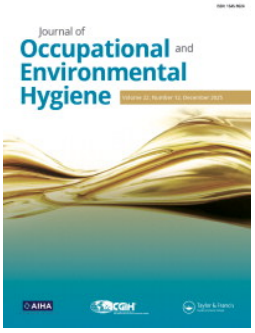

Journal of Occupational and Environmental Hygiene

ISSN: 1545-9624 (Print) 1545-9632 (Online) Journal homepage: www.tandfonline.com/journals/uoeh20

Quantitative headform fit evaluation and predictive modeling to assist with selecting N95 filtering facepiece respirators to mitigate respiratory hazards

Brooke Vollmer, Michael S. Bergman, Harold Boyles, Jordan Meyers, Nora Y. Payne, Jonisha Pollard & Ziqing Zhuang

To cite this article: Brooke Vollmer, Michael S. Bergman, Harold Boyles, Jordan Meyers, Nora Y. Payne, Jonisha Pollard & Ziqing Zhuang (2025) Quantitative headform fit evaluation and predictive modeling to assist with selecting N95 filtering facepiece respirators to mitigate respiratory hazards, Journal of Occupational and Environmental Hygiene, 22:12, 959-969, DOI: 10.1080/15459624.2025.2540829

To link to this article: https://doi.org/10.1080/15459624.2025.2540829

Published online: 27 Aug 2025.

JOURNAL OF OCCUPATIONAL AND ENVIRONMENTAL HYGIENE
2025, VOL. 22, NO. 12, 959–969
https://doi.org/10.1080/15459624.2025.2540829

Quantitative headform fit evaluation and predictive modeling to assist with selecting N95 filtering facepiece respirators to mitigate respiratory hazards

Brooke Vollmer, Michael S. Bergman, Harold Boyles, Jordan Meyers, Nora Y. Payne, Jonisha Pollard, and Ziqing Zhuang

National Personal Protective Technology Laboratory, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Pittsburgh, Pennsylvania

ABSTRACT

Ensuring that respiratory protection is effective for all can be difficult if limited resources are available to assist with selecting a well-fitting respirator model and user guidance. To better understand how various N95® filtering facepiece respirator models fit on a variety of face sizes, a quantitative fit evaluation was performed on 12 different N95 respirators distributed by the Strategic National Stockpile using five manikin headform sizes representative of most of the U.S. worker population (540 total tests). Manikin fit factor results varied depending on the respirator model and headform combination. Four respirator models achieved passing fit results across all headform sizes. Predictive modeling was then initiated, where the headform most closely aligned to an individual's facial dimensions is determined and then used to identify N95 respirators that may provide an acceptable fit. A multinomial logistic regression model was trained and tested using NIOSH's 2003 Anthropometric U.S. Survey and was found to have an accuracy of 85%. To address potential risks associated with predicting only a single headform size, a modified model allowing for multiple headform size predictions was also assessed and found to have an improved accuracy rate of 98%. With further human subject validation and field testing, this modeling approach could be used as a tool to aid in making the fit testing process more efficient, less burdensome, and better enable individuals to use respirators that fit more effectively, thereby adequately protecting them from hazards.

KEYWORDS
Filtering facepiece respirator; N95; NIOSH Approved N95; respirator fit; respirator selection

Introduction

Respiratory protection needs for the public and workers in occupational settings, with or without a respiratory protection plan in place, vary widely, and providing effective respiratory protection for all is a complex issue. Respiratory protective devices (RPDs) and face-worn products have become more readily available to the public in the wake of respiratory health emergencies, where the use of masks and respirators to mitigate the spread of respiratory diseases has gained widespread attention. An RPD is any device designed to protect from inhalation of a hazardous or inadequate atmosphere and must provide a demonstrated and referenceable level of protection. An N95 filtering facepiece respirator (FFR) is a type of RPD that protects the wearer by filtering particles out of the air the user is breathing. N95 FFRs are the most common type of FFR, filtering out at least 95% of particles from the air the user is inhaling when used properly (NIOSH 2021).

N95 FFRs must achieve a tight seal to the face to provide their expected level of protection. Numerous studies have observed that fit testing is necessary to ensure respirators provide their expected level of protection (CDC 1998; Coffey et al. 1998, 1999, 2004; Reponen et al. 2011). Other studies have reported that workers passing a fit test, compared to those who were not fit tested, achieved higher levels of respiratory protection (Colton 1999; Zhuang et al. 2003). The time burden to fit test an employee was estimated by OSHA to be 30 min (Federal Register 1998). In the workplace, performing an OSHA-accepted fit test to find a respirator make, model, and size that fits can require extensive time and costs.

With the various makes, models, and sizes of FFRs available, it may be difficult to find a well-fitting device without proper guidance or tools. This could

CONTACT Brooke Vollmer BVollmer@cdc.gov National Personal Protective Technology Laboratory, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Pittsburgh, Pennsylvania.

be more difficult for certain sexes and ethnic groups that are known to have trouble identifying FFRs that fit them properly. The 2020 Current Population Survey from the U.S. Bureau of Labor Statistics states Black and Asian individuals comprised 13% and 6% of the U.S. workforce (BLS 2021a). In 2020, women comprised approximately 45% of full-time wage and salary workers in the U.S. (BLS 2021b). A recent literature review by Chopra et al. (2021) reports that women and Asian individuals have difficulties in identifying respirators that fit them and may require multiple fit test attempts to obtain an adequately fitting respirator (Chopra et al. 2021). Two studies from South Africa with mixed cohorts of predominantly Black, Asian, and minority ethnic subjects observed low fit test passing rates of 13.8% (Spies et al. 2011) and 22% (Manganyi et al. 2017). Additionally, a 2023 quantitative fit test study of hospital staff using P2/N95 particulate filter respirators (a type of respirator approved to Australian standards) found that males had significantly increased odds of passing a fit test compared to females (Zhang et al. 2023).

Currently, manufacturers of NIOSH Approved N95 FFRs are not required to design their respirators to fit specific facial sizing or anthropometry standards of people. The U.S. Code of Federal Regulations for approving RPDs, 42 CFR Part 84.171, states that tight-fitting facepieces of air-purifying particulate respirators, which include N95 FFRs, must be designed and constructed to fit persons with varying face shapes and sizes either by providing multiple sizes or one size that fits varying face shapes and sizes (Federal Register 1995). Therefore, using a NIOSH Approved N95 FFR alone is not enough to guarantee an ideal fit with a secure seal when used without undergoing proper fit testing.

The National Academies of Sciences, Engineering, and Medicine conducted a study and found that not all workers are regulated by a respiratory protection program (RPP) managed by the Occupational Safety and Health Administration (OSHA) or any other governmental authority (NASEM 2022). Workers could also become exposed to inhalation hazards, such as wildfire smoke or emerging airborne viruses, outside of their workplace settings, and they may not be provided with properly selected and fit-tested respirators for these specific hazards, or be unaware of when and how to use them. Additionally, there is no formal, coordinated system in place to guide and educate the public on when and how to use a respirator and which make, model, and size individuals should obtain for proper respiratory protection.

No national or state-based frameworks exist within the United States to obtain a fit test when respiratory protection is used outside of occupational environments or in occupational settings without an RPP. Therefore, resources and methods are needed to enhance the identification of respirators expected to provide an adequate seal for each unique user. Moreover, these methods need to be low-cost and easily employed in low-resource settings without technical expertise.

In response to the recent pandemic, the U.S. Strategic National Stockpile (SNS) had distributed a variety of NIOSH Approved N95 FFRs to pharmacy partners across the country. This SNS distribution is intended to provide high-quality protection at no cost to the American people (ASPR 2022). This paper describes the methods and results from a quantitative fit evaluation study of SNS-distributed FFRs using a complete set of ISO static advanced manikin headforms (StAH), which were developed based on a 2003 anthropometric survey conducted by NIOSH to be representative of approximately 95% of the U.S. worker population (Zhuang and Bradtmiller 2005). This paper also describes an anthropometric analysis conducted, using NIOSH's 2003 anthropometric survey database, to evaluate the usability of a predictive model that determines an individual's StAH size from a limited set of facial dimensions to further identify N95 FFRs evaluated from the SNS that would have a higher probability of adequately sealing to their face.

Methods

Respirator selection

Twelve N95 FFR models distributed by the SNS were procured for testing. These models spanned across four different respirator manufacturers (3M [Saint Paul, MN, U.S.], Dräger [Lübeck, Germany], Honeywell [Charlotte, NC, U.S.], and Moldex [Culver City, CA, U.S.]) and varied in make (cup and flat-fold) and size (small, regular, medium/large, low profile, and 1-size only). The 3M FFR models included the 8110S (cup, small), 8210 (cup, regular), 8511 (cup, regular), 8511 P (cup, regular), AURA 9205+ (flat-fold, regular), and AURA 9210+ (flat-fold, regular). The Dräger and Honeywell FFR models were all one-size only and included the Dräger XPLORE 1750 (flat-fold) and the Honeywell DC300 and OneFit NBW95 (cup). It must be noted that since the time of this evaluation, the Honeywell OneFit NBW95 FFR (NIOSH Approval # TC-94A-4373) has been discontinued, and the NIOSH approval has been rescinded.

Three cup-shaped Moldex FFR models were evaluated, which included the 2200 (medium/large), 2201 (small), and the low profile (LP) 2207. Nine units of each of the 12 N95 FFR models obtained for evaluation had undergone one quantitative fit test for each of the five StAH sizes.

Headform fit evaluation test setup and procedure

A manikin fit factor (mFF) (Manikin Fit Factor (mFF) is an expression related to the amount of leakage measured through the face or neck seal of a respirator mounted to a manikin under specified airflow and environmental conditions. If the challenge to the seal is an airborne substance, it is the ratio of its airborne concentration outside the respirator divided by the concentration that enters the respirator through the seal (Janssen and McKay 2017).) refers to the amount of leakage measured through the face seal of a respirator being evaluated on a manikin headform (Janssen and McKay 2017). This differs from the traditional fit factor in that it involves a respirator being mounted onto a manikin headform that can perform certain breathing exercises of a quantitative fit test. Static, quantitative fit testing was completed using each of the five sizes of StAHs (i-bodi LTD, Buckinghamshire, England) with compressible silicone elastomer skin; these sizes include Small, Medium, Large, Short-Wide, and Long-Narrow (Figure 1). Zhuang et al. 2010 provide a more detailed overview of the StAHs used in this study (Zhuang et al. 2010). N95 FFRs were placed on the StAHs following the manufacturers' guidance for correct head strap placement and adjustment of the bendable nose clip (for models so equipped).

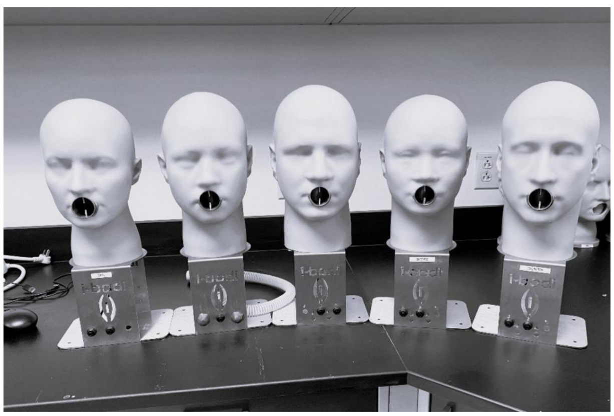

Figure 1. The complete set of ISO StAHs used for testing (five sizes from left to right: small, medium, large, short-wide, long-narrow).

StAHs were placed in a test chamber (48″ × 24″ × 54″ (L × W × H)) containing generated, charge-equilibrated, NaCl aerosol at a concentration of 11,000 ± 1,000 particles/cm^3 (Figure 2). Manikin fit was evaluated for the StAH under cyclic breathing conditions. A tube extended from the bottom of the StAH and was connected to a 3-L inflatable (non-latex, powder-free) bladder inside an isolated, airtight, acrylic breathing lung. This configuration prevents any particles potentially generated by the breathing simulator from entering the breathing zone of the StAH. A port on the breathing lung was connected to a breathing simulator (model: BRS; Koken Ltd., Japan) operating at 23 L/min (1-L tidal volume x 23 breaths/min). This breathing condition was selected to simulate a light-to-moderate breathing effort being outputted from the StAH, which can be related to a work rate similar to activities such as a healthcare

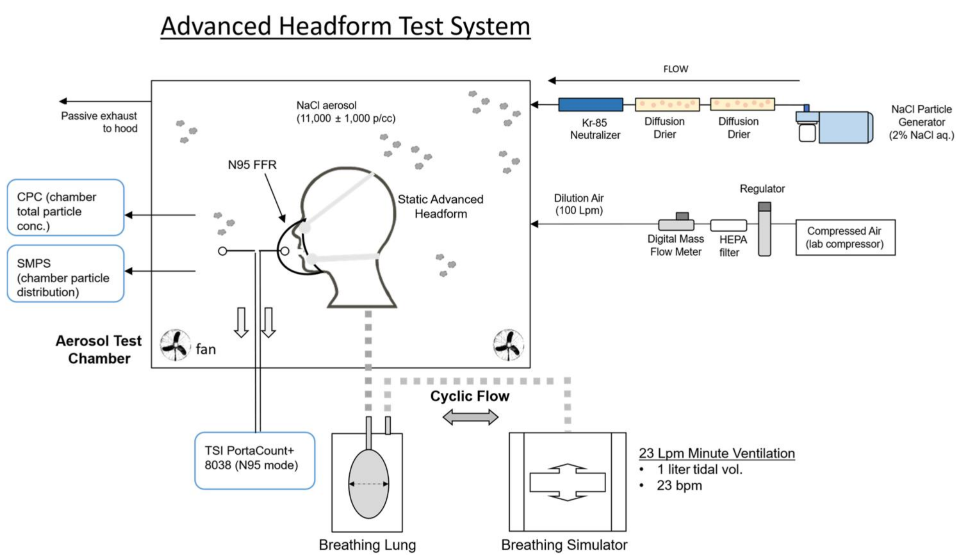

Figure 2. StAH quantitative fit evaluation test system.

worker performing patient care tasks, someone performing brisk walking, or standard janitorial tasks.

A PortaCount Pro + model 8038 Respirator Fit Tester (TSI, Inc., Shoreview, MN) operating in the N95-enabled mode was used. The PortaCount utilizes condensation nuclei counting technology to count individual particles and calculate quantitative mFFs. Fit evaluations were first performed utilizing a "screening" method that was developed to quickly evaluate the seal of the N95 FFR to the face of the StAH prior to beginning the official manikin fit evaluation. This screening method first involved placing the N95 FFR unit onto the StAH and making any necessary adjustments to the head straps and nose clip (for models so equipped). Then, with the breathing machine operating at 23 L/min, the test operator observed a graphic display of real-time mFFs on the PortaCount screen in real-time mode, where FFs are outputted approximately once per second. If the real-time outputs showed 10 consecutive FFs $\geq$ 100, then the test operator began the official manikin fit evaluation. If not, the N95 FFR was removed, replaced, adjusted, and reevaluated in real-time FF mode. If the screening criterion was not met after three attempts of the screening method, the official manikin fit evaluation was started after the third attempt.

The official manikin fit evaluation was a 2-min and 10-sec breathing exercise test, with the breathing machine operating at 23 L/min. This test outputted an overall mFF result for the N95 FFR unit and StAH combination, where the overall passing FF criterion is $\geq$ 100 when testing N95 FFRs.

Mean mFFs and standard error values for each headform size and N95 FFR model combination, as well as mean fit test passing rates and standard error values, were calculated. For each N95 FFR model, a Tukey HSD test was also performed to investigate pairwise differences in fit across the five StAH sizes. The familywise error rate was controlled at $\alpha = 0.05$.

Predictive model analysis

Anthropometric analysis

Utilizing the results from the quantitative headform fit evaluation, an effort to identify suitable respirator makes, models, and sizes based on an individual's unique facial dimensions was initiated, which first requires understanding the relationship between the five sizes of ISO headforms and the head and face sizes of individuals. While an individual's headform size can be determined using the ten principal component analysis (PCA) panel facial measurements, it is anticipated that it would be challenging for users to perform several of these measurements without specialized calipers or training. Therefore, using the 2003 NIOSH Anthropometric Database from Zhuang and Bradtmiller (2005), researchers sought to (1) identify a limited set of facial measurements that were sufficiently predictive of headform size and (2) fit and evaluate a predictive model using the select facial measurements to predict headform size (Zhuang and Bradtmiller 2005). The predictive model, if found to be sufficiently accurate, could then be used alongside the quantitative headform fit results as a tool to accurately predict an individual's headform size and identify respirators that are most likely to seal to their face.

To fit and assess the predictive model, two datasets were created using the NIOSH Anthropometric Database. Subjects missing one or more of the 19 facial dimensions (excluding neck circumference, which was deemed to be irrelevant to respirator fit) or headform size determinations were excluded (186 subjects removed; 3,811 subjects remained). Stratified random sampling with respect to sex, race group, and age group was then employed to ensure that these datasets were representative of the U.S. general population, allowing model performance statistics to be generalized. This resulted in a "training set" consisting of 913 subjects, which was used to fit the model, and a "test set" consisting of 289 subjects, which was used to assess the fitted model's predictive performance. The combined size of the training and test sets was chosen to maximize the number of subjects used for training and testing while still respecting the constraints imposed by sample stratification. There was no overlap of subjects between the training and test sets. Table 1 gives the frequency of each headform size in the training and test sets.

Since headform size is a categorical variable with more than two categories, a multinomial logistic regression modeling approach was taken. Through informed expert opinion regarding the difficulty of measuring each facial dimension and literature review, six facial dimension variables were ultimately selected for inclusion in the model: bizygomatic breadth (face

Table 1. Frequency of each headform size in the training and test sets.

Headform size	Training set	Test set
Small	99 (10.8%)	39 (13.5%)
Short-wide	124 (13.6%)	37 (12.8%)
Medium	485 (53.1%)	153 (52.9%)
Long-narrow	110 (12.0%)	29 (10.0%)
Large	95 (10.4%)	31 (10.7%)
Total number of subjects	913	289

width), menton-sellion length (face length), minimum frontal breadth, nose protrusion, nose breadth, and subnasale-sellion length. The model was then fitted to the training set using these six variables. The test set was used to estimate model accuracy in predicting a single headform size output, along with two headform size-specific measures of predictive performance, positive predictive value (PPV) (PPV. The model's PPV for a given headform size *H* is the proportion of subjects predicted to be *H* that are truly *H*.), and negative predictive value (NPV) (NPV. The model's NPV for a given headform size *H* is the proportion of subjects not predicted to be *H* that are truly not *H*.). Exact 95% confidence intervals were also computed for all headform size-specific estimates. A confusion matrix analysis of the predictive model was also conducted, which compared the model's predicted headform size classifications to the actual headform sizes pre-determined within the test set.

Performance comparison tests

To better contextualize the predictive performance of the model, we compared this model, which uses six of ten PCA panel facial dimensions to output a single headform size determination, to one that fits with all ten PCA panel facial dimensions. A principal components multinomial logistic regression model to the set of six facial dimensions in the model was also performed for comparison purposes. This was done to see if the use of principal components of the facial dimension variables rather than the facial dimension variables themselves would increase the predictive performance of the model.

Predictive sensitivity analysis

To account for a potential increase in measurement error, a sensitivity analysis was performed on the predictive model in which statistical noise was artificially added to the facial measurements in the test set. Random normal variates were added to each of the subjects' measurements in the test set so that the standard deviation of each facial dimension was increased by a multiplicative factor of $\sqrt{2}$ (i.e., the sample variance of each facial dimension was doubled).

Results

Headform fit evaluation

A total of 540 quantitative fit tests were performed within this evaluation (9 units × 12 N95 FFR models × 5 StAH sizes = 540 total tests). The mean mFF, mean fit test passing rate, and standard error results for each StAH size and N95 FFR model combination are presented in Table 2. Also displayed in Table 2 are Tukey HSD results for each N95 FFR model in

Table 2. Mean mFF, mean fit test passing rate, and standard error results for each headform size and N95 FFR combination ($n = 9$ respirator units per mean calculation), with pairwise comparisons (Tukey HSD test) of mean mFF by N95 model.

	Small		Short-wide		Medium		Long-narrow		Large	
	Mean mFF	SE	Mean mFF	SE	Mean mFF	SE	Mean mFF	SE	Mean mFF	SE
3M 8110S	94^{bc}	28	159^{ab}	16	172^{a}	16	64^{c}	21	22^{c}	7
3M 8210	95^{bc}	23	163^{ab}	15	183^{a}	16	201^{a}	0	73^{c}	26
3M 8511	185^{a}	9	201^{a}	0	190^{a}	7	177^{a}	10	201^{a}	0
3M 8511 P	171^{a}	13	178^{a}	10	177^{a}	16	179^{a}	12	201^{a}	0
3M AURA 9205+	105^{a}	19	138^{a}	16	137^{a}	12	152^{a}	11	118^{a}	5
3M AURA 9210+	142^{a}	13	172^{a}	11	156^{a}	19	111^{a}	24	176^{a}	12
Dräger XPLORE 1750	17^{b}	4	151^{a}	20	148^{a}	17	193^{a}	8	154^{a}	15
HONEYWELL DC300	3^{a}	1	65^{a}	20	51^{a}	16	52^{a}	24	15^{a}	5
HONEYWELL OneFit NBW95	4^{b}	1	166^{a}	22	115^{a}	30	132^{a}	23	21^{b}	7
MOLDEX 2200	46^{b}	13	187^{a}	5	153^{a}	16	174^{a}	14	59^{b}	20
MOLDEX 2201	22^{c}	11	201^{a}	0	201^{a}	0	131^{b}	29	88^{bc}	21
MOLDEX LP 2207	7^{b}	2	171^{a}	19	129^{a}	24	44^{b}	9	5^{b}	2
	Mean Rate	SE	Mean Rate	SE	Mean Rate	SE	Mean Rate	SE	Mean Rate	SE
3M 8110S	0.44	0.18	0.78	0.15	0.89	0.11	0.33	0.17	0.00	0.00
3M 8210	0.33	0.17	0.89	0.11	0.89	0.11	1.00	0.00	0.22	0.15
3M 8511	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
3M 8511 P	1.00	0.00	1.00	0.00	0.89	0.11	1.00	0.00	1.00	0.00
3M AURA 9205+	0.33	0.17	0.78	0.15	0.89	0.11	1.00	0.00	1.00	0.00
3M AURA 9210+	0.89	0.11	1.00	0.00	0.78	0.15	0.56	0.18	1.00	0.00
Dräger XPLORE 1750	0.00	0.00	0.78	0.15	0.78	0.15	1.00	0.00	0.89	0.11
HONEYWELL DC300	0.00	0.00	0.22	0.15	0.11	0.11	0.22	0.15	0.00	0.00
HONEYWELL OneFit NBW95	0.00	0.00	0.78	0.15	0.56	0.18	0.67	0.17	0.00	0.00
MOLDEX 2200	0.22	0.15	1.00	0.00	0.89	0.11	0.89	0.11	0.11	0.11
MOLDEX 2201	0.11	0.11	1.00	0.00	1.00	0.00	0.56	0.18	0.33	0.17
MOLDEX LP 2207	0.00	0.00	0.89	0.11	0.78	0.15	0.11	0.11	0.00	0.00

Note: Two headforms with the same letter grouping means their mean mFF results for the specific N95 FFR being evaluated are not significantly different.

the form of superscripted letter groupings. The Tukey HSD test generates groupings indicating groups of headforms on which the mean fit factor was comparable (i.e., for any two StAHs in the same Tukey grouping, the difference in mean mFF results was not found to be statistically significant).

For five of the 12 N95 FFR models in this study (3M 8511, 3M 8511 P, 3M AURA 9205+, 3M AURA 9210+, Honeywell DC300), there were no statistically significant differences in mean mFF results across all StAHs (adjusted p-values were greater than 0.05 for all pairwise comparisons for a given model). For four of these five N95 FFR models (the 3M models), the mean mFF for each of the headforms was above the passing threshold of 100, suggesting uniformly adequate fit on all headform sizes. For the fifth N95 FFR model (Honeywell DC300), the mean mFF was well below 100 on all five headforms, suggesting uniformly poor fit across all headform sizes.

The Dräger XPLORE 1750 model exhibited a high and comparable level of fit across the Long-Narrow, Large, Short-Wide, and Medium headforms. It exhibited significantly worse fit on the Small headform (adjusted p-values in pairwise comparisons of the Small headform with the other sizes were all less than 0.05). For the three Moldex N95 FFRs in this study (models 2200, 2201, LP 2207), the mean mFF was significantly higher on the Short-Wide and Medium headforms compared to the Small and Large sizes. The mean mFF results were variable across the three Moldex N95 FFRs when evaluated on the Long-Narrow headform. A similar pattern was observed with the 3M 8210, 3M 8110S, and Honeywell OneFit NBW95 FFR models.

Table 3. Headform size-specific estimates of predictive performance for the multinomial logistic regression model fit to the following six variables: bizygomatic breadth, menton-sellion length, minimum frontal breadth, nose protrusion, nose breadth, and subnasale-sellion length.

Headform size	Accuracy	PPV	NPV
Small	0.95 (0.92, 0.97)	0.83 (0.67, 0.94)	0.96 (0.93, 0.98)
Short-wide	0.96 (0.93, 0.98)	0.86 (0.71, 0.95)	0.98 (0.95, 0.99)
Medium	0.87 (0.83, 0.91)	0.88 (0.82, 0.93)	0.86 (0.79, 0.91)
Long-narrow	0.97 (0.94, 0.99)	0.83 (0.65, 0.94)	0.98 (0.96, 1.00)
Large	0.93 (0.90, 0.96)	0.66 (0.48, 0.81)	0.97 (0.94, 0.99)

All estimates are accompanied by 95% confidence intervals.

Predictive model analysis

Anthropometric analysis

Multinomial logistic regression modeling results are shown in Table 3. The model achieved an overall predictive accuracy of 85.1% [95% confidence interval (CI) = (79.4%, 88.1%)] and high accuracy rates (accuracy $\geq$ 80%) for each headform size (e.g., "Small" versus the rest of the sizes). The PPV is reasonably high (PPV $\geq$ 80%) across all headform sizes except for the Large headform size (PPV = 66%), although these estimates should be interpreted with caution, given the large confidence intervals for these estimates. Additionally, the NPV is reasonably high (NPVs $\geq$ 80%) across all headform sizes. Overall, across most headform sizes, the generally high PPVs and NPVs suggest that the model can correctly identify whether individuals are or are not of a given headform size.

The analysis of the confusion matrix (Table 4) reveals that when subjects were misclassified, in many cases, they were misclassified into "adjacent" sizes. For instance, of the 31 Large subjects in the test set, 7 were incorrectly predicted as either Long-Narrow or Medium, and 1 as Short-Wide, with none being incorrectly predicted as Small. Similarly, of the 39 Small subjects in the test set, 8 were incorrectly predicted as Short-Wide or Medium, and 1 as Long-Narrow, with none being incorrectly predicted as Large.

Performance comparison tests

Comparison results for this model to one fit with all 10 PCA panel facial dimensions in predicting a single headform size are shown in Table 5. As expected, the model using all PCA panel facial dimensions was more accurate, having an overall predictive accuracy of 91.7% [95% CI = (87.9%, 94.6%)]. Similarly, the model using all ten dimensions had comparable or higher PPVs and NPVs across all headforms. Although it exhibits better predictive performance, the model with all ten PCA dimensions requires more

Table 4. Confusion matrix for the predictive model when evaluated on the test set.

		Actual headform size				
		Small	Short-wide	Medium	Long-narrow	Large
Predicted headform size	Small	30	1	5	0	0
	Short-wide	2	31	2	0	1
	Medium	6	5	134	2	5
	Long-narrow	1	0	2	25	2
	Large	0	0	10	2	23

Table 5. Headform size-specific estimates of predictive performance for the multinomial logistic regression model fit to all ten PCA facial dimensions, for comparison only.

Headform size	Accuracy	PPV	NPV
Small	0.97 (0.95, 0.99)	0.90 (0.76, 0.97)	0.98 (0.96, 1.00)
Short-wide	0.97 (0.94, 0.99)	0.91 (0.76, 0.98)	0.98 (0.95, 0.99)
Medium	0.93 (0.90, 0.96)	0.92 (0.87, 0.96)	0.95 (0.89, 0.98)
Long-narrow	0.98 (0.95, 0.99)	0.87 (0.69, 0.96)	0.99 (0.97, 1.00)
Large	0.98 (0.96, 0.99)	0.96 (0.82, 1.00)	0.98 (0.96, 1.00)

measurements, including several that are more challenging to perform. The results of the comparison test, looking at a principal components multinomial logistic regression model of the set of six facial dimensions in the current model, exhibited highly similar predictive performance overall and within each headform size compared to the original model.

Sensitivity analysis

The sensitivity analysis conducted on the test set containing additional statistical noise showed that the predictive model's accuracy decreased to only 65.4% [95% CI = (59.6%, 70.9%)]. The standard deviations of each facial dimension in the test set, with this additional noise added, increased by at least 1 mm when compared to the standard deviations without additional noise. These differences in standard deviation values (mm) for each facial dimension, with no noise and with noise added, respectively, include: bizygomatic breadth (7.74, 10.89), menton-sellion length (7.36, 9.94), minimum frontal breadth (5.33, 7.48), nose protrusion (2.72, 3.82), nose breadth (4.48, 5.96), subnasale-sellion length (3.69, 5.06). With the noisy data, the model's PPV across all headform sizes is substantially decreased, suggesting that its predictive performance is sensitive to increased measurement error.

These sensitivity analysis results, along with the realization that misclassifications tended to occur when subjects were classified into "adjacent" sizes, inspired a modification to the predictive model to make it potentially more useful for real-world use. Instead of returning a single headform size prediction, the model was modified to output a set of headform sizes with a high estimated posterior probability (at least 95%) of containing an individual's true headform size, a well-recognized approach that can be valuable so long as the size of the predicted set is sufficiently small (Mortier et al. 2021). N95 FFR models fitting all headform sizes in the set would then be considered appropriate to output as a prediction to the individual. This set might contain a single headform size if the model places high posterior probability on a single headform size, but it might also contain two or more headform sizes if the model places high posterior probability on two or more sizes.

The modified predictive model developed was found to have a 98% overall accuracy rate, with 284 out of 289 test subjects' true headform sizes included in their prediction sets. Since large prediction sets could trivially contain an individual's true headform size (i.e., a set containing all five headform sizes always contains the individual's true headform size) but severely limit the number of respirators that could be predicted for the individual's unique facial dimensions, the sizes of the test subjects' prediction sets were evaluated. It was found that the majority of test subjects (92%) had a prediction set consisting of one or two headform sizes (128 and 137 subjects, respectively); 24 subjects had a prediction set consisting of three headform sizes, and no subjects had a prediction set consisting of four or more sizes. The modified model was also found to have higher overall accuracy compared to the original model in a sensitivity analysis—81% compared to the original model's accuracy rate of 65.4%. Additional details on this modification in comparison to the original model are provided in Appendix A.

Discussion

This study's quantitative fit evaluation results on 12 NIOSH Approved N95 FFR models distributed from the SNS showed that various N95 FFR models fit certain headform sizes better than others. These findings bring to light how complex achieving respiratory protection can be for a respirator user, since multiple factors influence fit results, including the user's unique head and face size, the N95 FFR's make, model (including the style of the model, such as cup and flat fold), and respirator size. Some N95 FFR models evaluated were developed as "one-size only" (i.e., Dräger, Honeywell), while other models evaluated are designated sizes as part of a series consisting of different sizes of N95 FFRs designed by the manufacturer (i.e., 3M, Moldex). These designated sizes may be designed to fit a more specific range of face shapes and sizes.

Only four out of the 12 N95 FFR models evaluated within this study, specifically the 3M 8511, 3M 8511 P, 3M AURA 9205+, and 3M AURA 9210+, could achieve mean mFF results above the passing threshold of 100 for all headform sizes. The Honeywell DC300

FFR resulted in significantly lower mean mFF results, well below 100, for all headform sizes. The remaining N95 FFR models evaluated displayed variable fit results across the five different headform sizes. If the Small and Large headforms are considered “extreme” sizes, then it may seem reasonable that poorer average fit was observed on these sizes compared to the Medium- and Short-Wide headforms. Manufacturers of NIOSH Approved N95 FFRs are currently not required to design their respirators to fit any specific facial sizing or anthropometry. This means that it can be difficult to know whether an N95 FFR will fit a user without having an idea of how the respirator fit correlates with head and face size. This difficulty is exacerbated when not performing a fit evaluation or an OSHA-accepted fit test prior to respirator use.

Within the PCA panel, the first principal component primarily captures the overall size of the face, while the second principal component captures key nasal features (Zhuang et al. 2007). Other researchers have noted the importance of nasal features in terms of respirator fit. A recent study, conducted by Khairul Hasni et al. (2023), performed quantitative fit testing of 135 subjects within Malaysia’s ethnically diverse population, including both male and female subjects. The authors specifically evaluated the design and fit of six NIOSH Approved FFR models of varying shapes/styles (i.e., cup, trifold, duckbill). They also investigated the association that sex and facial dimensions had on fit test results, utilizing face and anthropometric measurements from a 2020 Malaysian survey involving over 3,000 individuals (Khairul Hasni et al. 2023). The study found that certain FFR models fit a specific sex better than others, where larger facial sizes were mainly correlated with males and smaller facial sizes with female subjects. In addition, the study identified specific facial dimensions that appeared to positively or negatively influence fit factor results for various FFR models. Nose protrusion, nasal root breadth, and subnasale-sellion length were reported to have significantly positive effects on fit results (Khairul Hasni et al. 2023). This information supported the utilization of the simplified set of six facial dimensions to be considered within the predictive model when determining a user’s headform size classification, which were selected to best represent the overall size of the face and key nasal features that are relatively easy to measure and are important to consider for determining respirator fit.

Although further validation efforts are necessary to better understand the correlations between headform and human subject respirator fit results and prepare for real-world factors that could affect predictive accuracy, the modified predictive model developed in the current study (98% overall accuracy rate) was anticipated to be better equipped to handle measurement uncertainty and reduce real-world risk of recommending incorrectly-sized respirators for individuals compared to the original model. These results and user error considerations could further support the use of this predictive model within future tooling and respirator prediction and selection resources.

This paper’s findings help to increase respirator fit knowledge and contribute to the understanding of how fit correlates with varying head sizes and facial dimensions. Health equity capabilities could also be enhanced if future work demonstrates acceptable performance across diverse user populations by implementing an N95 FFR model identification tool that takes into account areas of the NASEM’s equity principle to include a population’s variation in facial size in achieving respiratory protection needs, utilizing this study’s predictive modeling approach (NASEM 2022). The utilization of the developed predictive modeling could also be done to protect respirator wearers in the form of a tool that requires minimal resources and can be accessible to many users through public computers and internet resources, without the need for a personal smart device. Respiratory protection compliance could also be enhanced using this approach as a form of guidance for users on the importance of a well-fitting respirator and how FFRs could provide optimal levels of protection if used and worn properly.

Lastly, respirator supply strategies and resource optimization could also be improved with the use of this study’s developments and findings to better prepare for respiratory health emergencies. An N95 FFR model identification tool developed utilizing this approach, although not intended as a tool to replace an OSHA-accepted fit test, could make the overall fit test process more efficient and less burdensome in relation to time and respirator supplies used.

Limitations

Some limitations involved with the quantitative fit evaluation conducted within this study include using StAHs as opposed to human subjects performing simulated workplace exercises, although the StAHs were developed based on actual human subject anthropometric data. It is possible that incorporating human subject movement into the fit evaluation could impact mFF measurements by causing potential shifting and breakage of the face seal, simulating a more real-world evaluation. Another limitation is that only one respirator type (N95 FFR) and limited models and sizes distributed by the SNS at the time of this study were evaluated. Future research could expand on this evaluation to include other respirator types and models that workers and the public may be using and have better access to.

Regarding the modified predictive model, a significant limitation includes the uncertain amount of measurement error that could be seen in real-world settings when measuring facial dimensions for use within the model. The sensitivity analysis performed within this study (i.e., adding normally distributed noise) does not simulate how measurement error might arise during facial measurement. Further evaluation and validation of the predictive model's functionality, usability, and performance with human subjects is necessary before real-world deployment. In addition, the development of proper training and guidance for individuals to measure these facial dimensions is critical to mitigate the potential for measurement error throughout these validation efforts. These limitations currently hinder this study's findings from providing a more complete understanding of respirator fit as it relates to anthropometric facial dimensions to achieve respiratory protection.

Conclusion

This quantitative fit evaluation study's findings indicate that certain N95 FFR models distributed by the SNS fit certain headform sizes better than others, revealing how variable fit can be across different makes, models (and their variable styles), and sizes. The modified predictive model developed and evaluated in this study had a high overall prediction accuracy of 98% when identifying the headform sizes of the test set subjects, showing encouraging results that could be taken into further human subject validation and field testing to better understand measurement variability and user experiences. Once this predictive modeling approach is sufficiently validated for real-world use, it could be further implemented and used as a respirator identification tool. This type of resource and tooling could make the official fit testing process more efficient and less burdensome in relation to time and respirator supplies used, and would enable the public and workers without an RPP to use respirator models more likely to fit and, thus, adequately protect them from hazards.

Attribution

N95 and NIOSH Approved are certification marks of the U.S. Department of Health and Human Services (HHS) registered in the United States and several international jurisdictions.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention. Mention of a company or product name does not constitute endorsement by NIOSH or the CDC.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Brooke Vollmer http://orcid.org/0000-0001-5894-1469
Michael S. Bergman http://orcid.org/0000-0002-2742-6197
Harold Boyles http://orcid.org/0000-0002-1454-8355
Jordan Meyers http://orcid.org/0000-0001-5866-5919
Nora Y. Payne http://orcid.org/0000-0003-4935-6465
Jonisha Pollard http://orcid.org/0000-0001-9282-9120
Ziqing Zhuang http://orcid.org/0000-0002-0771-5274

Data availability statement

The data set will be available on the NIOSH Data and Statistics Gateway once cleared by NIOSH.

References

ASPR. 2022. Strategic national stockpile makes high-quality masks available to Americans. U.S. Department of Health and Human Services. https://aspr.hhs.gov/SNS/Pages/SNS-Hgh-Quality-Masks.aspx.

BLS. 2021a. Labor force characteristics by race and ethnicity, 2020. BLS Reports. https://www.bls.gov/opub/reports/race-and-ethnicity/2021/home.htm.

BLS. 2021b. Highlights of women's earnings in 2020. BLS Reports. https://www.bls.gov/opub/reports/womens-earnings/2020/pdf/home.pdf.

CDC. 1998. Laboratory performance evaluation of N95 filtering facepiece respirators. MMWR. 47(48):1045–1049.

Chopra J, Abiakam N, Kim H, Metcalf C, Worsley P, Cheong Y. 2021. The influence of gender and ethnicity on facemasks and respiratory protective equipment fit: a systematic review and meta-analysis. BMJ Glob Health. 6(11):e005537. doi: 10.1136/bmjgh-2021-005537.

Coffey CC, Campbell DL, Myers WR, Zhuang Z, Das S. 1998. 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(12):852–861. doi: 10.1080/15428119891011027.

Coffey CC, Campbell DL, Zhuang Z. 1999. Simulated workplace performance of N95 respirators. Am Ind Hyg Assoc J. 60(5):618–624. doi: 10.1080/00028899908984481.

Coffey CC, Lawrence RB, Campbell DL, Zhuang Z, Calvert CA, Jensen PA. 2004. Fitting characteristics of eighteen N95 filtering-facepiece respirators. J Occup Environ Hyg. 1(4):262–271. doi: 10.1080/15459620490433799.

Colton CE. 1999. Filtering facepieces: study supports need for fit testing. 3M Job Health Highlights. 17(2):1–4.

Federal Register. 1995. 42 CFR Part 84—approval of respiratory protective devices. https://www.ecfr.gov/current/title-42/part-84.

Federal Register. 1998. Final rule. Respiratory protection. Occupational Safety and Health Administration. 29 CFR Parts 1910 and 1926.

Janssen L, McKay R. 2017. Respirator performance terminology. J Occup Environ Hyg. 14(12):D181–D183. doi: 10.1080/15459624.2017.1359018.

Khairul Hasni NA, Ismail R, Muhamad Robat R, Mohamad N, Suib FA, Pahrol MA, Mahmud H, Osman B, Lim YC, Seman Z, et al. 2023. The effect of N95 designs on respirator fit and its associations with gender and facial dimensions. PLoS One. 18(11):e0288105. doi: 10.1371/journal.pone.0288105.

Manganyi J, Wilson KS, Rees D. 2017. Quantitative respirator fit, face sizes, and determinants of fit in South African diagnostic laboratory respirator users. Ann Work Expo Health. 61(9):1154–1162. doi: 10.1093/annweh/wxx077.

Mortier T, Wydmuch M, Dembczyński K, Hüllermeier E, Waegeman W. 2021. Efficient set-valued prediction in multi-class classification. Data Min Knowl Disc. 35(4):1435–1469. doi: 10.1007/s10618-021-00751-x.

NASEM. 2022. Frameworks for protecting workers and the public from inhalation hazards. Washington (DC): The National Academies Press. doi: 10.17226/26372.

NIOSH. 2021. NIOSH Approved particulate filtering facepiece respirators. Centers for Disease Control and Prevention. https://www.cdc.gov/niosh/npptl/topics/respirators/disp_part/default.html#:~:text=N95%20%E2%80%93%20Filters%20at%20least%2095,Not%20resistant%20to%20oil.&text=Surgical%20N95%20%E2%80%93%20A%20NIOSH%2Dapproved,least%2099%25%20of%20airborne%20particles.

Reponen T, Lee SA, Grinshpun SA, Johnson E, McKay R. 2011. Effect of fit testing on the protection offered by N95 filtering facepiece respirators against fine particles in a laboratory setting. Ann Occup Hyg. 55(3):264–271. doi: 10.1093/annhyg/meq085.

Spies A, Wilson KS, Ferrie R. 2011. Respirator fit of a medium mask on a group of South Africans: a cross-sectional study. Environ Health. 10(1):17. doi: 10.1186/1476-069X-10-17.

Zhang M, Hackett L, Smith J, Pritchard Z, Casey M, Low C, Buntine P. 2023. Striving to be the fittest: quantitative P2/N95 respirator fit test results among hospital staff during the COVID-19 pandemic. Antimicrob Steward Healthc Epidemiol. 3(1):e233. doi: 10.1017/ash.2023.503.

Zhuang Z, Benson S, Viscusi D. 2010. Digital 3-D headforms with facial features representative of the current US workforce. Ergonomics. 53(5):661–671. doi: 10.1080/00140130903581656.

Zhuang Z, Bradtmiller B. 2005. Head-and-face anthropometric survey of U.S. respirator users. J Occup Environ Hyg. 2(11):567–576. doi: 10.1080/15459620500324727.

Zhuang Z, Bradtmiller B, Shaffer RE. 2007. New respirator fit test panels representing the current U.S. civilian work force. J Occup Environ Hyg. 4(9):647–659. doi: 10.1080/15459620701497538.

Zhuang Z, Coffey CC, Jensen PA, Campbell DL, Lawrence RB, Myers WR. 2003. Correlation between quantitative fit factors and workplace protection factors measured in actual workplace environments at a steel foundry. AIHA J. 64(6):730–738. doi: 10.1202/475.1.

Appendix A

In the original multinomial logistic regression model, predictions are based on five estimated posterior probabilities, corresponding to the five headform sizes. Each posterior probability represents the model's estimate of the probability that the individual has the given headform size based on their six facial dimension measurements. The model's headform size prediction for an individual is simply the size corresponding to the largest posterior probability. The modified model predicts as follows: First, the posterior probabilities are ordered from largest to smallest; then, the model returns the headform sizes corresponding to the first k posterior probabilities such that their cumulative sum is at least 0.95 (95%). Table A1 shows the predicted headform sets as a result of using the modified model, where 92% of the test set subjects received a prediction set consisting of just one or two headform sizes, and none received a prediction set of four or more. These prediction sets outputted by the modified model were found to be sufficiently small relative to the total number of headform sizes (five).

Table A2 demonstrates the real-world value of this approach by presenting a hypothetical example comparing predictions made by the original and modified models for a subject who is, in truth, Short-Wide. The original model would incorrectly predict the subject to be Small; this subject would then be recommended to use respirators found to fit the Small headform. The modified model would predict the subject to be either Small or Short-Wide; this subject would be recommended to use respirators found to fit both these headform sizes. Although the first set of recommended respirators may contain respirators that fit Short-

Table A1. Prediction set outcomes from the modified model in relation to the total subjects within the test set.

Size of prediction set	Percentage of test set subjects (%)
1	44.3
2	47.4
3	8.3
4	0
5	0
Total	100

Table A2. Posterior probabilities for the five StAH sizes and sets of predicted headform sizes for two hypothetical subjects.

Subject	Small	Short-wide	Medium	Long-narrow	Large	Set of predicted headform sizes
1	0.05	0.95	0.00	0.00	0.00	(Short-Wide)
2	0.46	0.53	0.00	0.00	0.01	(Small, Short-Wide)

Wide subjects, it may also contain respirators that fit Small subjects only and thus be a poor fit for the subject in question. In contrast, all recommended respirators in the second set will fit Short-Wide subjects. By expanding the set of headform sizes in cases where the model does not place high posterior probability on a single headform size, the modified model may decrease the risk of recommending an ill-fitting respirator.