OSHA Docket No. OSHA-2015-0015; (RIN) 1218–AC94

Comments on Specific Questions Posed by OSHA Regarding Additional PortaCount Quantitative Fit Testing Protocols

Submitted by:
Clifton D. Crutchfield, PhD
10851 N. Poinsettia Dr.
Tucson, AZ 85737
cdcrutch@comcast.net

QUESTION #1 - WERE THE THREE STUDIES DESCRIBED IN THE PEER-REVIEWED JOURNAL ARTICLES WELL CONTROLLED, AND CONDUCTED ACCORDING TO ACCEPTED EXPERIMENTAL DESIGN PRACTICES AND PRINCIPLES?

As described, the study designs did not follow the guidelines for conducting an evaluation of a new fit test protocol as set forth in Annex A2 of ANSI/AIHA Z88.10 – Respirator Fit Test Methods (AIHA, 2010). Section 7.2 of the Standard describes the "approved" particle counting instrument procedure. Section 7.2.1 states:

> *"Particle-counting instruments typically use the particles in the ambient air as the challenge aerosol. This eliminates the need for aerosol generators and test chambers."*

The three studies submitted by TSI were conducted in a test chamber using generated NaCl particles to "augment" the "ambient" challenge aerosol. A previous TSI pilot study [OSHA-2015-0015-0001] of possible test methods was also conducted in a test chamber using "augmented ambient" challenge concentrations >60,000 p/cm^3.

The reference method used for the studies is therefore not the same as the "approved" reference method described above. According to OSHA-2015-0015-0009:

> *For the Fast-Half method, the target ambient particle concentration was 10,000 p/cm^3 (range 5,000 - 20,000 p/ cm^3). In the experimental trials, the mean ambient particle concentrations were 9,073 - 41,916 p/cm^3, with overall mean 20,722 p/cm^3. For the reference method, the mean ambient particle concentrations were 9,304 - 37,520 p/cm3, with overall mean 20,889 p/cm^3.*
>
> *For the Fast-Full method, the target ambient particle concentration was 10,000 p/cm3 (range 5,000 - 20,000 p/ cm^3). In each experimental trial, the mean ambient particle concentrations were 11,268 - 36,604 p/ cm^3, with overall mean 21,586 p/ cm^3. For the reference method, the mean ambient particle concentrations were 11,494 - 36,730 p/ cm^3, with overall mean 21,741 p/cm^3.*

The minimum ambient concentration required by the PortaCount to test elastomeric respirators is 1,000 p/cm^3 [TSI RESFT 205]. The vast majority of the fit tests conducted with the >12,000

PortaCounts currently in use [OSHA-2015-0015-0012] are not conducted in test chambers. The ambient (non-augmented) challenge aerosol concentrations used for those tests can be 20 or more times lower than those used in the Fast-Half and Fast-Full studies. They can be up to 60 times lower than the pilot study used to determine the proposed Fast protocol exercises. Given the potential for such disparities, what surety can there be that the reported test results are applicable to PortaCount fit testing as it is done in the real world?

> *It is ironic that PortaCount measurements of generated aerosol in a test chamber were also used in the study that is most often cited as the "validation" study for the PortaCount fit test protocol (daRoza, 1991). In that study, PortaCount measurements of fit using ambient aerosol were also made during the same mask donning prior to test subjects entering the generated aerosol chamber. However, those ambient aerosol measurements were too variable to be analyzed or used. The use of generated aerosol was one of many aspects of the original PortaCount validation study that clearly did not meet the basic requirements of a new fit test method 'validation' study (Crutchfield, 1993). Of the three QNFT methods accepted by OSHA, only CNP technology has met the QNFT validation criteria established by the American National Standards Institute. The aerosol-based QNFT methods were essentially grandfathered into OSHA acceptance and ANSI's Respirator Fit Test Methods Standard Z-88.10 without validation studies. The ability of aerosol-based fit test systems to effectively measure respirator leakage has not been validated against known respirator leakage or primary reference standards (Allen, 2015).*

It is also ironic that, after publishing an Application Note in 2012 containing the statement below, TSI now seeks to reduce the "accepted" PortaCount test protocol time by significantly reducing its ambient/mask sample time and by eliminating the majority of currently required PortaCount fit test exercises. A big question is how much is potentially lost by trying to save 280 seconds?

> *The OSHA ReDon fit test protocol used by the OHD Quantifit fit tester appears to be significantly faster than the OSHA protocol used for the TSI Portacount®fit tester. However, if all the time spent by each fit test subject is taken into account, the time difference becomes trivial. [Application Note RFT-008, Speed of Fit Testing with the TSI Portacount Respirator Fit Tester vs. the OHD Quantifit Fit Tester, 2012]*

QUESTION #2 - WERE THE RESULTS OF THE THREE STUDIES DESCRIBED IN THE PEER-REVIEWED JOURNAL ARTICLES PROPERLY, FULLY, AND FAIRLY PRESENTED AND INTERPRETED?

Both the Fast-Half and Fast Full articles contain the following statement in their Methods section:

> *Zhuang et al. (2004) concluded that the ambient aerosol concentration did not affect the measured fit factor.*

What the authors did not disclose is the fact that the cited Zhuang (2004) study used PortaCount N-95 Companion fit test data from a companion NIOSH study of QLFT methods. Given the very narrow aerosol size range used by the N95 Companion, the conclusion stated above may make sense for N-95 fit tests of filtering facepiece respirators. However, the Fast-Half and Fast-Full articles in question are about fit testing elastomeric respirators.

The Introduction section of the same Zhuang article cited above states:

> *A data set of 20,974 PortaCount fit tests conducted at various workplaces was obtained and analyzed by Zhuang et al. (1999). The ambient concentrations were divided into five levels and Duncan's multiple range test was performed to determine if the fit factors vary among the five levels (Montgomery, 1984). Overall fit factors were found to be significantly dependent on the ambient concentration in both the half-mask and the full-facepiece data sets (p-value < 0.01).*

The statement about challenge agent concentration not affecting measured fit factors should also be considered in the context of the numerous workplace protection factor (WPF) studies that are documented in the occupational health literature. A strong correlation between measured fit factors and WPFs has been very elusive for a very long time (Crutchfield et al., 1993). The greatest challenge confronted by WPF research efforts has been the low concentrations of ambient challenge agent available in actual work places. The only alternative to insufficient challenge agent concentration is to greatly increase mask sample time and try to increase the sensitivity of mask leakage detection (Janssen, 2014). The very sensitive laboratory-based analytical techniques that have been applied to WPF studies conducted in actual work environments have generally produced limited numbers of very hard to get in-mask samples with incredibly high geometric standard deviations.

> *The resulting FFs had a geometric mean (GM) of 400 (range=10–6010) and a geometric standard deviation (GSD) of 6.1. Of the 55 valid donnings, 43 were good fitting (FFs≥100) and 12 were poor fitting (FFs<100).* ***The WPFs had a GM of 920 (range=13–230,000) and a GSD of 17.8.*** *(Zhuang, 2003)*

The low reported correlations between measured fit factors and workplace protection factors are not surprising, and do not offer much of a basis for validating fit test systems.

QUESTION #3 - DID THE THREE STUDIES TREAT OUTLIERS APPROPRIATELY IN DETERMINATION OF THE EXCLUSION ZONE?

OSHA has already addressed the issue of how the exclusion zone around the required fit factor was handled in the studies submitted by TSI.

A second major issue involves the introduction and use of a **5-sec** "normal breathing" (NB) in-mask sample. The wide range of measured fit factors reported in the three studies was obviously based on a wide range of test mask leakage (much of it related to the planned use of poorly-fitting masks). The reported result for the Half-Fast study was 54 reference fit factor failures

(3 with fit factors < 5). Eleven fit test pairs were omitted from the analysis because the ratio of maximum to minimum NB fit factors was > 100.

An obvious question not addressed in the study is the validity of a 5-sec in-mask sample when the following factors are considered:

- The actual sample volume read by the CNC detector would be <30 cm^3.

- What portion of ***one*** breathing cycle was actually sampled, and with what consistency? In-mask airflow dynamics associated with factors such as aerosol streaming, leak location, and lung deposition (Myers et al., 1986; Oestenstad et al., 1990; Crutchfield et al., 1997) cause in-mask concentrations to vary significantly between the inhalation and exhalation cycles of normal breathing.

- Given the numerous in-mask aerosol sampling biases that have been identified in the scientific literature, how much was each 5-sec NB sample impacted by the in-mask flow dynamics at the moment of collection?

- A previous study involving 10-sec in-mask ambient aerosol samples (Sreenath, 2001) did not meet ANSI requirements for test sensitivity.

Another question not addressed in the submitted studies is the impact of eliminating the normal/reference protocol's 11-sec mask purge between each of the four fit test exercises specified for the new Fast protocols. Without the mask purge, how can the results and impact of each exercise be independently evaluated or substantiated?

QUESTION #4 - WILL THE PROPOSED PROTOCOLS GENERATE REPRODUCIBLE FIT-TESTING RESULTS?

The issue of reproducibility of fit test results was not directly addressed in the papers submitted by TSI to OSHA. However, a study of a modified PortaCount test protocol involving shortened in-mask sample periods was previously conducted by TSI. (Sreenath et al., 2001) That study involved collecting and recording in-mask aerosol samples at 10-sec intervals during 671 PortaCount fit tests. The fit tests were conducted in the field on half-mask respirators using the standard OSHA fit test protocol. Study data analysis involved comparing overall fit factors derived from shorter in-mask sample times (ranging from 10 to 50 sec) to the simultaneously-sampled 60-sec sampling period.

> *"It is evident from Table I that the correlation coefficient increases as the mask sample time is increased. The Pearson's correlation coefficient is fairly high, even for a sample time of l0 seconds and increases to as high as 0.99 when the sample time is increased to 50 seconds. The data for the l0-second mask sample has a larger standard deviation than the data with 60-second sample time. This is illustrated in Figure 3, a plot of the ratio of overall fit factor obtained using a l0-second mask sample to the overall fit factor obtained using a 60-second sample time."* (Sreenath, 2001; pg 983)

The log plot of 10-sec vs. 60-sec overall fit factors referred to above as Figure 3 clearly shows that a large number of the 10-sec fit factors were substantially higher (less respirator leakage detected) than their corresponding 60-sec fit factors.

An examination of Table VIII in the submitted Fast-Half article also sheds some light on the question of reproducible results. When the ratio of maximum to minimum fit factor for each exercise set was analyzed, ~65% of the Reference Method fit tests had ratios exceeding one order of magnitude. More than 10% of those tests had ratios exceeding two orders of magnitude.

Ratios for the Fast-Full Method showed >30% exceeding one order of magnitude and ~5% exceeding two orders of magnitude. That is a lot of within fit test variability. The unanswered question is whether the source of the variability was due primarily to poor test mask fit, exercise effect, or method measurement biases.

QUESTION #5 - WILL THE TWO PROPOSED PROTOCOLS RELIABLY IDENTIFY RESPIRATORS WITH UNACCEPTABLE FIT AS EFFECTIVELY AS THE QUANTITATIVE FIT TESTING PROTOCOLS, INCLUDING THE STANDARD PORTACOUNT QNFT PROTOCOL, ALREADY LISTED IN APPENDIX A OF THE RESPIRATORY PROTECTION STANDARD?

Given the assumptions made and methods used to conduct the submitted studies, I see no clear evidence that the two proposed protocols will reliably identify respirators with unacceptable fit. The issue of applicability of studies conducted in a test chamber using "augmented" ambient aerosol to PortaCount fit testing in the real world was addressed in Question #1 above.

Additional concerns arise from the methods, assumptions and rationale used by TSI to develop the proposed Fast protocols. The process was essentially spelled out in two forums:

- A presentation - Gregory Olson, Richard Remiarz, and Jeff Weed: *The Evolution and Evaluation of a Fast Fit Test Protocol Based on the TSI PortaCount® Respirator Fit Tester.* Presented at the International Society for Respiratory Protection European Section Meeting; November 18, 2015; Delft, Netherland.

- A White Paper submitted to OSHA titled *Exercise Selection Rationale for the Fast-full, Fast-half, and Fast-FFR Studies* [TSI, 2015].

It appears that a major problem with the pilot and subsequent studies arose from respirator selection.

> *A pilot study was conducted in 2012-2013 to determine if a shortened fit test protocol could pass ANSI criteria, which requires a minimum of 50 failed fit tests for the analysis. If respirators are properly sized and donned, a large number of tests is required to achieve enough failures. To achieve more failures some test subjects were given incorrect size respirators or were intentionally donned improperly. No user seal checks permitted.*

Assuming that the most appropriate respirator class has been chosen to protect a worker's health, respirator size is obviously the next most influential determinant of respirator performance. Respirator donning has also proven to be one of the major determinants of the protection afforded by the respirator. Using improper sizing or donning technique to achieve poor fit introduces the potential for major instability and variation in a 'validation' study that is supposed to accredit all future uses of a new fit test protocol in the real world. TSI rejected the idea of using measurable known fixed leaks into "properly sized and donned" respirators because *"The use of artificial leak paths (eg., capillary tubes, etc.) would not have allowed evaluation of dynamic leaks that are caused by exercising."*

Concerns about the selection and omission of fit test exercises for the proposed Fast protocols are addressed in Question #8 below. The issue of evaluation of dynamic leaks, as well as respirator leakage in general, is another area of major concern that has been around for a long time. A 1982 review of qualitative fit test methods conducted by NIOSH to support the inclusion of qualitative fit testing in OSHA's Lead Standard (OSHA, 1982) defined a false positive error as a QLFT method failing a fit test that was passed by the generated-aerosol (standard/reference/criterion) method.

> *One important factor increases the costs of QLFT relative to QNFT. The record shows that the rate of false positives resulting from QLFT exceeds the rate of false positives from QNFT (Ex- 48). This means that the QLFT methods reject a greater number of respirators that actually have acceptable fits than the QNFT methods do, and thus result in more retesting of subjects to attain adequate fit. Estimates of the rate of false positives in QLFT range from 38 percent to 90 percent (Ex- 34E). Using available studies in the record (Ex. LANL, Ex. 6-40, Ex.16, and Ex. 16-A), OSHA calculates an average false positive incidence for QLFT of about 39.5%. The cost implications of this are significant. For instance, on the basis of one study (Ex- 5-38), OSHA found that 149 tests were required to qualitatively fit test 98 employees. (OSHA, 1982)*

I disagree with OSHA's interpretation of the QLFT false positives, which was that 39.5% of the time the QLFT methods failed good fitting respirators (**based on assumed validity of generated aerosol QNFT results**). The basic presumption supporting Qualitative Fit Testing (QLFT) is if a test subject can taste, smell, or otherwise detect a QLFT test agent during a fit test, the respirator being tested has a leak rate higher than the QLFT pass level and should therefore fail the fit test. QLFT methods are generally considered to be less sensitive and more subjective than QNFT methods. However, the data cited above by OSHA shows that the standard/reference/criterion generated aerosol method exhibited very poor fit test sensitivity by passing 40% of the fit tests that the QLFT methods had failed when the test subjects tasted, smelled, or otherwise detected a QLFT challenge agent.

Similar results and conclusions are seen in a comparison of N95 respirator fits using Bitrex Qualitative and TSI PortaCount+® Quantitative Fit Testing (Clapham et al., 2000). Of the 79 total fit test pairs, the Bitrex QLFT method passed 11 and failed 68 fit tests, while the PortaCount+ passed 44 and failed 35. When the study data are assessed using Bitrex as the Reference method, the PortaCount+ method is assigned a test sensitivity of 0.46. The authors expressed a similar concern about QLFT false positives to the one expressed by OSHA (1982).

> *The results of this study indicate that the use of bitrex during qualitative fit testing of N95 disposable filtering facepieces results in an increase in failure and/or rejection in cases where a TSI PortaCount (plus N95 companion accessory) quantitatively establishes an acceptable fit.* [Clapham et al., 2000]

Bitrex also exhibited greater test sensitivity than the PortaCount in studies reported by Coffey (1999) [N = 125, Bitrex failed 41% of fit tests passed by PortaCount]; Janssen (2002) [N = 75, CNP failed 24, Bitrex failed 22, PortaCount failed 15]; and by Nelson (2004) [Table V].

Why did the aerosol-based QNFT methods used as the test reference method in the comparison studies cited above fail to detect the respirator leakage sensed and reported by the QLFT method? A large number of factors that bias (generally by decreasing) aerosol-based measurements of respirator leakage have been reported by multiple researchers. They have been tabulated, summarized, and discussed (Crutchfield, 1993, 2009) and are very relevant to the current evaluation of the proposed Fast protocols.

Multiple comparisons of matched pairs of PortaCount and controlled negative pressure (CNP) fit test results strongly reinforce the concerns about PortaCount respirator leak measurement capabilities reported in the studies cited above. During a number of comparative studies conducted with both fixed (head form/breathing machine) systems and human subjects, CNP fit test systems using static exercise protocols have consistently measured far more respirator leakage with substantially greater accuracy, precision, and speed than aerosol-based systems using dynamic exercise protocols (Crutchfield, 2009). A summary of major findings includes:

- A study of the ability of PortaCount and CNP fit test methods to detect known levels of leakage introduced into test subject respirators during the same mask donning (Crutchfield et al., 1995) showed that the PortaCount method detected an average of 21% of the known leak with a coefficient of variation (CV) of 62%, while the CNP method detected an average of 105% of known leakage with a CV of 10%.

- The same study design was employed using a breathing machine-headform system (Crutchfield and Park, 1997). A series of fixed leaks were introduced via matched hypodermic needles into three separate locations (bridge of nose, cheek, and chin) in both half-mask and full-face respirators mounted on the headform. The CNP system detected an average of 98.4% of the known leakage introduced into the respirators during 96 separate fit tests, with a coefficient of variation of 4.3%. An analysis of variance revealed that CNP measurements were unaffected by leak location. PortaCount measurements of the same leaks averaged 40.3% of the known leak rate, with a coefficient of variation of 46.9%. An analysis of variance detected significant differences in the ambient aerosol system's measurements of leakage as a function of leak location. The observed pattern of leak detection strongly supports the theory of in-mask sampling bias related to particle streamlining reported by Myers (1986) and Oestenstad (1990). The overall increase in PortaCount known leak detection (40.3% vs. 20.8% detected in a similar study done with human subjects (Crutchfield et al., 1995)) was attributed to the absence of challenge aerosol lung deposition in the bellows of the breathing machine.

- A meta-analysis of the results reported in seven different comparison studies from multiple sources was performed using a total of 754 paired CNP and ambient aerosol fit tests (Crutchfield, 2009). All data included in the meta-analysis met the following selection criteria:
    - **a.** All fit tests were conducted on **human subjects wearing respirators of appropriate size**
    - b. All data points represent results from **paired sequential fit tests** conducted with each fit test system **during the same mask donning**
    - c. **No fixed leaks were introduced** into subject respirators during the fit tests
    - d. **No paired fit test results meeting the criteria outlined above were excluded**.

- Results of the meta-analysis are shown in Table 1 below.

Table I. Results of Meta-Analysis of Paired Overall Fit Factors from Seven CNP – Ambient Aerosol Fit Test System Comparison Studies. (Crutchfield, 2009)

Parameter	Ambient Aerosol	CNP		RATIO Variable
Mean FF	24638	1534		78.32
Median FF	11101	1141		11.05
Geo. Mean FF	9646	953		10.46
5^{th} Percentile FF	575	115		4.98

The lack of fit test sensitivity exhibited by the PortaCount is consistently displayed throughout the range of fit factors measured during the comparison studies. Even the lower five percent of that range showed a 5-fold difference between paired PortaCount and CNP fit factors. Such disparities in leak measurement capability exhibited by the two methods raise important questions about the lack of validated sensitivity of the ambient aerosol fit test method. (See also QUESTION #6 response below)

QUESTION #6 - DID THE PROTOCOLS IN THE THREE STUDIES MEET THE SENSITIVITY, SPECIFICITY, PREDICTIVE VALUE, AND OTHER CRITERIA CONTAINED IN THE ANSI/AIHA Z88.10-2010, ANNEX A2, CRITERIA FOR EVALUATING FIT TEST METHODS?

The test-sensitivity value of 0.95 is the only test statistic designated by ANSI in its Fit Test Methods standard as a criterion value that "shall" be met when accepting new fit test methods. The other criteria specified in Annex A2 of ANSI's Respirator Fit Test Methods standard are designated as evaluation criteria that "should" be met.

A good deal of confusion exists in the fit test literature regarding the test sensitivity of fit test methods. In simple terms, test sensitivity is an expression of the probability of avoiding false positives (which occur in the fit test world when a new method or protocol passes a respirator that was failed by the standard/reference fit test method). A lower test-sensitivity means a higher probability of producing false positives (i.e. passing unacceptable respirators).

Trying to evaluate the sensitivity of a revised PortaCount fit test protocol by using the standard OSHA PortaCount protocol presents a real quandary because the sensitivity of the standard PortaCount protocol has itself not been established. An extensive literature search, including queries to OSHA and the U.S. Army (Department of the Army, 1996), did not produce a study that meets ANSI criteria in terms of validating PortaCount capability to effectively measure respirator leakage (penetration) or fit. That search also did not produce a validation study for the generated aerosol fit test method, which has been used for decades as the "reference standard method" to which most other fit test methods (qualitative and controlled negative pressure) have been compared. Unfortunately, the results produced by aerosol-based fit test methods cannot be directly traced to primary calibration standards, so the quandary is not easily resolved. However, the fact that it exists should not be overlooked, or simply dismissed as the way things are, by regulatory and standards-setting organizations, or by the respiratory protection research community.

The comparison data (Crutchfield, 2009) used to conduct the meta-analysis cited in the response to Question #5 above can also be used to examine the test sensitivity of the PortaCount fit test method currently accepted by OSHA. Using the validation criteria specified by ANSI and the OSHA accepted CNP method as the reference method, the following test results were obtained.

A total of 16 fit test pairs (same mask donning; 14 CNP, 2 PortaCount) were excluded from the total of 754 test pairs because they fell within one standard deviation of the reference fit factor. The distribution of the remaining 736 fit test results are shown in Table 2 below. The results of the ANSI validation test are shown in Table 3.

Table 2. 2 x 2 Contingency Table Specified by ANSI Z88.10-2010 Standard

	Failed Reference (CNP) Test	Passed Reference CNP) Test
Passed PortaCount OSHA Protocol	A = 104	B = 610
Failed PortaCount OSHA Protocol	C = 15	D = 7

Table 3. ANSI Validation Test Criteria and Results for PortaCount OSHA Protocol

Statistic		Criterion	Result
Test Sensitivity	C / (A+C)	> 0.95	0.13
Predictive Value of a Pass	B / (A+B)	$\geq$ 0.95	0.85
Test Specificity	B / (B+D)	$\geq$ 0.50	0.99
Predictive Value of a Pass	C / (C+D)	$\geq$ 0.50	0.68
Kappa Statistic		> 0.7	0.17

The findings of the comparison studies of PortaCount vs other fit test systems discussed in the response to Question #5 as well as above raise additional questions and concerns about the proposed new Fast protocols.

- Does the currently accepted (without validation) PortaCount method have the proven level of fit test sensitivity to qualify as the reference method for validation of the proposed new Fast PortaCount protocols? What evidence supports such a conclusion?
- What reduction in fit test sensitivity could potentially occur when a new fit test protocol 'validated' in a test chamber using an "augmented" ambient aerosol challenge agent is subsequently used in the wide variety of places and conditions under which fit tests are conducted in the truly ambient real world?

QUESTION #8 - DOES THE ELIMINATION OF CERTAIN FIT-TEST EXERCISES (*E.G.*, NORMAL BREATHING, DEEP BREATHING, TALKING) REQUIRED BY THE EXISTING OSHA-APPROVED STANDARD PORTACOUNT® PROTOCOL IMPACT THE ACCEPTABILITY OF THE PROPOSED PROTOCOLS?

Elimination of the normal breathing, deep breathing, and talking fit test exercises from the proposed Fast protocols has significant potential for adverse impact on PortaCount fit test results in the real world. It is ironic that TSI now seeks to eliminate those exercises after ruling out such exercise elimination in previous studies.

> *Even though the correlation between exercises was found to be high,* ***we recognize the importance of all eight exercises****; hence, our attempt to develop a faster test protocol will not alter or exclude any of the eight exercises currently required by the OSHA standard, except to shorten exercise duration for good-fitting respirators only.* (Sreenath et al., 2001)

TSI's rationale for fit test exercise elimination was based on 1) a literature review, 2) undocumented informal conversations with industry fit test experts, and 3) *additional analysis of TSI data* [OSHA-2015-0015-0001] generated for a separate purpose that *uncovered an unexpected trend within the data for the talking exercise*.

1) **Reported Literature Review Results:**

> *There was little useful data available. One of the most recent and also most thorough research papers on the subject was from NIOSH: Zhuang (2004) concluded that Talking, Bending and Head-up-down were the three most critical exercises for all 3 respirator types.*

The TSI Rationale Paper and submitted study data analyses neglected to mention that the Zhuang (2004) paper was based on fit test data generated using a PortaCount N95-Companion. The Rationale Paper also did not mention or cite a number of published studies with findings directly related to Question #8. The first normal breathing (donning), deep breathing, and talking fit test exercises have been found to generally produce the lowest ambient aerosol fit factors by a number of researchers.

> *The correlation factors are fairly high for all exercises except for the first normal breathing exercise and the talking exercise. Other researchers have also observed this effect and the first normal breathing exercise is often referred to as the donning fit factor.* [18-19] **…..** *The donning fit factor is typically less than fit factors obtained by other exercises for a given sample time. The first normal breathing exercise is also significantly lower than the second normal breathing exercise.* (Sreenath et al., 2001) [Study conducted by TSI]
>
> *For overall fit factors greater than or equal to 100, the fit factor for "reading" or "the first normal breathing" was found to be the lowest fit factor among fit factors for each exercise for about 60% of the fit tests for half-mask respirators using six exercises; the normalized fit factors for these two exercises were also the smallest and second smallest. Fit factors for any exercise were found to be highly correlated with fit factors for other exercises. Fit factors for the first normal breathing were significantly smaller than those for the second normal breathing. Similar results were observed with the fit test data for full-facepiece respirators.* (Zhuang et.al, 1999)
>
> *Figure 4 is a comparison of individual exercise log fit factors at l0 seconds and 60 seconds (for half masks).The donning fit factor was significantly lower than the other exercise fit factors except the talking and bending exercises. The talking exercise yields the lowest fit factor (similar to the results by Zhuang et al.,*[18] *Crutchfield et al.,*[19] *and Bentley et al.*[21] *and it is evident from Figure 4 that the fit factors after the talking exercise begin to recover and increase as the exercise regime is continued. These trends do not change as the in-mask sample time is increased from 10 seconds to 60 seconds.* (Sreenath et al., 2001)

An in-depth review of the effects of mask donning and test exercises on measured respirator fit can be found in an article titled "Effect of Exercise and Mask Donning on Measured Respirator Fit" (Crutchfield et al., 1999). The effects of modified (shortened) fit test protocols are also examined. Conclusions drawn from the study that was based on 840 separate fit tests include:

- Respirator donning has a greater effect on respirator fit than do fit test exercises.
- The only test exercises that produced fit factors significantly lower than donning fit factors were the talking and bend over exercises.
- The lower fit factors produced by the talking exercise appear to be more consistent with sampling artifact than with actual exercise dynamics.
- The bend over exercise appears to be predictive of poor mask fit.
- A modified aerosol fit test protocol made up of fewer, more targeted fit test exercises identified unacceptable respirator fit as effectively as the longer OSHA-specified protocol.

- The controlled negative pressure fit test system, using the same modified test protocol, measured significantly more leakage in test respirators than the ambient aerosol system using either protocol.

- The time saved by reducing the number of fit test exercises could be better utilized to test more than a single mask donning.

2) Undocumented Informal Conversations:

Informal conversations with industry fit test experts can be expected to contribute little to evaluations of new respirator fit test protocol validity and acceptability if they are not documented.

3) Additional Analysis of TSI Data:

The expressed rationale for elimination of the normal breathing (donning), deep breathing, and talking fit test exercises from the proposed Fast protocols seems to be primarily based on TSI's additional analysis of the Talking White Paper data [OSHA-2015-0015-0001]. Their Rationale is confusing in places and not well substantiated. For example, TSI begins its Exercise Selection Rationale paper [OSHA-2015-0015-0008] by confusing fit test sensitivity with specificity, and by expressing the objective of respirator fit testing as simply a search for poor respirator fit.

> *The ANSI Z88.10-2010 method for evaluating new fit test methods was studied in detail. It became apparent that the* ***specificity requirement of 0.95*** *would be very difficult to achieve, given the inherent variability of respirator fit. In order to have a good chance of passing ANSI, any new protocol as a whole would have to be more rigorous than the reference (OSHA) method.*
>
> *The objective of the fit test is not to determine the protection level afforded by the respirator during typical use. The objective is to determine if the respirator fits well, or does not. The best way to do that is to identify and use rigorous exercises that have been shown to expose poor fits, without unnecessarily failing good fits. Exercises that do not expose poor fits are not useful for fit testing.*

It appears that the demonstrated major causes of poor respirator fit (ie. respirator selection and donning) strongly influenced the results of the study and data analysis conducted by TSI to formulate its rationale for test exercise elimination, as evidenced by the following statements:

- *Male and female fit test subjects with little or no respirator experience were recruited for fit testing.*

- *In an effort to generate a high percentage of poor fits, researchers would sometimes intentionally mis-match the respirator size and face size (via visual estimation).*

- *Subjects were not permitted to perform user seal checks.*

- *Two complete OSHA fit tests were run without pausing in-between. Ambient samples were recorded before and after each exercise just like is done with the commercial version of the PortaCount software. This had the effect of creating one very long (14-exercise) fit test.*

- *Research conducted more recently by TSI [analysis of talking exercise] reveals that in order to determine the most critical exercises, fit test results must be separately analyzed for poor-fitting and good-fitting respirators because the good-fitting data biases the results. The TSI white paper submitted to OSHA: "Analysis of the Talking Exercise Used for Respirator Fit Testing," shows that the talking exercise is not effective for identifying poor-fitting respirators.*

In addition to issues involving actual data independence and the limited/missing (1 – 60 sec?) sampling of respirator donning effect, the decision to segregate the pilot data by test result and to focus their analysis primarily on the failure data generated by the deliberate poor fits included in the test data was questionable.

> *Pilot study data was analyzed in the same manner used by Zhuang (2004) in that the frequency of occurrences where each exercise produced the lowest fit factor was computed. Except that unlike Zhuang, fit test data was separated into three groups: All fit tests, good-fitting fit tests, and poor-fitting fit test. A poor-fitting fit test was defined as any test where at least one exercise failed (using pass=500 for full-face and 100 for half-face and FFR). The importance of isolating poor-fitting respirators/fit tests is* ***readily apparent****.* **[??]** *The data group for poor-fitting respirators is of primary interest because we are attempting to identify the exercises that are the best for challenging the face seal and identifying poor fitting respirators. Exercises with fit factors that are above the pass level do not help identify poor fits.*

Based on their *Table FF1*: *Percentage of Times Exercise was Lowest for Full-face Respirators* poor fitting only data (N = 32), *Normal breathing (NBl, NB2), Deep breathing (DB) and Talking (T) were deemed non-rigorous for full-face respirators.* In addition:

> *The anomaly between the first and second normal breathing, NB1 (0%) vs. NB2 (19%), is puzzling. Zhuang (2004) found both NB1 and NB2 to be non-critical. We reasoned that if normal breathing was a truly effective exercise, this would have been seen during NB1. We concluded that the in-mask particles measured during NB2 had to be residual particles left over from bending, exacerbated by the high in-mask volume and subsequent slow purging of full-face respirators. The NB2 results were thus discounted, leaving SS, UD and B as the most rigorous exercises and the primary candidates for the new protocol.*

The poor fitting only data (N = 21) shown for Half-face respirators in Table HF1 exhibit a very different exercise result pattern from the one shown for Full-face respirators.

> *For half-face respirators, bending (B) was again the most rigorous for poor-fitting respirators (19%) (Table HF1). However, none of the other exercises stood out (10 -14%).*

The discrepancy between NB1 and NB2 did not appear, which was attributed to the low in-mask volume and quick purging compared to full-face respirators.

Due to the lack of data suggesting that half-face respirator fit tests should use different exercises than full-face respirators, we decided to move forward with an identical protocol for both types. Only the pass level is different. The new protocol has been named the Fast-Half Protocol.

Since the proposed Fast-Full and Fast-Half protocols start with a 30 sec Bending exercise, the following questions need to be considered and answered:

- Which exercise is supposed to assess the effect of donning the test respirator? Can an assessment of fundamental respirator fit be made using Fast Protocol data?

- If the generally assumed purpose of the second (Jogging) exercise is to assess the potential for a shift in fundamental respirator fit during conduct of the fit test, what should the Jogging fit factor be compared against? There are no NB1 or NB2 tests in the proposed protocols.

- Given the assumptions about differences in mask purging between full-face and half-face respirators used above to explain the observed anomaly/discrepancies in NB1 and NB2 test data, what are the anticipated effects of **omitting between exercise mask purging** from future PortaCount fit tests conducted with the proposed Fast protocols in the real world?

- The cited rationale for omitting the Talking exercise from the proposed Fast-Full/Half Protocol but including it in the Fast-FFR Protocol is contradictory and far from convincing, especially given the poor fitting data (N = 43) results for the Talking exercise (2%) shown in Table FFR1.

I do not agree with TSI's interpretation (*shown in italics below)* of the reason that the Talking fit factor often appears as the lowest fit factor measured during PortaCount fit testing. I think a more viable explanation can be found in Crutchfield et al. (1999), pg. 835.

The authors suspect that the phenomenon regarding the talking exercise is due to body generated particles. Talking involves the high-frequency vibration of wet vocal chords. This may be generating particles (droplets) in a way that does not occur during the other exercises where the subject is not talking. Sub-micrometer body-generated particles are likely to be misinterpreted as leakage particles by particle counting instruments

Sub-micrometer droplets???

REFERENCES

Allen, L. (2015) RESPIRATORY PROTECTION: Professional Trust Should Be Earned. Occupational Safety and Health, May, 2015.

American National Standards Institute. (2010) American National Standard Respirator Fit Testing Methods. Am. Ind. Hyg. Assoc., Fairfax Va. ANSI Z 88.10.

Clapham SJ, Stephenson DJ, Wallace DO, Lillquist DR, and Suruda AJ. (2000) Comparison of N95 Disposable Filtering Facepiece Fits Using Bitrex Qualitative and TSI PortaCount® Quantitative Fit Testing. Int. J. Occup. Envr. Health 6(1).

Coffey C, Zhuang Z, and Campbell D. (1999) The Efficacy of the Bitrex Qualitative Fit Test Method with N95 Filtering-facepiece Respirators. Int. Soc. Resp. Prot. International Conference, Oct, 1999.

Crutchfield, C., and Van Ert M. (1993) An Examination of Issues Affecting the Current State of Quantitative Respirator Fit Testing. J. Int. Soc. Resp. Prot. 11(2):5-18.

Crutchfield C, Park D, Henshel J, et al. (1995) Determinations of Known Respirator Leakage Using Controlled Negative Pressure and Ambient Aerosol QNFT Systems. Am. Ind. Hyg. Assoc. J. 56(1):16-23.

Crutchfield C, and Park D. (1997) Effect of Leak Location on Measured Respirator Fit. Am. Ind. Hyg. Assoc. J. 58(6):413-417.

Crutchfield C. (2009) The Relationship Between Fit Factors Measured with Controlled Negative Pressure and Ambient Aerosol Fit Test Systems. Submitted to J. Int. Soc. Resp. Prot. MS#09-195.

da Roza RA, Biermann AH, Foote KL, McCormack C, and Sachett CR. (1991) Evaluation of PortaCount for Determining Respirator Fit Factors, Part III: Human Subject Tests and Comparison with an Aerosol Photometer. J. Int. Soc. Resp. Prot. 9(3):22-38.

Department of the Army (1996): Letter to Senator John McCain, September 17, 1996.

Janssen L, Michael D. Luinenburg M, Mullins H, and Nelson T. (2002) Comparison of Three Commercially Available Fit-Test Methods. Am. Ind. Hyg. Assoc. J. 63(6).

Janssen L, Zhuang Z, and Shaffer R. (2014) Criteria for the Collection of Useful Respirator Performance Data in the Workplace. J. Occup. Environ. Hyg. 11(4): 218–226.

Myers, W.R., J. Allender, R. Plummer and T. Stobbe: "Parameters that Bias the Measurement of Airborne Concentration Within a Respirator". Am. Ind. Hyg. Assoc. J. 47(2):106-114 (1986).

Oestenstad RK, Perkins JL, and Rose VE. (1990) Identification of Faceseal Leak Sites on a Half-Mask Respirator. Am. Ind. Hyg. Assoc. J. 51(5):280-284.

Nelson T and Mullins H. (2004) Recommendations for the Acceptance Criteria for New Fit Test Methods. J. Int. Soc. Resp. Prot. 21(1).

OSHA (1982): Occupational Exposure to Lead: Respirator Fit Testing. Federal Register Vol. 2, No. 219, Friday, November 12, 1982, Rules and Regulations.

Sreenath A, Weed J, and Church T. (2001). A Modified Protocol for Quantitative Fit Testing using the PortaCount. AOEH 16(10):979-988.

Zhuang Z, Coffey C, Myers W, McMullin D and Jensen P. (1999) The Effect of Ambient Aerosol Concentration and Exercises Used For Fit Testing on Quantitative Fit Factors. J. Int. Soc. Resp. Prot. 17(3):68.

Zhuang Z, Coffey CC, and Lawrence RB. (2004) The Effect of Ambient Aerosol Concentration and Exercise on PortaCount® Quantitative Fit Factors. J. Int. Soc. Resp. Prot. 21:11-20.

Janssen L, Zhuang Z, and Shaffer R. 2014 Criteria for the Collection of Useful Respirator Performance Data in the Workplace. J Occup Environ Hyg.; 11(4): 218–226.