Journal of Occupational and Environmental Hygiene, 11: 77–84
ISSN: 1545-9624 print / 1545-9632 online

DOI: 10.1080/15459624.2013.843780

Determining Sample Size and a Passing Criterion for Respirator Fit-Test Panels

D. Landsittel,[1] Z. Zhuang,[2] W. Newcomb,[2] and R. Berry Ann[2]

[1]Center for Research on Health Care Data Center, Division of General Internal Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania
[2]National Personal Protective Technology Laboratory, National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Pittsburgh, Pennsylvania

Few studies have proposed methods for sample size determination and specification of passing criterion (e.g., number needed to pass from a given size panel) for respirator fit-tests. One approach is to account for between- and within-subject variability, and thus take full advantage of the multiple donning measurements within subject, using a random effects model. The corresponding sample size calculation, however, may be difficult to implement in practice, as it depends on the model-specific and test panel-specific variance estimates, and thus does not yield a single sample size or specific cut-off for number needed to pass. A simple binomial approach is therefore proposed to simultaneously determine both the required sample size and the optimal cutoff for the number of subjects needed to achieve a passing result. The method essentially conducts a global search of the type I and type II errors under different null and alternative hypotheses, across the range of possible sample sizes, to find the lowest sample size which yields at least one cutoff satisfying, or approximately satisfying all pre-determined limits for the different error rates. Benchmark testing of 98 respirators (conducted by the National Institute for Occupational Safety and Health) is used to illustrate the binomial approach and show how sample size estimates from the random effects model can vary substantially depending on estimated variance components. For the binomial approach, probability calculations show that a sample size of 35 to 40 yields acceptable error rates under different null and alternative hypotheses. For the random effects model, the required sample sizes are generally smaller, but can vary substantially based on the estimate variance components. Overall, despite some limitations, the binomial approach represents a highly practical approach with reasonable statistical properties.

Keywords fit-test panels, fit-testing, sample size

Address correspondence to: Ziqing Zhuang, Respiratory Protection Team Leader, Technology Research Branch, National Personal Protective Technology Lab (NPPTL), National Institute for Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention (CDC), 626 Cochrans Mill Road, Building 13, P.O. Box 18070, Pittsburgh, PA 15236; e-mail: zaz3@cdc.gov

INTRODUCTION

Assessing Fitting Characterisics in the Approval of Respirators in the United States

Prior to 1972, the U.S. Bureau of Mines (USBM) was responsible for testing and approval of respirators used in the United States. Respirators for the protection of miners from inhaling harmful coal dust were tested for fit using a qualitative coal dust fit-test with a minimal number of USBM employees as test subjects. In 1970, the National Institute for Occupational Safety and Health (NIOSH) was mandated to co-approve respirators with the USBM. NIOSH took over the administration of the program as well as the USBM's activities for respirator performance testing. At that time coal dust was replaced with isoamyl acetate vapor (an organic vapor) in the fit-test. In 1995 the fit-testing of particulate filtering respirators was abandoned with the transition to a new respirator approval regulation because of the difficulty and lack of appropriate fit-testing techniques.

Also at that time, NIOSH became the sole agency responsible for approving most respirators, with those functions now carried out by the NIOSH National Personal Protective Technology Laboratory (NPPTL). A number of studies have been conducted to compare different fit-test methods and assess characteristics associated with fit.[1–5] As NIOSH revises its respirator approval standard, re-establishment of an assessment of fitting characteristics in the approval process is anticipated. The goal of any proposed fit-test criterion would be to demonstrate the ability of a respirator to fit the facial sizes and shapes for which it was designed. To achieve this, it is necessary for the method be able to reject a high percentage of ineffective respirators, while still passing a high percentage of highly effective respirators.

The Role of Facial Dimensions in Respirator Testing

One assumption for interpreting fit-test results is that the sample, or test panel, represents a random sample of the

targeted population. Over the last 40 years, a variety of approaches have been utilized to approximate a representative sample. Initially, the USBM program tested particulate filtering respirators on three subjects with "varying facial shapes and sizes," including one full, one average, and one lean, a provision that was criticized as being too vague.[6] In 1972, Los Alamos National Laboratory (LANL) developed recommendations for test panels to evaluate respirator fit, which, based on a 1967–68 USAF survey, led to the LANL proposal of 25-subject fit-test panels.[7,8]

Following criticisms of the LANL specifications [9], NIOSH initiated a study to develop a more representative anthropometric database for civilian respirator users[10] and produce a panel more representative of the age and racial/ethnic distributions of the current civilian population. Based on the NIOSH anthropometric survey, Zhuang et al.[11–13] defined two new test panels, including 1) the NIOSH bivariate Respirator Fit-test Panel (NRFTP) with ten cells based on face length and face width, and 2) the Principal Component Analysis (PCA) Panel, which was defined from the first two principal components (which are the independent linear combinations that explain the maximum degree of variance) from 10 dimensions of face size.

The goal of this current study is to outline criteria for both the sample size requirement and cutoff for achieving passing results, with the assumption that the NRFTP, PCA, or other such panel is being utilized to achieve a sample that is representative of the underlying worker population.

Goals of the Current Study

The previously described history behind fit-testing serves to motivate the need for determining how many subjects should be tested, and how many should be required to achieve a passing fit factor. Discussing the role of facial dimensions in fit-testing is also necessary to describe the basics of how subjects might be selected for fit-testing a representative sample. The goals of this current study are, however, strictly limited to determining the number of subjects who should be required in a respirator test panel, and the number who should be required to achieve a passing result based on criteria that can be easily implemented in practice. Numerous other considerations, such as choosing between the NRFTP and the PCA (or some combination), determining how the manufacturer instructions affect the selection of a test panel, and deciding on a cutpoint for an acceptable fit factor, must also be considered in regulation decisions. However, the current study is not meant to outline the regulatory process, but, instead, aims to assess the key statistical properties of the binomial test as it relates to sample sizes for testing the percentage of workers achieving a given fit factor.

Boostrapping and Random Effects Models

To assess quantitative fit of respirators, the Food and Drug Administration (FDA) [14] recommends use of a random effects model and bootstrap intervals, where fit factors are log-normal with two components of variability: within- and between-subject [15,16]. Zhang and Kolz [17] subsequently presented a closed-form normal approximation and showed that their approximation compares well with the more computationally intensive bootstrap methods. They provide sample size estimates for testing the proportion of users who receive a specified level of protection. Subsequent discussion will focus on the closed-form approximation.

An Alternative Approach Using Binomial Tests

Rather than use the repeated measures data on individual subjects, and random effects modeling to account for within-subject correlations, the proposed binomial approach categorizes each subject as pass or fail based on the average fit factor across multiple donnings. The type I and type II errors can then be calculated directly from the binomial distribution[18]. A global search can then be implemented to find the required sample size, and corresponding cutoffs for the number of subjects needing to pass.

METHODS AND MATERIALS

Overview of the Methods

The approach is divided into four different sections. These are: 1) estimating the parameter of interest as the proportion of users achieving a sufficiently high fit, 2) defining the null and alternative hypotheses and associated type I and type II error rates, 3) sample size estimation, and 4) analysis of the NIOSH benchmarking respirator fit-test data.

Each section outlines the needed terminology and concepts, and equations and methods for both the random effects model and binomial approach.

Estimation

Overall Estimation Concepts

The fit factor is defined as the ratio of the measured challenge agent concentration outside the respirator to its concentration inside the respirator. The primary parameter of interest from a quantitative fit-test is the proportion of users who achieve a sufficiently high fit. Using the same notation as given by Zhang and Kolz,[17] let θ denote the proportion of subjects who achieve that fit factor, and let x denote the log-fit factor. To illustrate the method, 100 is specified as an example for the minimum acceptable fit factor (based on OSHA requirements 29 CFR 1910.134); using the notation of Zhang and Kolz[17], $x = \log_e(100)$ and $\theta =$ the proportion of subjects achieving a fit factor of x. Other values, such as 10, could be specified for x.

Estimation with the Random Effects Model

Assuming the log-fit factor for a given user of a given respirator model is normal with mean μ and random intercept a, the terms ${\sigma_a}^2$ and ${\sigma_e}^2$ represent the between-subject and within-subject (or error) variance, respectively. Further, let 1-q denote the probability with which we require the respirator achieve the log-fit of x. θ can then be estimated using Equation 1 from Zhang and Kolz[17] (which is also given below).

Estimating the proportion achieving sufficient fit

$$\hat{\theta} = \Phi\left(\left(\hat{\mu} - z_q\hat{\sigma}_e - x\right)/\hat{\sigma}_a\right) \quad \textbf{(1)}$$

Estimation for the Binomial Method

An alternative method is to dichotomize each of the n subjects' log-fit factors as being at least x (denote as $y_i = 1$) or below x ($y_i = 0$). The method for doing so—e.g., taking the mean or maximum log-fit factor for each subject (if the fit factor is measured multiple times per subject) before comparing to x—depends on the goals of the given fit-testing study. For instance, if one wishes to assess potential fit of the respirator, the maximum may be a more appropriate summary measure (see Discussion). The estimate of θ is then the simple proportion of subjects who achieve a log-fit factor of x. Note that without utilizing individual repeated measures data on a given subject, this method does not involve the probability q, but instead dichotomizes the log-fit factor.

Type I and Type II Error

General Definitions

To calculate type I and type II error rates, the null and alternative hypotheses are defined as H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$, respectively, where $\theta_1 > \theta_0$ and passing the respirator equates to rejecting the null. Type I error, or falsely rejecting the null, thus equates to falsely passing a respirator which in truth achieves a log-fit of at least x for only $100 \times \theta_0\%$ of the population. The maximum probability of a type I error (α) is typically set equal to 0.05. This study also considers lower values of α for lower values of θ_0. For instance, we may want to control α at 5% for a reasonably ineffective model, e.g., with $\theta_0 = 0.6$ but also control α at a more strict level, e.g., 1%, for an even less effective model, e.g., with $\theta_0 = 0.5$. Type II error is defined as falsely failing to reject the null, or falsely failing a respirator model which in truth does achieve a log-fit of at least x for $100 \times \theta_1\%$ of the population. The maximum probability of a type II error, or β, is typically set equal to 0.20; subsequent analyses also consider $\beta = 0.1$.

Calculations of Error Rates for Each Method

For the random effects model, α and β are directly incorporated into the sample size calculation, as described below. For the binomial method, the type I and type II error rates are instead calculated for each possible cutoff value for the number passing (defined as Y^*) and the hypothesized value of θ. Type I error rates are calculated from Equation 2a and type II error rates are calculated from Equation 2b. Each follows directly from the binomial distribution.

Probability of falsely rejecting the null (passing under the null)

$$P(Y \geq Y* \,|n, \theta_0) = \sum_{k=Y*}^{n} \binom{n}{k} \theta_0^k (1-\theta_0)^{n-k} \quad \textbf{(2a)}$$

Probability of falsely not rejecting the null (failing under the alternative)

$$P(Y \geq Y* \,|n, \theta_1) = \sum_{k=0}^{Y*} \binom{n}{k} \theta_1^k (1-\theta_1)^{n-k} \quad \textbf{(2b)}$$

Sample Size Estimation

Sample Size for the Random Effects Model

For the random effects model, the required sample size is given by Equation 3a below, which is the same formula as Equation 3 in Zhang and Kolz[17] (using n instead of m to denote the number of subjects); Φ represents the standard normal density, and z_α and z_β denote the critical values of the standard normal distribution for the given type I and type II error rates, respectively. Assuming a constant number of measurements ($m = m_i$) per subject, $\boldsymbol{V}$ represents the Fisher information matrix (Equation 3b), $\boldsymbol{g}$ denotes the gradient for the inverse normal function of θ (Equation 3c), and the variance is estimated by $\boldsymbol{g^T V g}$, where z_q denotes the critical value of the standard normal distribution corresponding to the probability with which we require the respirator achieve the log-fit of x. Zhang and Kolz define $\boldsymbol{V}$ and $\boldsymbol{g}$ in their appendix (unnumbered, p. 722) with some differences in notation (i.e., using n instead of m to denote measurements per subject); w is defined as $m/(1 + m\rho)$. Conducting the matrix multiplication of quantities in Equations 3b and 3c (of dimension 3×3 and 3×1, respectively) yields a scalar for $g^T V g$.

Required sample size for the random effects model

$$n \geq \frac{\left(z_\alpha + z_\beta\right)^2 g^T V g}{\left(\Phi^{-1}(\theta_1) - \Phi^{-1}(\theta_0)\right)^2} \quad \textbf{(3a)}$$

The Fisher information matrix

$$V = \begin{bmatrix} \sigma^2/w & 0 & 0 \\ 0 & 2\sigma_e^4\left\{w^{-2} + m^{-2}(m-1)^{-1}\right\} & -2\sigma_e^4/\{m(m-1)\} \\ 0 & -2\sigma_e^4/\{m(m-1)\} & 2\sigma_e^4/(m-1) \end{bmatrix} \quad \textbf{(3b)}$$

Gradient for the inverse normal function of θ

$$g = [1/\sigma_a, -(\mu - z_q\sigma_e - x)/(2\sigma_a^3), -z_q/(\sigma_a\sigma_e)]^T \quad \textbf{(3c)}$$

Sample Size for the Binomial Method

For the binomial method, the required sample size is defined as the minimum sample size where at least one value of Y^* meets the proposed type I and type II error rates, i.e., the probability of observing the given number of successes or greater is below α under null θ_0, and the probability of observing less than the given number of successes is below β under the alternative θ_1. A global search over all critical values for all sample sizes is then conducted to determine which sample sizes and which cutoffs for the number passing satisfy a given set of values for α and β. The required sample size is then defined as the minimum value that simultaneously achieves error rates within the specified limits.

Analysis of Empirical Data

Findings of the two methods were then illustrated using fit factors collected on 98 different respirator models, which included both filtering facepiece and half-mask respirators. These data were collected by NIOSH for benchmarking purposes. These data are presented here for illustrative purposes only, and are not necessarily reflective of the distributions we would expect to see when respirators are fit to test panels which reflect their intended target population. For each model, regardless of its intended size (e.g., small, medium, large, or any size), benchmarking data were collected using the NIOSH bivariate Respirator Fit-test Panel (NRFTP) with 25 subjects and 3 fit factors measured on each subject. To first summarize the overall distributions, mean log-fit factors (MLFF) were calculated on two levels: 1) data for the MLFFs on a given subject were averaged over the 3 donnings, and 2) the MLFF and standard deviation of the log fit-factor (SDLFF), across subjects for a specific respirator were calculated and summarized (with the quartiles of the distribution over those subject-specific MLFF transformed back to the original scale).

These findings were summarized across the 98 models by reporting results for the models representing quartiles of the overall distribution of the MLFF. Selected q-q plots[18] were also displayed to assess the normality of the log-fit factors (LFF). In terms of the individual donnings, estimates of the between- and within-subject variability were also summarized across the 98 models. Results were then compared in terms of how the between- and within-subject variability to calculate the required sample size (with the random effects model) compares with the sample sizes based on binomial probabilities.

RESULTS

Overview of Results

Critical values and corresponding error rates and sample sizes for the binomial method are presented first since they are independent of empirical data; since tests and required sample sizes for the random effects model depend on the mean log-fit factors and variability estimates, those findings are presented after analysis of the empirical data. Analyses of empirical data are divided into the assessment of normality, and the summary of mean log-fit factors and variance components. Estimates of required sample sizes are then presented for the random effects model.

Error Rates across Sample Sizes and Critical Values for the Binomial Method

Table I shows the type I and type II error rates for sample sizes of 25 to 50 in increments of 5. Results are shown for only a selected subset of results, where the critical value for number passing is between 70% and 80% to highlight the critical values with more optimal results.

Results show that requiring only 17 or 18 of 25 subjects to pass yields elevated type I error rates of over 15%. In contrast, increasing the critical value to 19 or 20 of 25 yields high type II error rates, of over 20% for $\theta_0 = 0.6$. A sample size of 30 yields similar results. For a sample size of 35, the lower critical values, i.e., requiring 24 or 25 to pass, again yielded elevated type I levels and critical values of 27 or 28 again yielded elevated type II levels. For the critical value of 26, however, results nearly met the pre-specified optimal levels (i.e., type I error rates below 1% or 5% and type II error rates below 20% or 10% for the different null and alternative hypotheses), except for a type I error of 5.8% (instead of ≤5%) for $\theta_0 = 0.6$. Sample sizes of 40, 45, and 50 all yielded at least one critical value that met the pre-specified criteria, including 30 of 40, 33 or 34 of 45, and 37 or 38 of 50. Results were similar for this range of sample sizes. In summary, a sample size of 40 is necessary to achieve the desired error rates, with a sample size

TABLE I. Type I and II Error Rates Using Binomial Probabilities for Selecting Critical Values

Sample Size	Critical Value (%)	Type II Error Rates		Type I Error Rates	
		$\theta_1 = 0.9$	$\theta_1 = 0.8$	$\theta_0 = 0.6$	$\theta_0 = 0.5$
25	17 (68%)	**<0.1%**	**4.7%**	27.4%	5.4%
	18 (72%)	**0.2%**	**10.9%**	15.4%	2.2%
	19 (76%)	**0.9%**	22.0%	7.4%	**0.7%**
	20 (80%)	**3.3%**	38.3%	**2.9%**	**0.2%**
30	21 (70%)	**<0.1%**	**6.1%**	17.6%	2.1%
	22 (73%)	**0.2%**	**12.9%**	9.4%	**0.8%**
	23 (77%)	**0.8%**	23.9%	**4.4%**	**0.3%**
	24 (80%)	**2.6%**	39.3%	**1.7%**	**0.1%**
35	24 (69%)	**<0.1%**	**3.4%**	19.5%	2.0%
	25 (71%)	**<0.1%**	**7.5%**	11.2%	**0.8%**
	26 (74%)	**0.2%**	**14.6%**	5.8%	**0.3%**
	27 (77%)	**0.6%**	25.5%	**2.6%**	**0.1%**
	28 (80%)	**2.0%**	40.1%	**1.0%**	**<0.1%**
40	28 (70%)	**<0.1%**	**4.3%**	12.9%	**0.8%**
	29 (73%)	**<0.1%**	**8.8%**	7.1%	**0.3%**
	30 (75%)	**0.1%**	**16.1%**	**3.5%**	**0.1%**
	31 (78%)	**0.5%**	26.8%	**1.6%**	**<0.1%**
	32 (80%)	**1.5%**	40.7%	**0.6%**	**<0.1%**
45	31 (70%)	**<0.1%**	**2.5%**	14.3%	**0.8%**
	32 (71%)	**<0.1%**	**5.2%**	8.4%	**0.3%**
	33 (73%)	**<0.1%**	**9.9%**	**4.5%**	**0.1%**
	34 (76%)	**0.1%**	**17.4%**	**2.2%**	**<0.1%**
	35 (78%)	**0.4%**	28.0%	**0.9%**	**<0.1%**
	36 (80%)	**1.2%**	41.2%	**0.4%**	**<0.1%**
50	35 (70%)	**<0.1%**	**3.1%**	9.6%	**0.3%**
	36 (72%)	**<0.1%**	**6.1%**	5.4%	**0.1%**
	37 (74%)	**<0.1%**	**11.1%**	**2.8%**	**<0.1%**
	38 (76%)	**0.1%**	**18.6%**	**1.3%**	**<0.1%**
	39 (78%)	**0.3%**	28.9%	**0.6%**	**<0.1%**
	40 (80%)	**0.9%**	41.6%	**0.2%**	**<0.1%**

Note: **Bold font** denotes error rates meeting the pre-specified criteria (of ≤10%, 20%, 5%, and 1%, respectively).

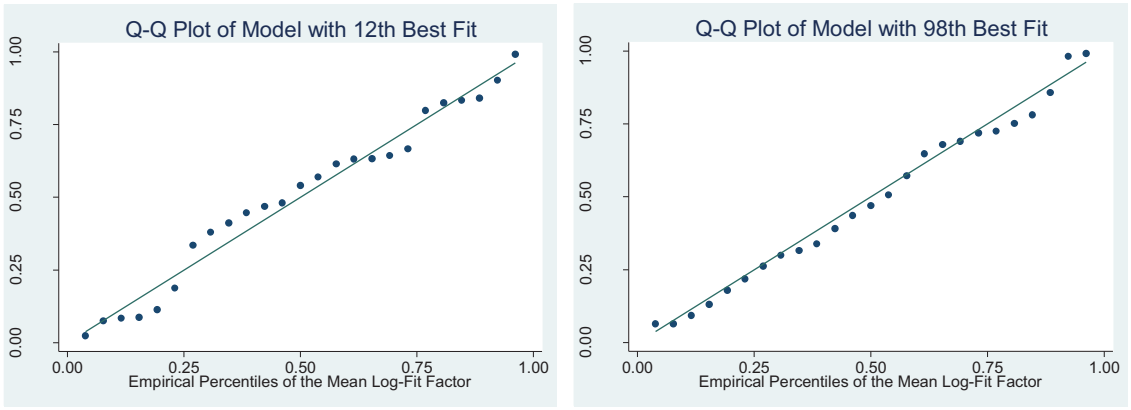

FIGURE 1. Q-Q plots of two models with good to excellent correspondence to normality (color figure available online).

of 35 yielding only a slight increase over the stated α-level for $\theta_0 = 0.6$.

Summary of Empirical Data

Fit factors were collected on 98 models, each tested on the NIOSH bivariate Respirator Fit-test Panel (NRFTP) with 25 subjects and 3 donnings per subject. Results were not identified by individual respirator model since 1) that was not the goal of this article or this data collection effort, and 2) data were collected using the NRFTP regardless of intended size of the respirator model.

Normality of Subject-Specific Log-Fit Factors by Respirator Model

The q-q plots (which plot percentiles of the ordered empirical data versus the percentiles of the standard normal distribution) were evaluated to assess normality of (subject-specific) mean log-fit factors for a given respirator. For 58.2% of the respirator models (57/98), the distribution was consistently near normality, i.e., the percentile for the empirical data was within 5% or at most 10% of the standard normal percentile across the range of data. See Figure 1 for two such examples; in one example the correspondence to normality is near perfect, and in the other example there is relatively close agreement with some minor discrepancies.

In another 25.5% (25/98) of respirator models, the deviations from normality were noticeably larger, i.e., the percentile for the empirical data was at least 10% or 15% off from the standard normal percentile for at least some of the data. Figure 2 gives two examples of models which illustrate these moderate to reasonably large deviations from normality. The remaining 16.3% (16/98) of respirator models showed large deviations in the range of a 20% to 25% difference in the respective percentiles; see the two examples in Figure 3. Results of these plots show clear deviations from normality for around 40% of the respirator models tested. To evaluate whether these deviations could be explained by testing models of a particular size on the overall NRFTP, q-q plots of only the 33 one-size-fits models were also evaluated separately. The relative frequencies of deviations from normality (i.e., the percent of large, moderate, and small deviations) were very similar to the entire sample of models; 4/33 (12.1%) had large deviations from normality, 10/33 (30.3%) had moderate deviations, and 19/33 (57.5%) had small deviations.

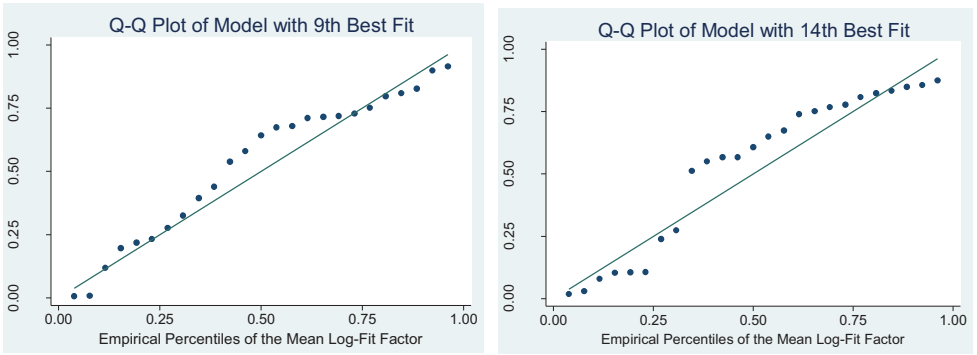

FIGURE 2. Q-Q plots of two models showing moderately large deviations from normality (color figure available online).

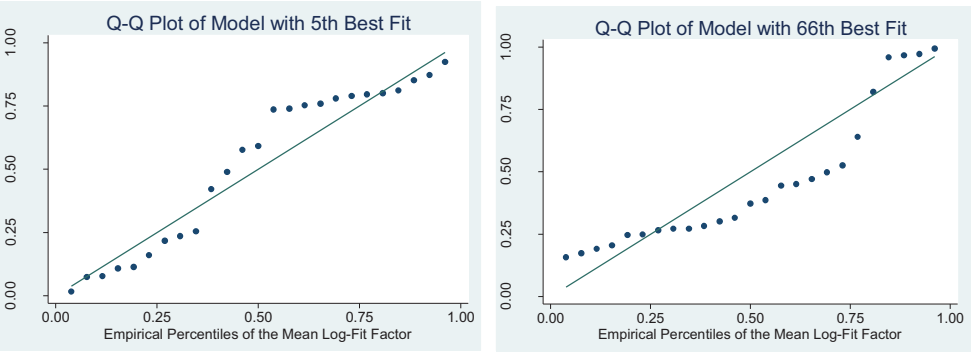

FIGURE 3. Q-Q plots of two models showing large deviations from normality (color figure available online).

Summary of Mean Log-Fit Factors

Table II summarizes the distributions of LFF (and FF in the original scale using the geometric mean, denoted by GM) by displaying summary statistics for the models representing quartiles when ranked by the GM.

The median value (across the 98 models) for the GM FF was 81.6; the worst fitting respirator had a GM of 5.6 and the best fitting respirator had a GM of 2207. The middle 50% of respirators had a GM between approximately 30 and 290. As described above, caution is warranted in interpreting these results since all respirator models (for purposes of this data collection effort) were tested on the entire NRFTP regardless of intended size of the respirator model. The purpose of describing these data was to simply provide an illustration of the sample size methods as described below.

Summary of Variance Components

Table III displays the between- and within-subject variability estimates, and ratio of variances, when sorted by either the between or the within estimate (so that columns 2, 3, and 4, or 4, 5, and 6 represent variability estimates from a single respirator model). These estimates are subsequently utilized to calculate sample size requirements and significance tests. They also clearly illustrate the wide range of within- and between-subject variances and the variability in that ratio, although some fraction of the observed variability may be due to the experimental design used to collect the benchmarking data.

Sample Size Estimates with the Random Effects Model

The estimates for the mean log-fit factor (MLFF), and $\sigma_a{}^2$ and $\sigma_e{}^2$ (between- and within-subject variance), were used to illustrate potential results for the required sample size using the random effects model based on Equation 3. More specifically, to illustrate the range of potential sample sizes with this method, we used the combination of the MLFF, $\sigma_a{}^2$, and $\sigma_e{}^2$ corresponding the empirical data for the model with the minimum, first quartile, median, third quartile, and maximum values for ρ ($= \sigma_a{}^2/\sigma_e{}^2$). Results show that the required sample size for most models is in the range of 25 – 42, but substantial variability can exist in the required sample size when the within-subject variability is large and the ratio of between- to within-subject variability is low. Sample sizes were calculated separately with q = 0.8 and q = 0.9, where q denotes the percentage of the distribution of fit factors required to be above 100, and are summarized in Table IV.

TABLE II. Summary of Model Fit Factor by Quartiles of the Geometric Mean

		Fit Factor (transformed back to original scale)		
Model Rank	**GM**	**Median**	**1st Q**	**3rd Q**
Minimum	5.6	5.3	3.6	8.3
1st Q	29.7	20.5	13.3	74.3
Median	81.6	57.7	25.4	347.4
3rd Q	290.1	283.8	65.3	1066.6
Maximum	2207.2	3229.5	1529.4	4782.6

GM = geometric mean.
SD = standard deviation.
1st Q and 3rd Q = the inter-quartile range.

TABLE III. Summary of Between and Within Subject Variance for the Log-Fit Factors

	Variances sorted by $\sigma_a{}^2$			Variances sorted by $\sigma_e{}^2$		
Model Rank	$\sigma^2{}_a$	$\sigma^2{}_e$	ρ	$\sigma^2{}_a$	$\sigma^2{}_e$	ρ
Minimum	0.083	0.057	1.438	0.083	0.057	1.438
1st Q	1.224	0.457	2.682	2.558	0.495	5.168
Median	2.515	1.757	1.431	6.018	0.916	6.573
3rd Q	3.407	2.149	1.586	2.331	1.534	1.520
Maximum	7.955	0.987	8.062	1.369	2.733	0.501

1st Q and 3rd Q = the inter-quartile range.
σ_a and σ_e = between and within subject variance (using data in the log scale).
$\rho = \sigma_a{}^2/\sigma_e{}^2$.

TABLE IV. Sample Sizes Using the Random Effects Approach to Test $\theta_0 = 0.6$ versus $\theta_1 = 0.8$

Quartile of ρ	MLFF (GM)	Variability			Sample Size	
		σ_a^2	σ_e^2	ρ	q = 0.8	q = 0.9
Minimum	4.601 (99.58)	0.524	2.713	0.193	416	899
1st quartile	5.966 (389.94)	2.114	1.390	1.521	25	27
Median	3.881 (48.47)	3.195	1.299	2.459	33	42
3rd quartile	3.951 (51.99)	2.969	0.774	3.835	28	34
Maximum	2.789 (16.26)	4.079	0.259	15.78	30	33

$\rho = \sigma_a^2/\sigma_e^2$.
MLFF = mean log-fit factor.
GM = geometric mean.
q = percentage of donnings required to surpass the required fit of $\log_e(100)$.
σ_a^2 and σ_e^2 = between and within subject variance (using data in the log scale).

DISCUSSION

This article proposes the use of a simple binomial test for determining whether a given respirator achieves sufficient fit for a panel of subjects. The given calculations show that sample sizes of 35 to 40 achieve, or approximately achieve (within <1%) the targeted type I and II error rates (of 1% or 5%, and 10% or 20%, depending on the given null and alternative hypothesis, respectively) to discriminate between respirators that provide a sufficiently high average fit factor (e.g., of 100) to at least 80% of the population (versus the null of sufficiently fitting 60% or less of the population). This method gives a simple cutoff for passing or failing a respirator using a given panel size.

Analysis of the NIOSH Benchmarking Data

The NIOSH benchmarking data were used in this study to illustrate how the variance components might affect results of sample size calculations with the random effects model. These data, however, were not collected for purposes of making inferences about specific respirators, and were not collected for purposes of this study. Therefore, these data are only useful for providing an illustrative example of how variance components and sample size estimates might vary across different respirators. Given these limitations of the data, results provide only an illustration for the possible impact of within- and between-subject variability that may be an overestimate of what would be observed in practice.

Comparison of Methods

Although results were also illustrated for the random effects model, it is critical to note that findings are not meant to investigate which of these two approaches is optimal. In fact, the two methods are not directly comparable based on empirical data, as the random effects model represents a more rigorous criterion, where $100 \times (1-q)\%$ of the within-subject distribution must meet the log-fit of x. In contrast, agencies such as NIOSH may only wish to assess whether a given respirator model is capable of providing a sufficient fit for a sufficiently high percentage of subjects (as discussed at public meetings in 2007 [19]). In those cases, the subject may be classified as passing if the maximum log-fit factor exceeds the given value of x. Hence, the empirical data and results of the random effects are presented strictly to illustrate the degree of variability that may be observed and how that variability can affect sample size calculations, not to specifically compare methods.

In terms of practical implementation, use of the random effects model introduces substantial challenges. As noted, within- and between-subject variability estimates can greatly affect sample size calculations and subsequent significance tests. Use of this method does not lead to a universal cutoff for achieving a passing result, which may be very problematic for agencies involved in fit-testing. This represents a major limitation for public acceptance of any subsequent criterion for respirator approval, especially since manufacturers may wish to pretest their models and interpret those pre-test findings. Different respirators, with different variability estimates, would potentially have different cutoffs for passing. In contrast, the binomial test yields a single criterion for a given panel size and a single estimate from the necessary panel size.

Another specific concern for the random effects model is the assumption of normality. For the NIOSH benchmark data, many of the distributions varied substantially from normality. It is unclear how substantially this might affect the findings. The binomial test, however, is not dependent on any distributional assumptions.

Selecting the Final Sample Size and Passing Percentage Cutoff

In general, requiring at least 26 of 35 test subjects to achieve sufficient fit nearly meets the above considerations while minimizing the required sample size. For a sample size of 25, requiring at least 19 to pass yields the optimal cutoff, but this criterion still yields a type II error over 20% for $\theta = 0.80$, and a type I error of over 5% for $\theta = 0.60$. Increasing the sample size to 30, and selecting the optimal cutoff of 23 of

30, does yield a sufficiently low type I error, but still equates to a type II error over 20%. Increasing the sample size to 35 nearly reaches all specified requirements, except that the type I error for $\theta = 0.60$ is slightly over 5%.

CONCLUSION

The proposed approach uses binomial probabilities to calculate error rates across different null and alternative hypotheses about the respirator's underlying effectiveness, different sample sizes, and different minimum requirements for the percentage of subjects required to pass fit-testing. An inward leakage test is used as an illustration for applying this approach. This approach, as compared to the previously proposed random effects model, substantially simplifies the problem of sample size estimation for respirator fit-testing, simultaneously identifies a specific cutoff for the percentage of test subjects required to pass, and achieves reasonable statistical properties that correspond to interpretable probability statements of scientific interest. The method makes minimal assumptions and does not require any preliminary data or knowledge about the underlying distribution of fit factors or partitioning of the variance components; development of the test criteria is based solely on achieving adequately high probabilities for rejecting ineffective respirators, and passing effective respirators, and is thus well suited for estimating sample sizes and passing criteria that are easily interpreted and implemented in practical settings. For the binomial approach, probability calculations show that a sample size of 35 to 40 yields acceptable error rates under different null and alternative hypotheses. For the random effects model, the required sample sizes are generally smaller, but can vary substantially based on the estimate variance components. Overall, despite some limitations, the binomial approach represents a highly practical approach with reasonable statistical properties.

DISCLAIMER

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

REFERENCES

1. **Coffey C.C., D.L. Campbell, W.R. Myers, Z. Zhuang, and S. Das:** Comparison of six respirator fit-test methods with an actual measurement of exposure in a simulated health care environment: Part I—Protocol development. *Am. Ind. Hyg. Assoc. J. 59*:852–861 (1998).
2. **Coffey C.C., D.L. Campbell, W.R. Myers, and Z. Zhuang:** 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*:862–870 (1998).
3. **Zhuang, Z., C.C. Coffey, P.A. Jensen, D.L. Campbell, R.B. Lawrence, and W.R. Myers:** Correlation between quantitative fit factors and protection factors measured under actual workplace environments at a steel foundry. *Am. Ind. Hyg. Assoc. J. 64*:730–738 (2003).
4. **Zhuang, S., C. C. Coffey, and R. Berry Ann:** The effect of subject characteristics and respiratory features on respirator fit. *J. Occup. Environ. Hyg. 2*:641–649.
5. **Roberge, R. J., Z. Zhuang, and L. M. Stein**,: Association of body mass index with facial dimensions for defining fit-test panels. *J. Intern. Soc. Respir. Protect. 23*:44–52.
6. **Hyatt, E.C., J.A. Pritchard, B.J. Held, et al.:** *Respiratory studies for the National Institute for Occupational Safety and Health, July 1, 1972–June 3, 1973, LA-5620-PR*. Los Alamos, N.M.**:** Los Alamos Scientific Laboratory, 1994..
7. **Hack, A.L., E.C. Hyatt, B.J. Held, T.O. Moore, C.P. Richards, and J.T. McConville:** *Selection of respirator test panels representative of U.S. adult facial size, LA-5488*. Los Alamos, N.M.**:** Los Alamos Scientific Laboratory of the University of California, 1974.
8. **Hack, A.L., and J.T. McConville:** Respirator protection factors: Part I-Development of an anthropometric test panel. *Am. Ind. Hyg. Assoc. J. 39*:970–975 (1978).
9. **National Institute for Occupational Safety and Health (NIOSH)**: "Preamble to revised *42* CFR Part 84." *Federal Register 60*:30355 (8 June 1995).
10. **Zhuang, Z., and B. Bradtmiller:** Head-and-face anthropometric survey of U.S. respirator users. *J. Occup. Environ. Hyg. 2*:567–576 (2005).
11. **Zhuang, Z., J. Guan, H. Hsiao, and B. Bradtmiller:** Evaluating the representativeness of the LANL respirator fit-test panels for the current U.S. civilian workers. *J. Int. Soc. Resp. Prot. 21*:83–93 (2004).
12. **Zhuang, Z., B. Bradtmiller, and R.E. Shaffer**,: New respirator fit-test panels representing the current U.S. civilian work force. *J. Occup. Environ. Hyg. 4*:647–659 (2007).
13. **Zhuang, Z., D. Groce, H.W. Ahlers, et al.:** Correlation between respirator fit and respirator fit-test panels by respirator size. *J. Occup. Environ. Hyg. 5*:617–628.
14. **Food and Drug Administration (FDA):** Class II special controls guidance document: Filtering facepiece respirator for use by the general public in public health medical emergencies. *Guidance for Industry and FDA Staff*, July 3, 2007.
15. **Nicas M., and J. Neuhaus:** Variability in respirator protection and the assigned protection factor. *J. Occup. Environ. Hyg. 1*:99–109 (2004).
16. **Crump K.S.:** Statistical issues with respect to workplace protection factors for respirators. *J. Occup. Environ. Hyg. 4*:208–214 (2007).
17. **Zhang, Z., and R.M. Kotz**,: Inference and sample size calculation in the fit assessment of filtering facepiece respirators. *J. Biopharm. Stat. 1*:713–723 (2008).
18. **Rosner, B.:** *Fundamentals of Biostatistics. 6th Edition*. Boston, Mass.**:** PWS-Kent. (2006).
19. "Total Inward Leakage Requirements for Respirators," *Federal Register 74*:209 (2010). pp. 56144–56146.