyg. 4 (1), D1–D4.