Letter to the Editor

September 24, 2018

Journal of the International Society for Respiratory Protection
Attn: Ziqing Zhuang, Ph.D.
105 EagleBrook Court
Venetia, PA 15367
JISRPEditor@ISRP.com

To the Editor:

We are writing in regard to the recent articles by Wu, et al. (2017, 2018). The articles present details of studies performed to validate respirator fit testing methods using a new optical particle counter. The authors state that these studies were conducted in accordance with the requirements of the validation protocol in ANSI Z88.10-2010, Annex A2. Unfortunately, it appears that the authors have misunderstood the requirements of this validation method and as a result we believe they have reached an incorrect conclusion in at least one case.

The first flaw in their procedure was the selection of a reference fit test method. Clause A2.1 of the standard states:

> *"While the ideal comparison procedure has yet to be proven, this annex provides a specific procedure for evaluating fit test methods against the current body of knowledge." … "It involves a statistical comparison of a new fit test method against the generated aerosol method described in this standard."*

The generated aerosol method referred to is described in Clause 7.1 of that document ("Generated Aerosol Quantitative Fit Test Procedure"). The PortaCount® method is defined as a separate method, as described in Clause 7.2 ("Particle-counting Instrument Quantitative Fit Test Procedure"). In general, generated aerosol methods use a photometer to measure the concentration of particles inside and outside of the respirator. Han, Willeke, and Colton (1997) describe several photometric-based fit test methods, although they are not explicitly referred to as "generated aerosol" methods, since that term was introduced by OSHA in 1998. This vagueness in the definition of the generated aerosol method in the ANSI document is a weakness of that standard.

Clause A2.1 of the standard does permit one exception to the selection of the reference test method:

> "A proposed modification to an accepted QNFT procedure can be evaluated using the accepted protocol for that instrument as the reference standard."

If the MT-05U and the PortaCount® were comparable instruments then one might make the argument that substituting the MT-05U is a simple modification of an accepted QNFT procedure. However, as the authors note, there are multiple differences between the instruments in terms of the detection method, the lowest detectable particle size and the maximum measurable particle concentration. In view of these significant differences, one must conclude that the generated aerosol method is the only acceptable reference test method to be used for validation of a fit test method employing the MT-05U.

The second, more subtle, but equally important flaw in the procedure by Wu, et al. is in the treatment of outliers in their analysis of the data. Specifically, the authors used a statistical test to identify and eliminate outliers from their data sets. In doing so, they excluded pairs of data which affect the results. This is evident

most significantly in Figure 5C of the 2017 publication. As shown in that figure, there were at least five instances in which the reference fit test yielded a fit factor less than 500 (the Reference Fit Factor for full facepiece respirators) while the new method yielded fit factors greater than 500. The authors flagged these data as statistical outliers and eliminated them from the data set before calculating the contingency table. In the published figure the outliers are plotted with red Xs. As a result of excluding the outliers, the authors used values of A=0 and C=67 rather than A=5 and C=69 in calculating the Test Sensitivity. When the outliers are included in the calculation, the Test Sensitivity is 93.2%, which fails the ANSI performance criteria.

Our disagreement is not with the specific statistical method that was applied to identify outliers but rather with the automatic, uncritical elimination of data points that were deemed to be outliers. While the statistical test revealed that the data pairs were outliers, they do not reveal why. The authors infer from the statistical test that the outliers "originated from instrument errors or sampling bias"; however they offer no evidence to support those diagnoses. Outliers are a signal that something unusual has happened, e.g., equipment malfunction, data recording error, failure to follow the experimental protocol, etc. Statistical textbooks (e.g., Rawlings, 1988) counsel investigating the cause of outliers before eliminating them. It is not apparent from the article that the authors made such an investigation before eliminating the outliers. The charts in Figure 3 of the 2017 publication compare simultaneous readings from two different PortaCounts®. In each case, the regressions show excellent agreement between the two instruments. So, if an error occurred with one of the instruments during fit testing, then it was probably with the MT-05U. When a person is conducting a fit test in the field with that instrument, how will they know when an instrument error occurred and the fit test should be deemed invalid?

While the ANSI protocol does not explicitly address the handling of outliers, it does address the issue indirectly. As noted by the authors, the ANSI protocol establishes the requirement: "Any reference fit factor below 10% of the required fit factor accepted by the new fit test method shall disqualify the method." Presumably a statistical test would identify the point (FF_{Ref}<0.1 x RFF, FF_{New}>RFF) as an outlier; however, the ANSI method does not permit exclusion of that outlier. Rather, if such a point exists in the data set then the new method is rejected. If that point can't be excluded from the data set, then less obvious outliers such as those identified in Figure 5C, should not be excluded either.

In general, the simultaneous sampling method employed in these studies is preferable to the sequential sampling method specified in the ANSI protocol. It offers the key advantage that the results obtained from the two instruments are from the same donning of the respirator at the same time. However, the fact that the two instruments were measuring the same aerosol sample during a fit test makes it difficult to see how an outlier can be attributed to sampling bias. In addition, the authors reported that the instruments passed the daily check procedures, so unless some error or irregularity was noted by the test operator at the time the test was conducted (e.g., if a sample hose dislodged from the respirator during the test or if the hose was kinked), there's no reason to believe that the results of that test are not valid. That is, there's no reason to believe that the results are not indicative of the normal instrument performance.

By comparison, Richardson, et al. (2013) conducted sequential fit tests. To ensure that the fit of the respirator did not change from one fit test to another, they adopted a screening technique that involved measuring the concentration of particles inside the respirator for 5 seconds during normal breathing before the first fit test, between the first and second fit test, and after the second fit test. If this "baseline fit" changed by more than a factor of 100 from the start of a given test to the end, then it was assumed that the fit of the respirator had shifted over the course of the test, so the assumption of a consistent fit during the two tests was not valid. In that case the data was not used in the analysis. Insofar as the goal of the ANSI protocol is to compare two test methods *for the same respirator fit*, the screening method was a reasonable approach to validate that assumption.

It is worth noting that the Sibata MT-05U is a count-based instrument. The authors may find it useful to define an acceptable margin of error for a fit factor of FF_{New}=500 (e.g., ±5%) and then work backwards to establish a minimum required chamber concentration $C_{out,min}$ to achieve that margin of error, as discussed by TSI (2018). This will permit a *post hoc* analysis of the data, eliminating values of FF_{New} for which $C_{out} < C_{out,min}$, which may, in turn, eliminate some of the data points that were identified as statistical outliers. This approach would be justifiable because it is directly traceable to the raw data of the measurement and the variability associated with low particle counts.

In summary, the studies do not conform to the ANSI validation protocol because they did not use the generated aerosol method as the reference method. The published studies do provide interesting comparisons between the MT-05U and the PortaCount® (with and without the N95 Companion). We commend the authors for providing a complete accounting of the data that they collected and the logic they used in their analysis. Reanalysis of the data taking into the consideration the variability associated with count statistics may provide a legitimate method to identify and eliminate outliers.

Sincerely,

Andrew Viner
Craig Colton, CIH
Ginny Agresti
3M Personal Safety Division

REFERENCES

American National Standards Institute. (ANSI) (2010) Respirator Fit Testing Methods. American National Standards Institute /American Industrial Hygiene Association, Fairfax, VA. ANSI Z88.10-2010.

Han D-H, Willeke K, and Colton CE. (1997) Quantitative Fit Testing Techniques and Regulations for Tight-Fitting Respirators: Current Methods Measuring Aerosol or Air Leakage, and New Developments. Am. Ind. Hyg. J., 58(3): 219-228

OSHA (1998) 29 CFR 1910.134, Appendix A.

Rawlings, JO (1988) Applied Regression Analysis: A Research Tool. 1st Ed., Wadsworth, Inc.

Richardson AW, Hofacre KC, Weed J, Holm R and Remiarz R. (2013) Evaluation of a Faster Fit Testing Method for Full-Facepiece Respirators Based on the TSI PortaCount®. J. Int. Soc. Resp. Prot. 30: 116–128.

TSI (2018) An Analysis of Utilizing Condensation Nuclei Counter (CNC) vs. Optical Particle Counter (OPC), Application Note RFT-023. (http://www.tsi.com/uploadedFiles/_Site_Root/Products/Literature/Application_Notes/RFT-023_QRFT_CNC_vs_OPC_US-web.pdf)

Wu B, Leppänen M, Yermakov M and Grinshpun SA. (2017) Evaluation of a new instrument for aerosol quantitative fit testing. J. Int. Soc. Res. Prot. 34(2): 111-127.

Wu B, Leppänen M, Yermakov M, and Grinshpun SA. (2018) Utility of an Optical Particle Counting Instrument for Quantitative Respirator Fit Testing with N95 Filtering Facepieces. J. Int. Soc. Resp. Prot. 35(1):14-25.

Response to Letter to the Editor

RE: Papers “Evaluation of a New Instrument for Aerosol Quantitative Fit Testing” and “Utility of an Optical Particle Counting Instrument for Quantitative Respirator Fit Testing with N95 Filtering Facepieces”

Bingbing Wu[1*], Maija Leppänen[2], Michael Yermakov[1], and Sergey A. Grinshpun[1#]

[1] Center for Health-Related Aerosol Studies, Department of Environmental Health, University of Cincinnati, PO Box 670056, Cincinnati, OH 45267-0056, USA
[2] University of Eastern Finland, Department of Environmental and Biological Sciences, PO Box 1627, 70211 Kuopio, Finland

#Author whom correspondence should be addressed to.
Tel: +1-513-558-0504; fax: +1- 513-558-2263; e-mail: sergey.grinshpun@uc.edu

SUBMITTED TO *Journal of International Society for Respiratory Protection*
November 16, 2018

We appreciate the opportunity to respond to the Letter to the Editor, which was submitted by Andrew Viner, Craig Colton, and Ginny Agresti of the 3M Personal Safety Division regarding our two papers, Wu et al. (2017) and Wu et al. (2018), recently published in the *Journal of ISRP*. The referred papers present our studies on the evaluation of a new instrument for respirator fit testing, MT-05U. We are grateful to the authors of this Letter for their interest in our research and for taking the time to share their thoughts. Our responses are presented below.

The main concerned voiced in the Letter is that our studies might have reached an incorrect conclusion at least in one case. The authors seem to attribute it to our possible misunderstanding of the requirements of the evaluation protocol specified in ANSI Z88.10-2010, Annex A2 (ANSI, 2010) as well as to an alleged “weakness” of the ANSI standard.

First, the Letter questioned whether an appropriate reference fit test method (required by the ANSI standard) was used in our evaluation of the new MT-05U. We fully acknowledge that the generated aerosol method described in Clause 7.1 “Generated Aerosol Quantitative Fit Test Procedure” (QNFT) of the ANSI standard is the one to be used, and indeed implemented this very method in our studies. Our investigation was designed as a comparative study involving the new OPC-based fit testing instrument and the reference instrument (PortaCount®, TSI Inc., Shoreview, MN). The latter is overwhelmingly used in the field. The PortaCount® design allows using not only generated aerosol but also the ambient aerosol as the challenge. The PortaCount® fits in the Clause 7.2 (“Particle-counting Instrument Quantitative Fit Test Procedure”) of the ANSI standard that suggests utilizing the ambient aerosol. It is to note that in our studies the PortaCount® served as the “aerosol detection system” and “device for recording fit test results” as listed in the “Equipment” section of Clause 7.1. The procedures required by Clause 7.1 for the generated aerosol QNFT were carefully followed in our study design. Finally, the concentration of aerosol generated in our tests (4,000 – 16,000 particles/cm^3) was always within the operational concentration range of the PortaCount®: 0.01 – 25,000 particles/cm^3 (TSI Inc., 2015).

The Letter points out to a “weakness” in the ANSI standard associated with the definition of the aerosol detection system in the generated aerosol method. Apparently, the authors of the Letter would like to see a standard specifying a single detection method, e.g., photometry. First of all, even if the ANSI standard had some shortcomings or “weaknesses”, this is the only one providing guidelines for evaluating a new fit testing method. In absence of alternatives, it was utilized in our studies. We acknowledge that ANSI (2010) does not define a specific aerosol detection system. Typically, it is a flame or forward light scattering photometry utilized in the frameworks of the generated aerosol method, described in Han, Willeke, and Colton (1997) as well as in the International Organization for Standardization (ISO) standard 16975-3 (ISO, 2017). We believe that it is an advantage of the ANSI standard, not its “weakness,” that it is not limited to a specific aerosol detection system. This leaves a niche for measurement devices other than photometers to be deployed for respirator fit testing.

Second, the Letter implies that the treatment of outliers in our data analysis was not sufficiently justified, which could have affected the final conclusion of the study. It is stated in the Letter that there was no sufficient explanation as to how the outliers were identified and what the causes of the outliers are. We agree that eliminating an outlier requires a proper explanation, and should probably have provided more details in our papers. We are using this Response as an opportunity to elucidate the justification of the outlier treatment. The Letter specifically questioned the five outliers marked in zone A, Figure 5C of Wu et al. (2017). In these, the reference fit factor (FF_{ref}) was below the required fit factor (RFF) while the new method produced FF_{new} in excess of RFF. These data points were indeed critical to the validation of the new method because if they were not eliminated but included in calculations, the comparison statistics sensitivity would have dropped to 93.2%, which is below the mandatory requirement of ≥ 95%.

It was observed that the five points, which later were identified as outliers, presented a subject-specific pattern. Having come from two subjects, three from X and two from Y, these five data points formed two clusters shown in the graph (Figure 5C of Wu et al., 2017). The exercise-specific and overall FF-values representing these data points are listed in Table I below. In all cases obtained with these two subjects, the exercise-specific FF values recorded by the MT-05U were considerably greater than those recorded by the PortaCount®. It was reported in Wu et al. (2017) that when testing these individuals, we observed an excessive amount of water condensed in sampling lines. This must have resulted from the high humidity of air exhaled by these subjects. The block or kink in a sampling tube impedes the normal performance of the instrument, and the MT-05U was found to be particularly sensitive to the humidity effect, which made the MT-measured FF values higher. One value, 14,657, generated during exercise 3 by subject X, was even above the upper operating limit of MT-05U (10,000).

The Letter also made an inference that if there was any instrument error, it would probably come from MT-05U. This judgement was based on the readings from the two identical PortaCounts® that were tested in parallel that showed an excellent agreement (Figure 3 of Wu et al., 2017). We are hesitant to come to this conclusion because the sample sizes were fairly different. For the comparison of two PortaCounts®, the sample size was 22 for half facepiece and full facepiece respirators and 24 for P100 FFR; for the comparison of MT-05U versus PortaCount®, the sample size was 122 for half facepiece, 135 for full facepiece, and 148 for P100 FFR. In other words, while no direct instrumental error was discovered in the course of the tests involving two PortaCounts®, can one be assured that, likewise, no instrumental error occurred also in a much larger set of experiments?

The Letter referred to the screening approach for validating a consistent fit during the two tests (Richardson et al., 2013). While we acknowledge the features of the approach and its ability to check whether the respirator fit changed between sequential tests involving different fit testers (each requiring a separate donning), we do not believe that this approach is applicable to our study design. The reason is that, unlike the sequential fit testing conducted by Richardson et al. (2013), we performed the simultaneous measurement with two fit testing instruments operating in parallel (during the same donning). This design eliminates the need in “screening.”

Table I. Original Fit Factor Data Representing the Outliers in Zone A, Figure 5C of Wu et al. (2017)

Subject code	Instrument	Fit factor							
		Exercise 1	Exercise 2	Exercise 3	Exercise 4	Exercise 5	Exercise 7	Exercise 8	Overall
X	PortaCount®	368	623	462	387	215	287	509	365
X	MT-05U	1335	4108	14657	2727	595	1560	3386	1731
X	PortaCount®	291	359	341	384	258	474	606	361
X	MT-05U	1180	9542	4899	929	538	1691	7838	1455
X	PortaCount®	189	312	335	347	248	572	518	317
X	MT-05U	855	3993	3904	1301	880	1858	1638	1480
Y	PortaCount®	137	200	122	125	184	100	73	121
Y	MT-05U	602	778	507	443	738	596	485	570
Y	PortaCount®	86	223	110	109	168	150	129	128
Y	MT-05U	539	802	310	403	688	756	525	519

The authors of the Letter were not sure whether we were able to disqualify the new method when a statistical test identified an outlier point in case

$$FF_{Ref} < 0.1 \times RFF \text{ and } FF_{New} > RFF \quad (1)$$

following ANSI (2010). This concern is probably rooted in a lack of understanding of our protocol with respect to the data collection order (Figure 1 of Wu et al., 2017). In our studies, the data analysis always started with checking for conditions described in Eq. (1) (see Criterion 1 of the protocol presented in Figure 1 of Wu et al., 2017). Therefore, we are confident that our study design provided a full capability to disqualify the new method in case the data would fit Eq. (1).

Last, the Letter suggested us to define an acceptable margin of error for a fit factor of $FF_{New} = 500$ (e.g., ± 5%) and then work backward to establish a minimum required chamber concentration $C_{out.}$ to achieve that margin of error, as discussed by TSI Inc. (2018). We appreciate the suggestion, but in our experiments, the aerosol concentration C_{out} was sufficiently high to assure a reasonable margin of error. The calculation in the referred TSI document was conducted for a very low level of C_{out} = 3 particles/cm^3 while in our tests involving MT-05U the C_{out} level was much greater: 60 –760 particles/cm^3 in Wu et al. (2017) and 122 – 528 particles/cm^3 in Wu et al. (2018). We determined that even in the worst-case scenario the margin of error generated in our studies was within ± 8.45%, which is rather low. Below, we present the calculation that yields this value.

In our calculation, we conservatively used the lowest concentration measured in the fit tests performed with full facepiece respirators (C_{out} = 60 particles/cm^3). For this type, the fit factor required for passing the fit test is RFF = 500 according to 29 CFR 1910.134 (OSHA, 1999). Consequently, $C_{in} = C_{out}/500 = 0.12$ particles/cm^3. The margin of error for counting particles in the worst-case scenario at the 95% confidence level can be calculated according to the TSI Inc. (2018) algorithm as follows:

Total Particles (n) = (C_{in}) × (flow rate) × (time conversion) × (time of exercise) = (0.12 particles/cm^3) × (1000 cm^3/ min) × (1 min / 60 sec) × (70 seconds total mask sample) = 140 particles. Note that the sampling flow rate of MT-05U is 1000 cm^3/min. In the worst-case scenario, the margin of error at 95% confidence level is

$$\pm\frac{1}{\sqrt{n}} = \pm\frac{1}{\sqrt{140}} = \pm 8.45\%$$

In case someone wants to achieve a lower margin of error, e.g., 5%, the minimum ambient aerosol concentration should be 171 particles/cm^3. By reviewing the data obtained by Wu et al. (2017), we found that the ambient concentration fell below 171 particles/cm^3 only on three occasions when the following paired fit factor values (FF_{Ref}, FF_{New}) were generated: (371, 142), (215, 81), and (2513, 1248). However, none of these pairs was identified as statistical outliers, contrary to the expectation articulated in the Letter.

To summarize, the studies of Wu et al. (2017) and Wu et al. (2018) did utilize the generated aerosol method as the reference, which we believe is appropriate. The outlier analysis included two steps: statistical screening and investigation of the causative factors. The testing was conducted in accordance with the ANSI (2010) standard for the evaluation of new methods.

We hope that this Response contains appropriate information to essentially eliminate the concerns expressed in the Letter. We sincerely appreciate this opportunity to discuss our findings and trust that this exchange will bring additional attention to the development and evaluation of alternative respirator fit testing methods.

REFERENCES

American National Standards Institute. (ANSI) (2010) Respirator Fit Testing Methods. American National Standards Institute /American Industrial Hygiene Association, Fairfax, VA. ANSI Z88.10-2010.

Federal Register. (2016) Additional PortaCount® Quantitative Fit-Testing Protocols: Amendment to Respiratory Protection Standard (Proposed Rules and Notice of Public Hearing). Federal Register 81:195, 69740 (October 7, 2016).

Han D-H, Willeke K, and Colton CE. (1997) Quantitative Fit Testing Techniques and Regulations for Tight-Fitting Respirators: Current Methods Measuring Aerosol or Air Leakage, and New Developments. American Industrial Hygiene Journal 58(3): 219-228.

International Organizational for Standardization. (ISO) (2017) Respiratory protective devices -- Selection, use and maintenance -- Part 3: Fit-testing procedures. ISO, Geneva, Switzerland. ISO 16975-3:2017

OSHA. (1999) Respiratory Protection. Title 29 Code of Federal Regulation, Part 1910.134. U.S. Retrieved October 27, 2018, from https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=9780.

Richardson AW, Hofacre KC, Weed J, Holm R and Remiarz R. (2013) Evaluation of a Faster Fit Testing Method for Full-Facepiece Respirators Based on the TSI PortaCount. Journal of the International Society for Respiratory Protection 30: 116–128.

TSI Inc. (2015) PortaCount® Pro 8030 and PortaCount® Pro+ 8038 Respirator Fit Testers. Operation and Service Manual. P/N 6001868, Revision P, March 2015. TSI Inc., Shoreview, MN.

TSI Inc. (2018) An Analysis of Utilizing Condensation Nuclei Counter (CNC) vs. Optical Particle Counter (OPC). Application Note RFT-023. Retrieved October 27, 2018, from http://www.tsi.com/uploadedFiles/_Site_Root/Products/Literature/Application_Notes/RFT-023_QRFT_CNC_vs_OPC_US-web.pdf

Wu B, Leppänen M, Yermakov M and Grinshpun SA. (2017) Evaluation of a New Instrument for Aerosol Quantitative Fit Testing. Journal of the International Society for Respiratory Protection 34(2): 111-127.

Wu B, Leppänen M, Yermakov M, and Grinshpun SA. (2018) Utility of an Optical Particle Counting Instrument for Quantitative Respirator Fit Testing with N95 Filtering Facepieces. Journal of the International Society for Respiratory Protection 35(1): 14-25.