

Recommendations for the Acceptance Criteria for New Fit Test Methods

Thomas J. Nelson¹ and Haskell E. Mullins²

¹ NIHS Inc., 2401 East Mall, Ardentown DE 19810, USA
Email: tjnelson1@comcast.net

² 3M Occupational Health and Environmental Safety Division, 3M Center, Building 235, St. Paul, MN 55144-1000, USA

ABSTRACT

The American National Standards Institute (ANSI) standard on fit testing, Z88.10 includes a provision for the acceptance of new fit test methods. The new test method is compared to one of the ANSI Z88.10 accepted quantitative fit tests (QNFT). To determine if the underlying distribution of fit factors affects the Z88.10 sensitivity calculation, four different possible fit factor distributions were examined. Pass/fail rates were estimated based on a logistic regression analysis of actual fit test performance. When a simple sensitivity test is used as recommended by the ANSI standard, the apparent performance of a new fit test can be affected by the distribution of data. It is proposed that a binary logistic regression analysis be used to determine the probability of passing with a new fit test method. An acceptable fit test method should have a low probability of passing the test with low reference method fit factors. Passing criteria with probabilities of less than 0.2 at a fit factor of 50 and less than 0.5 at a fit factor of 100 are recommended by the authors. These criteria are consistent with the current Z88.10 standard's accepted fit test methods, in that the irritant smoke test does not meet the criteria while the other tests are found to be acceptable.

Keywords: Fit Test, Fit Test Evaluation

INTRODUCTION

Fit tests are used to help select a specific respirator for a person. If a person passes a fit test, they are assumed to have a respirator that "fits". It is assumed that subsequent respirator use will provide a level of protection equal or greater than the assigned protection factor.

A number of fit test methods are accepted (CFR, 2003b; American National Standards Institute, 2001). They include both qualitative and quantitative types. Since we fit test once a year with a single test, it is important that the accepted method reject inadequately fitting respirators.

The American National Standards Institute (ANSI) standard on fit testing, Z88.10 includes a provision for the acceptance of new fit test methods (American National Standards Institute, 2001). The result of the new test method (pass/fail) is compared to pass/fail results of one of the accepted quantitative fit tests (QNFT). A minimum of 100 sequential paired tests from at least 25 different subjects must be conducted. At least 50 tests must have reference method fit factors less than the required fit factor. The required fit factor is defined as the numeric value that is the pass criterion. For a half facepiece respirator it is generally set at a value of 100 (CFR, 2003b). A cumulative distribution plot or histogram is constructed to visually confirm that the fit factors obtained from the reference method bracket the required fit factor.

Test statistics are calculated from the results of paired fit tests. The new method must demonstrate a sensitivity of 0.95 (equivalent to a beta (β) statistic of 0.05). Test sensitivity is the proportion of the reference method fit tests with fit factors less than the required fit factor that also failed the new fit test method. In the two by two contingency table shown in Table I, the sensitivity is equal to block C divided by the sum of blocks A and C.

Table I. Two by Two Contingency Table

Result	Failed Reference Test (Fit Factor < Required Fit Factor)	Passed Reference Test (Fit Factor > Required Fit Factor)
Passed New Test	A	B
Failed New Test	C	D

The remaining statistics listed in the standard are the predictive value of a pass, test specificity or alpha (α) statistic, predictive value of a fail and kappa statistic. They are to be reported for comparison purposes.

The test specificity measures the fraction of those that pass reference fit test that also pass the new fit test. The fraction should be 0.50 or greater to ensure that most wearers with acceptable fit tests will pass the new test. It is calculated by dividing the number in block B by the sum of the numbers in blocks B and D in Table I.

Details on calculating the predictive value of a pass, predictive value of a fail and kappa statistic are found in the ANSI Z88.10 standard.

Janssen et al. have shown that the choice of the reference fit test method can have an apparent effect on how a candidate test appears to perform (Janssen et al., 2002). The ambient aerosol, controlled negative pressure and bitter aerosol fit test methods were each used as the reference method to determine how the other two test methods performed. The sensitivity of each method was calculated using each of the other two methods as the reference. None of the test methods met the ANSI sensitivity criterion of 0.95 or greater when compared with either of the other two methods. Janssen concluded that fundamental differences exist among the fit test methods evaluated and it was not surprising that they did not agree with one another in all cases. It should be pointed out that each of the three methods did meet the sensitivity criterion during their validation.

It has been assumed by some researchers that the sensitivity statistic is applied uniformly across a range of fit factors. Campbell used a mathematical model to predict the percentage of people who would not achieve a level of respirator performance consistent with the assigned protection factor (Campbell et al., 2001). Respirator fitting characteristics and theoretical fit test performance as defined by the α and β statistics formed the basis for the calculation. The results of the model showed that some respirators would not perform as well as others. Contrary to experience and common expectations, Campbell's model has a higher proportion of people tending to pass the fit test as the number of people with inadequate respirator fits increases.

A casual review of fit test data from studies that compare a new fit test method to an existing method refute the concept that the α and β statistics are uniform across all fit factors. Marsh reported pass/fail results for an irritant smoke test in terms of ranges of generated aerosol reference fit factors as shown in Table II (Marsh, 1984). The percent of irritant smoke test failures in each range of generated aerosol fit factors decreased as the reference fit factor increased.

The purpose of this analysis is to examine the influence that the distribution of the reference method fit factors has on the acceptability of a fit test method. If the distribution has an effect on the statistics that are calculated, a more robust method of analyzing data that is not affected by the distribution of fit factors should be used to judge how well a fit test method performs.

Table II. Fit Test Result Versus Range of Reference Fit Factors (Marsh Irritant Smoke Data)

Result	Reference Fit Factor Range			
	<10	10-100	101-500	>500
Passed Irritant Smoke Test	0	4	9	55
Failed Irritant Smoke Test	38	66	13	15
Percent Failing of Total in Fit Factor Range	100%	94%	59%	21%

METHODS

Five fit test methods were evaluated, the American Iron and Steel Institute (AISI) and Marsh irritant smoke, the Bitrex, Ambient Counting and Controlled Negative Pressure Fit Tests. The AISI irritant smoke fit test data were obtained by comparing the result of an irritant smoke qualitative fit test with a generated aerosol QNFT (American Iron and Steel Institute, 1981). In this test, a person's ability to respond to the irritating properties of a low concentration of the stannic chloride used in the smoke tube is first demonstrated. Then while wearing a respirator the person is exposed to a higher concentration of the irritating smoke. In the AISI test, passing a 200 ml flow of air through the tube generated the irritant smoke. If no irritation is noted, the person passes the test.

The Marsh irritant smoke test data were obtained by comparing the result of the irritant smoke test to a generated aerosol QNFT (Los Alamos National Laboratory, 1983). The irritant smoke was generated using a 50 ml rubber squeeze bulb. At the start of the test, the bulb was held at a distance of approximately one meter from the person and was moved closer until the tube was approximately 15 cm. from the person's face.

The Bitrex, Ambient Particle Counting and Controlled Negative Pressure (CNP) fit test data were compared to generated aerosol QNFT reference data (Mullins et al., 1995; da Roza et al., 1991; Crutchfield et al., 1994). The procedures for the fit tests are essentially those published in the Occupational Safety and Health Administration (OSHA) respirator standard (CFR, 2003b). In each of the comparison tests six exercises were used rather than the eight listed in the OSHA standard. The bending over and grimace exercises were not performed.

The ANSI Z88.10 standard simply states that the fit factors obtained from the reference method bracket the required fit factor. Figures 1-3 show possible hypothetical distributions of fit factors for half facepiece respirators that could be obtained in a study of a new fit test method. Figure 1 has the fit factors uniformly spread across the range. Figure 2 shows typical distributions of fit factors. (Nelson et al., 1984; Mullins et al., 1995). Distribution 1 has a mode at a fit factor of 20, distribution 2 has a mode at 120. Both distributions are approximately lognormal. Figure 3 represents a distribution that could occur if a leak was added to good fitting facepieces to arrive at a narrow range of fit factors near the passing value. In each case the fit factors bracket the required fit factor of 100 but the distributions are very different.

This relationship between increasing reference method fit factors and the probability of failing new method fit tests can be quantified with a regression equation. Since the outcome of a fit test is pass/fail, a binary logistic regression analysis can be used to assess the probability of passing the new fit test as a function of the reference method fit factor. The regression analysis function in Minitab was used (Minitab Inc., 2000). The reference fit factor is compared to the outcome of the fit test under evaluation (pass/fail).

For each distribution, the best fitting line for the binary regression analysis (Table III) was used to determine the number of passes for each range of fit factors shown in Figures 1-3. The center point of the histogram bin was used; the theoretical number of passes was then calculated based on the probability of passing at that fit factor. For example, for the AISI fit test, the bin range 17-25 had 10 people. The mid point of the bin is 21, the probability of passing the AISI fit test with a fit factor of 21 from

the equation is 0.541. Thus we would expect 5 ($10 * 0.541$) people to pass the fit test and 5 people (10-5) to fail. The calculated number of passes and fails was determined for each bin range, the sum of the passes and fails for people with fit factors less than and greater than 100 were then determined.

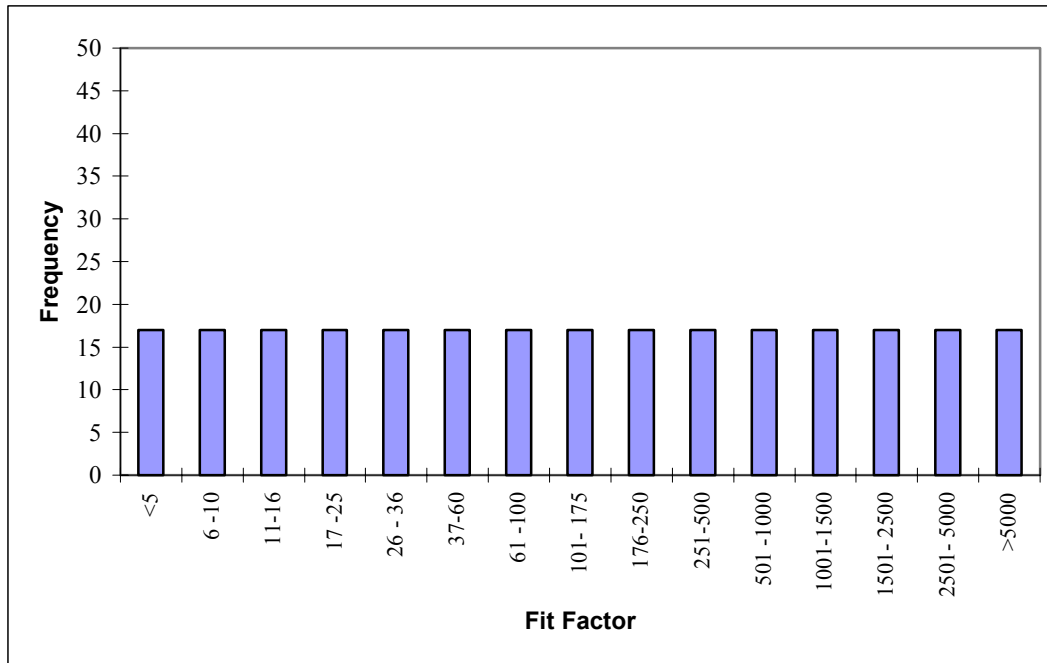


Figure 1. Histogram, even distribution of fit factors.

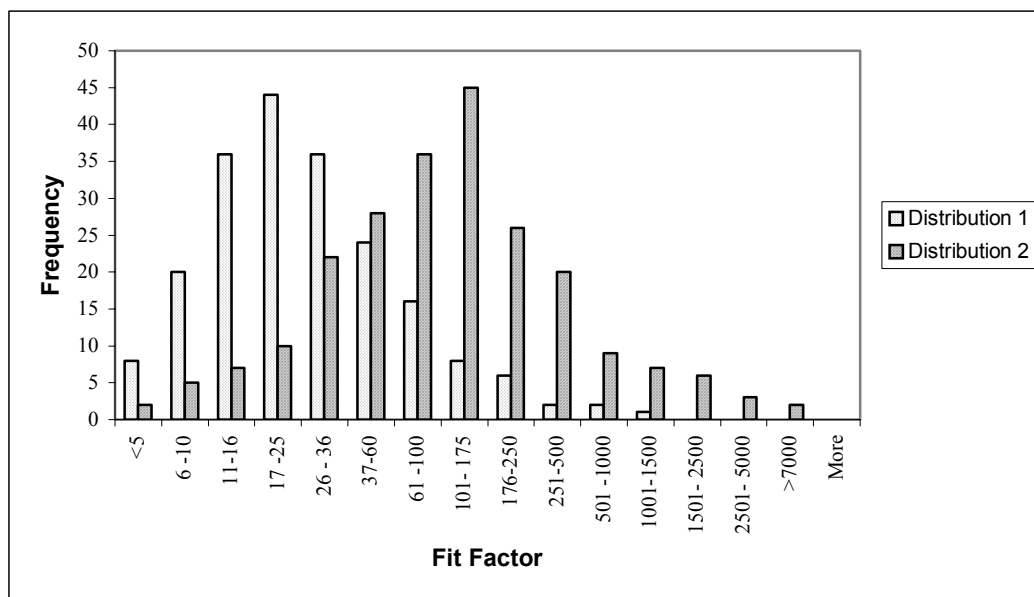


Figure 2. Histogram, lognormal distributions of fit factors.

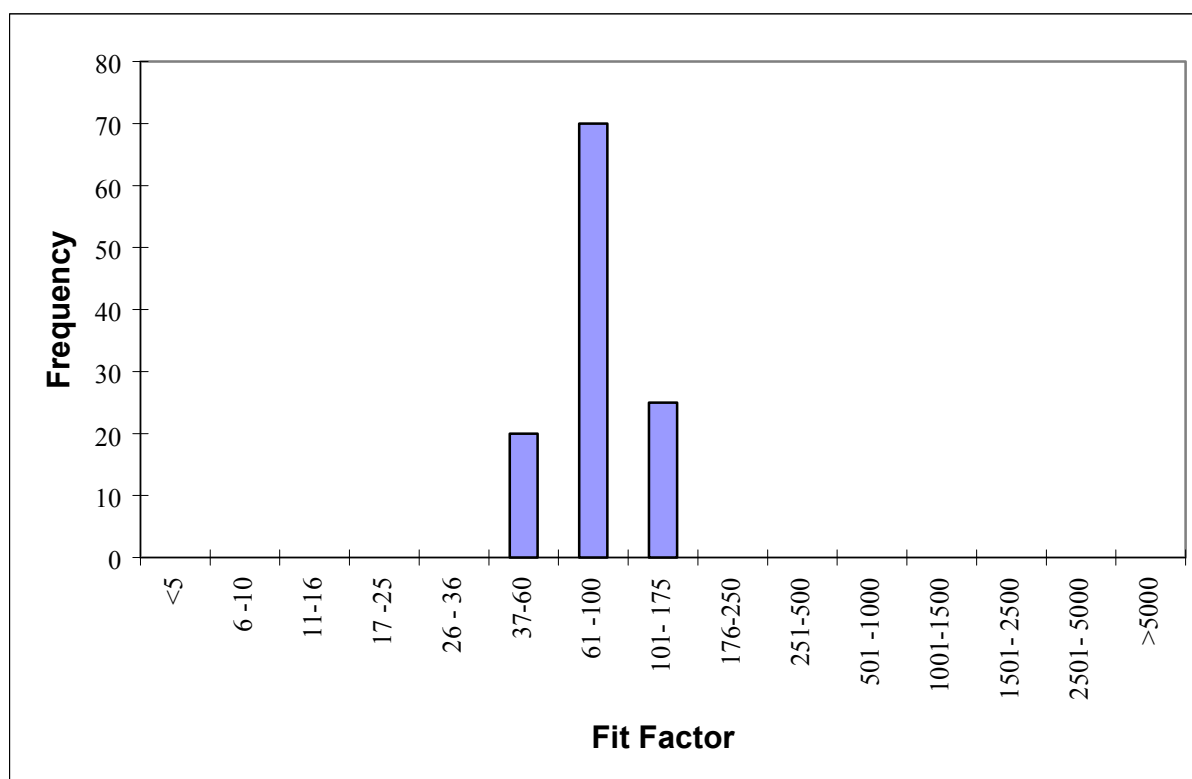


Figure 3. Histogram, distribution of fit factors with induced leak.

The sensitivity and specificity for each fit test method was calculated from these data for each of the four fit factor distributions. Specificity is defined in the ANSI Z88.10 standard as the value in block B divided by the sum of blocks B and D in Table I.

Table III. Best Fit Regression Equations

Fit Test Method	Best Fitting Regression Equation	
	$Y = 1 / \{1 + [\text{EXP}(-A - (B * X))] \}$ where Y = probability of passing and X = Log10(Fit Factor)	
	A	B
AISI Irritant Smoke	-4.538	3.566
Marsh Irritant Smoke	-4.202	1.429
Bitrex	-12.342	5.792
Ambient Particle Counting	-8.143	3.898
Controlled Negative Pressure	-18.553	7.133

RESULTS AND DISCUSSION

A summary of the statistical properties for each fit test method is shown in Table IV. The best fitting regression line for each of the five fit tests is shown in Figure 4. The calculated test sensitivity and specificity for each distribution and fit test are shown in Tables V and VI.

Table IV. Binary Probabilistic Regression Analysis of Selected Fit Test Method Validation Studies

Fit Test Method	Number of Comparative Fit Tests	Probability of Passing at Indicated Reference Fit Factor		
		10	50	100
AlSI Irritant Smoke	110	0.27	0.82	0.93
Marsh Irritant Smoke	199	0.06	0.15	0.21
Bitrex	150	0.0	0.08	0.32
Ambient Particle Counting	26	0.01	0.18	0.41
Controlled Negative Pressure	100	0.0	0.00	0.01

Table V. Test Sensitivity for Different Fit Factor Distributions Based on Binary Regression Analysis

Fit test method	Distribution type			
	Even	Lognormal 1	Lognormal 2	Induced leak
AlSI Irritant Smoke	0.49	0.46	0.28	0.12
Marsh Irritant Smoke	0.90	0.90	0.87	0.83
Bitrex	0.96	0.97	0.91	0.82
Ambient Particle Counter	0.95	0.92	0.91	0.72
Controlled Negative Pressure	1.0	1.0	1.0	1.0

Table VI. Test Specificity for Different Fit Factor Distributions Based on Binary Regression Analysis

Fit test method	Distribution type			
	Even	Lognormal 1	Lognormal 2	Induced leak
AlSI Irritant Smoke	0.99	0.97	0.97	0.96
Marsh Irritant Smoke	0.49	0.31	0.36	0.24
Bitrex	0.90	0.71	0.75	0.52
Ambient Particle Counter	0.79	0.71	0.58	0.56
Controlled Negative Pressure	0.69	0.25	0.34	0.04

The values for test sensitivity vary by distribution type for each fit test method except CNP. The CNP test has a sensitivity of 1.0 for each distribution. Test sensitivity for the induced leak distribution has the lowest value for the remaining fit test data sets. Test specificity varies by distribution for each fit test method. The induced leak distribution has the lowest value compared to the other distributions.

Test sensitivity measures how well the candidate fit test method agrees with the reference fit test method. It is a measure of safety of the new test method. Fit test methods with a high value for sensitivity will not allow people with reference fit factors below the required fit factor to pass. Test

specificity is a measure of a possible economic burden associated with the new test method. Lower values of specificity mean that fit tests with reference fit factors greater than the required fit factor will be failed by the new fit test method. This forces the person to undergo a new fit test even though respirator fit would be deemed acceptable by the reference fit test method.

The binary logistic regression analysis shows that the probability of passing a fit test varies with reference fit factor value. The calculation of test sensitivity for various distributions of fit factors shows that the calculated sensitivity value can vary as a function of the distribution of reference fit factors. A simple sensitivity calculation may generate different results from different studies using the same reference fit test method. This is not a desirable artifact of an experiment aimed at determining the performance of a fit test method.

When an evaluation of a fit test method is performed, it is difficult to generate a specific number of fit factors in specific ranges (Janssen et al., 2002). Attempts are made to provide people with incorrect sizes of respirators to achieve lower fit factors. A test method that is evaluated with an even distribution of reference fit factors will tend to have a higher value for sensitivity. In this distribution, more reference fit factors are very low, increasing the probability that a candidate test will correctly identify them as unacceptable.

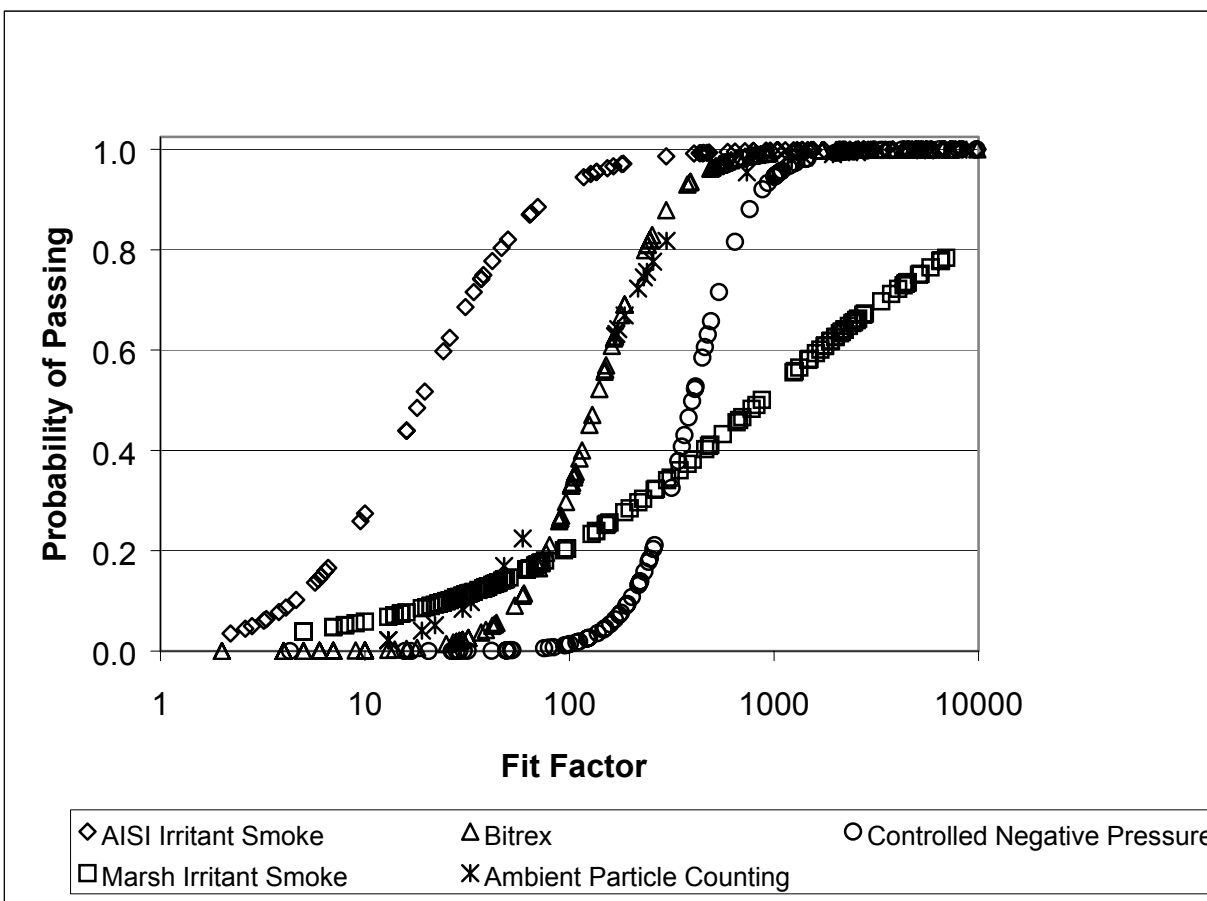


Figure 4. Probability of passing a fit test versus fit factor.

Introducing a leak that groups fit factors near the required fit factor would be expected to show the smallest value for sensitivity. For example, with the ambient aerosol counting fit test method for fit

factors from 60 to 100 the probability of passing is 0.22 to 0.37. If most of the fit factor data from an evaluation is in this range, the resulting sensitivity will be low.

The required fit factor for half or full facepiece respirators has varied over time. For half facepiece respirators, it has varied from 10 to 100. In the late 1970's, OSHA standards provided little information on how to conduct a quantitative fit test. The acrylonitrile standard simply states that fit testing results in the respirator achieving the assigned protection factor listed in the standard (CFR, 2003a). This implies that a quantitative fit test only be required to show a fit factor equal to the assigned protection factor.

In his evaluation of the irritant smoke test Marsh defined inadequate fit as a fit factor less than 10 for half facepiece respirators (Marsh, 1984). The 1980 ANSI Z88.2 standard allowed the results of QNFT to determine the assigned protection factor (American National Standards Institute, 1980). Three tests were required for each person. The lowest of the three fit factors became the individual's protection factor, with a limit of ten times the assigned protection factor for the class of respirator.

When OSHA began requiring QNFT in the substance specific health standards, OSHA inspectors were fit tested by personnel from the OSHA training institute (Colton, 2002). Employers could set any required fit factor for their facility and if an OSHA inspector did not meet this minimum requirement, the inspector could be refused entry. The training institute set a required fit factor of 100 for half facepieces (1000 for full facepieces) since this was consistent with the highest required fit factor that could be set in the ANSI Z88.2 standard (1980). The minimum passing fit factor became 100 for a half facepiece respirator in OSHA substance specific health standards and continues today as shown in the OSHA respiratory protection standard (CFR, 2003b).

Coffey determined the relationship between fit factor measurements and a biological index of exposure (Coffey et al., 1998). People were fit tested by various methods and wore a half facepiece respirator in a chamber containing Freon® 113 gas. After exposure, the exhaled breath was analyzed to determine their actual exposure. It was found that when the required fit factor was set at 50 or above, all subjects achieved protection above the assigned protection factor of ten.

Zhuang measured program protection factors (PPF) for half facepiece respirators in a steel mill to determine if the results correlated with fit factor measurements (Zhuang et al., 2003). Fit factors were measured before and after each PPF determination. Iron was the contaminant measured. Since the level of airborne iron was below the OSHA permissible exposure limit, workers with fit factors less than the required fit factor of 100 could participate. When the average of the pre- and post fit factors and the geometric mean of the PPFs were examined, the correlation is significant ($r^2 = 0.60$). The relationship is shown in Figure 5. Based on the regression analysis, an average fit factor of 30 will result in a geometric mean PPF of 50.

Nelson determined the relationship between average respirator performance and the possibility of being exposed above the occupational exposure limit (Nelson et al., 2000). Data from workplace studies were used to estimate the "worst-case" persons' respirator performance and monitoring data from the literature was used to estimate the likely range of workplace exposures. When geometric mean workplace protection factors (WPFs) are greater than 44, very few exposures would be greater than the exposure limit. These fit factor and respirator performance studies demonstrate that respirator workplace performance of 30 to 50 will result in low exposures.

CONCLUSIONS

ANSI Z88.10 recommends that the acceptability of a new fit test be determined by comparison with an existing quantitative fit test. When a simple sensitivity test is used to describe a new test, the result can be affected by the distribution of data. In several cases using the theoretical distributions described in this paper, the outcome of a sensitivity test for the Bitrex and Ambient Particle Counter fit tests could have failed to meet the ANSI Z88.10 sensitivity requirement. The method used to determine acceptability should be independent of specific data collected. It is proposed by the authors that a binary logistic regression analysis be used to determine the probability of passing with a new test.

To generate the data, the generated aerosol fit test should be used as the "gold" standard (Janssen et al., 2002). Fit factors should be distributed over a range from below to well above the required fit factor. No attempt should be made to introduce artificial leaks since the resulting fit factors do not represent actual fit factors that would occur.

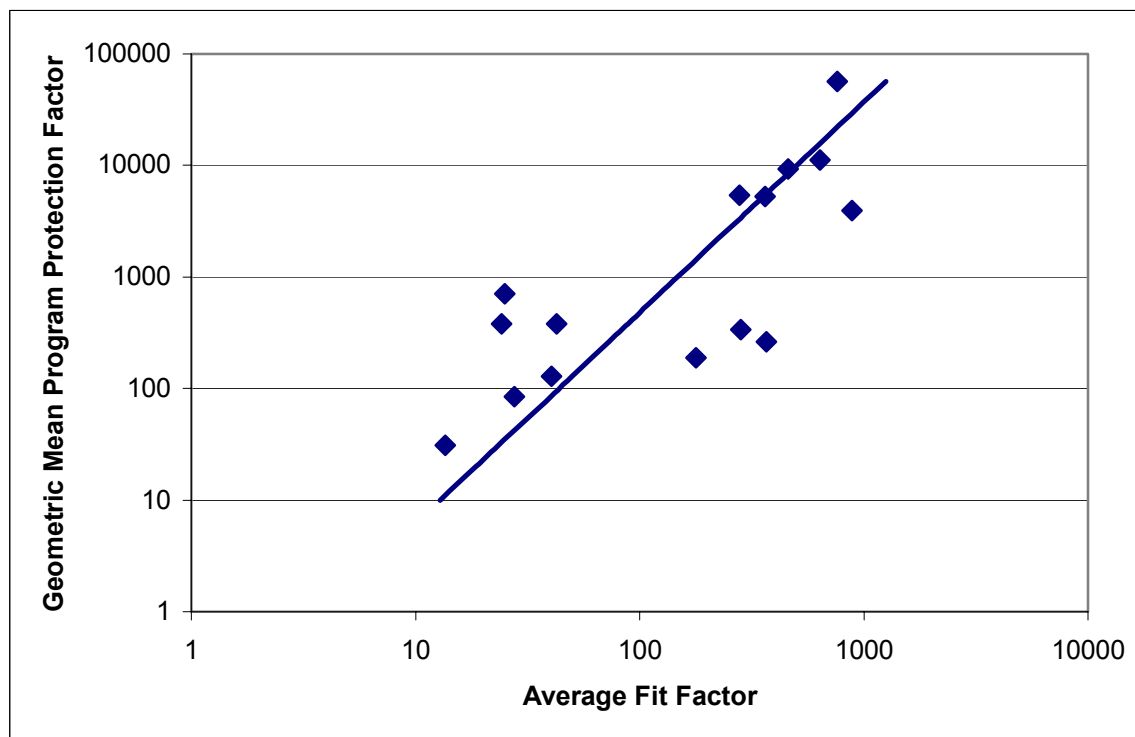


Figure 5. Average Fit Factor versus geometric mean program protection factor.

Published data indicate that fit tests that reject fit factors less than 30-50 result in adequate respirator performance. An acceptable fit test should have a low probability of passing the test with low fit factors. Passing criteria with probabilities of less than 0.2 at a fit factor of 50 and less than 0.5 at a fit factor of 100 are recommended by the authors. The outcome using these criteria are consistent with the current Z88.10 standard's recommended fit tests, in that the AISI irritant smoke test does not meet the criteria while the other tests are found to be acceptable.

Acknowledgements

The authors wish to thank Cliff Crutchfield and Ziqing Zhuang for providing access to the data from their studies.

REFERENCES

American Iron and Steel Institute. (1981) *Respirator Fit Testing, A Qualitative Protocol using Irritant Smoke. (Unpublished Information)* American Iron and Steel Institute; 1101 17th Street NW #1300; Washington DC 20036.

- American National Standards Institute.(1980) *American National Standard for Respiratory Protection*. American National Standards Institute, Inc. New York, New York ANSI Z88.2.
- American National Standards Institute. (2001) *American National Standard Respirator Fit Testing Methods*. American Industrial Hygiene Association, Fairfax Va. ANSI Z 88.10.
- Campbell DL, Coffey CC and Lenhart SW. (2001) Respiratory Protection as a Function of Respirator Fitting Characteristics and Fit-Test Accuracy. *Am. Ind. Hyg. Assoc. J.* 62(1):36-44.
- Code of Federal Regulations. (2003a) "Acrylonitrile" .U.S. Government Printing Office, Office of the Federal Register, Washington, DC. Title 29, Section 1910.1045.
- Code of Federal Regulations. (2003b) Respiratory Protection. U.S. Government Printing Office, Office of the Federal Register, Washington, DC. Title 29, Section 1910.134.
- Coffey CC, Campbell DL, Myers WR, Zhuang Z, and Das S. (1998) Comparison of Six Respirator Fit-test Methods with an Actual Measurement of Exposure in a Simulated Health Care Environment: Part II—Method Comparison Testing. *Am. Ind. Hyg. Assoc. J.* 59(12):862–870.
- Colton CE. February 11, 2002. [Private Conversation]. "Origin of the Required Fit Factor" 3M, 3M Center, St. Paul MN 55144.
- Crutchfield CD, Ruiz AM, Van Ert MD. (1994) A Validation Study of Respirator Fit Testing by Controlled Negative Pressure. *Appl. Occup. Environ. Hyg.* 9(5): 362-366
- 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. Respir. Prot.* 9(3):22-38.
- Janssen L, Luinenburg M, Mullins H, and Nelson T. (2002) Comparison of Three Commercially Available Fit Test Methods, *Am. Ind. Hyg. Assoc. J.* 63 (6): 762–767.
- Los Alamos National Laboratory. (1983) Evaluation of Irritant Smoke Qualitative Fitting Test for Respirators by JL Marsh (LA 9778-MS) Los Alamos, N.M. Los Alamos National Laboratory.
- Marsh JL. (1984) Evaluation of Irritant Smoke Qualitative Fitting Test for Respirators. *Am. Ind. Hyg. Assoc. J.* 45(6): 371-376.
- Minitab Inc. (2000) MINITAB™, [Computer Software] State College, Pennsylvania Office The Pennsylvania State University.
- Mullins HE, Danisch SG, and Johnston AR . (1995) Development of a New Qualitative Test for Fit Testing Respirators. *Am. Ind. Hyg. Assoc. J.* 56(11):1068-1073.
- Nelson TJ, Skredtvedt OT, Loschiavo JL, and Dixon SW. (1984) Development of an Improved Qualitative Fit Test Using Isoamyl Acetate. *J. Int. Soc. Respir. Prot.* 2(2):225-248.
- Nelson TJ, Jayjock MA, and Colton CE. (2000) How Protective are Respirator Assigned Protection Factors: An Uncertainty Analysis. *Am. Ind. Hyg. Assoc. J.* (61):388-393.
- Zhuang Z, Coffey CC, Jensen P, Campbell DL, Lawrence RB, and Myers WR. (2003) Correlation Between Quantitative Fit Factors and Workplace Protection Factors Measured in Actual Workplace Environments at a Steel Foundry." *Am. Ind. Hyg. Assoc. J.* 64(6) 730-738.