PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A prospective observational study of gender and ethnicity biases in respiratory protective equipment for healthcare workers in the Covid-19 pandemic
AUTHORS	Carvalho, Clarissa; Schumacher, Jan; Greig, Paul; Wong, Danny; El Boghdadly, Kariem

VERSION 1 – REVIEW

REVIEWER	Sozkes, Sarkis Namik Kemal Universitesi - Degirmenalti Kampusu, Biomedical Engineering Biomaterials Department
REVIEW RETURNED	09-Jan-2021

GENERAL COMMENTS	Proper fit anf face seal is a highly important factor in the performance of PPE in respiratory protection of healthcare providers. The Occupational Safety and Health Administration (OSHA) of the United States requires fit testing of respirators prior to extended use. For sufficient protection from aerosols good fit and seal is essential. Eventhough the importance of good seal and proper fit is highlighted in literature, there are limited sources and data investigating the quantitative fit testing. This original research is well designed and sharing valuable information for future research and development strategies of respirator manufacturing. Design and certification of the respirators should consider demographic charectersitics of face. Limitations of the study is discussed adequately, and future nationwide or international multicentral studies may be designed by the quidance of this paper. I believe this valuable paper will have many citations and be in great interest of occupational safety of healthcare providers. This manuscript may be published as presented without any revisions.

REVIEWER	Park, Sun Catholic University of Korea College of Medicine, Department of Internal Medicine
REVIEW RETURNED	25-Jan-2021

GENERAL COMMENTS	It is highly appreciated that authors made efforts in performing the fit tests for healthcare workers (HCWs) and analyzed the data in a timely manner during the COVID-19 pandemic. This is about the gender and ethnicity difference in fit test results of respiratory protective equipment (RPE) among healthcare workers at a single hospital in London during the COVID-19 pandemic.

	The study results reflect the current limitations of RPE in protective female healthcare workers or those of black, Asian and minority ethnic (BAME) background. The design of respirators was based on the result of the cohort where females and Asians were under-represented. The study is timely and appropriate to be addressed during the COVID-19 pandemic. However, this issue has been raised by previous studies. Also, there are several methodological issues to be addressed. First, the fit tests were repeated on HCWs who had failed to pass the test. Thus, the fit tests should not be treated as "independent" and they should be treated "dependent". Also, experiencing the first test may influence the results of the second tests as HCWs often learn how to participate in fit testing during the first test. Therefore, the test data needs to be analyzed accordingly. For example, separate analysis of the first test and the subsequent tests or mixed effect logistic regression may be options for the statistical method. Second, in this study, RPE with various designs were used for fit testing. However, there was no description about the specific designs, such as shape, size, head-band or early loop, etc. Also, the shapes and designs of re-suable RPE are largely different from disposable ones. Furthermore, different shapes of RPE may fit better for female HCWs or HCW of BAME background, which may confuse the readers whether HCWs did not find the better fitted RPE or the designs of RPE are not appropriate for female or non-white HCWs in general. The proportion of fit-test pass using the Design C model appeared to be generally lower in most HCWs whereas the design I better appeared to be well fitted for most HCWs. If HCWs with BAME background had been preferentially tested using the generally not-well-fitted RPE, it may be difficult to conclude that there are gender and ethnicity biases in RPE. For these reasons, the following is advised. 1) The detailed description of designs and filtering function of RPE used should be provided 2) In addition to the overall analysis, it is advised to analyze s on the gender and ethnicity differences among HCWs who were tested using the same design. Third, other factors which may influence the fit testing were not investigated nor adjusted for. Prior experience in respirator use or working experience of HCWs may influence the fit test results. So, demographics and work-related characteristics of HCWs should be provided and be adjusted for.

VERSION 1 – AUTHOR RESPONSE

Editor and Reviewers comments	Authors comments	Page number
Reviewer: 1 Dr. Sarkis So zkes,	Dear Dr Sozkes Thank you for reviewing our manuscript and for your comments. We agree this is of importance to the occupational safety of healthcare	

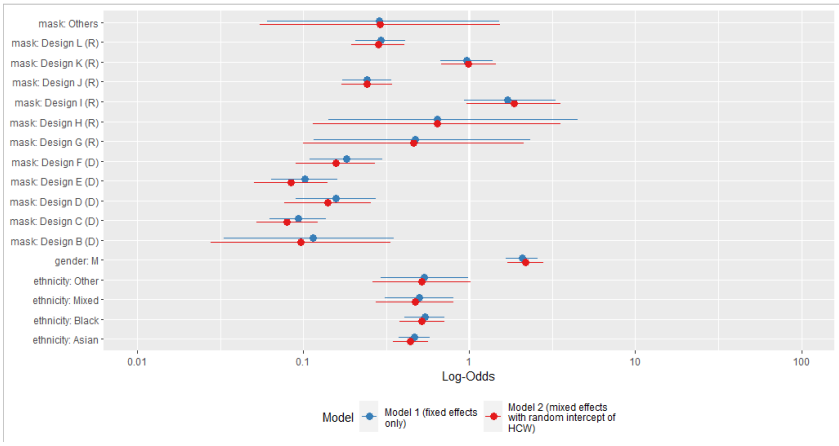
Namik Kemal Universitesi - Degirmenalti Kampusu Comments to the Author: Proper fit anf face seal is a highly important factor in the performance of PPE in respiratory protection of healthcare providers. The Occupational Safety and Health Administration (OSHA) of the United States requires fit testing of respirators prior to extended use. For sufficient protection from aerosols good fit and seal is essential. Even though the importance of good	workers and hope publication will raise awareness and trigger a change in mask design.	

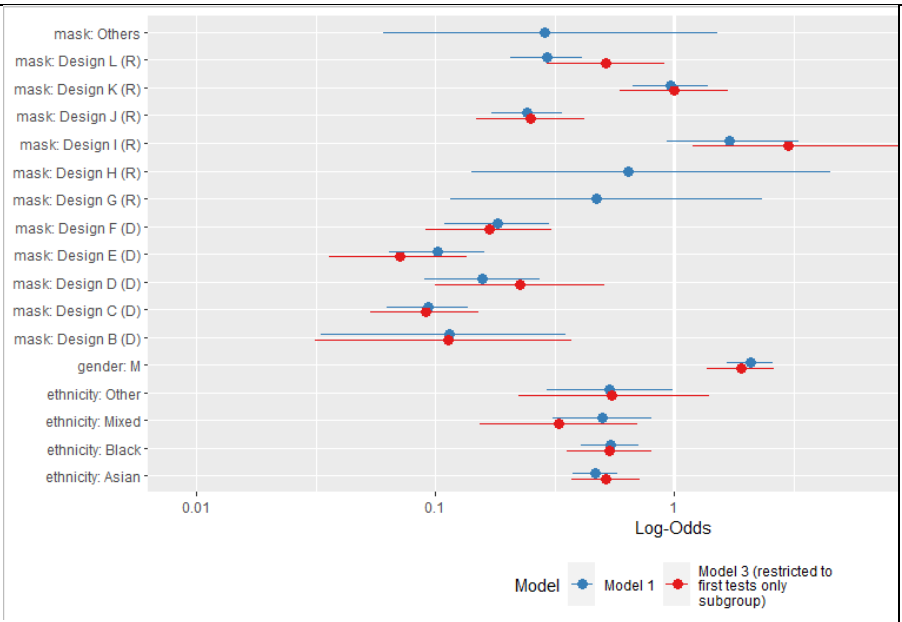
seal and proper fit is highlighted in literature, there are limited sources and data investigating the quantitative fit testing. This original research is well designed and sharing valuable information for future research and development strategies of respirator manufacturing. Design and certification of the respirators should consider demographic charectersitics of face. Limitations of the study is discussed adequately, and future nationwide		

or international multicentral studies may be designed by the quidance of this paper. I believe this valuable paper will have many citations and be in great interest of occupational safety of healthcare providers. This manuscript may be published as presented without any revisions.		
It is highly appreciated that authors made efforts in performing the fit tests for healthcare workers (HCWs) and analyzed the data in a timely	Dear Dr Park Thank you for reviewing our manuscript and for your detailed analysis.	

manner during the COVID-19 pandemic. This is about the gender and ethnicity difference in fit test results of respiratory protective equipment (RPE) among healthcare workers at a single hospital in London during the COVID-19 pandemic. The study results reflect the current limitations of RPE in protective female healthcare workers or those of black, Asian and minority ethnic (BAME) backgroun d. The design of respirators was based on the result of the cohort where females and Asians		

were under-represented. The study is timely and appropriate to be addressed during the COVID-19 pandemic.		
However, this issue has been raised by previous studies. Also, there are several methodological issues to be addressed.	Although this issue has been raised by previous articles, we did not see any with our data. We believe our research adds to the current literature and will be used to inform further research and subsequent improvements in RPE design.	

First, the fit tests were repeated on HCWs who had failed to pass the test. Thus, the fit tests should not be treated as "independent" and they should be treated "dependent".	Thank you for raising this possible statistical issue. On the reviewer's recommendation we re-ran a mixed effects logistic regression model with a random intercept for individual healthcare workers (i.e. considering each fit test as nested within individual HCWs). We found that while there was a small improvement to model fit—Akaike Information Criterion [AIC] reduced from 2682.5 for the model without the random intercept (Model 1) to 2675.4 for the model with the random intercept (Model 2)—this did not substantially alter our fixed effects estimates for the variables under investigation (see Figure A below). We also separately repeated the modelling, analysing only first tests (Model 3) and found no appreciable difference with the estimates obtained in our original analysis (see Figure B, below). We therefore argue that even if HCWs could learn how to participate in fit testing during the first test, this negligibly impacts on the results of the fit test, which depend much more greatly on the variables under consideration in our study. We have added text in the methods and in the results indicating that both a mixed effects model and a model fitted using only first fit attempts was also attempted. We have also included the Figures A and B (below) in supplementary materials.  **Figure A: Forest plot comparing the fixed effects point estimates and 95% confidence intervals the original model in the manuscript (Model 1) compared to a revised model fitted using mixed effects logistic regression with a random intercept for HCW (Model 2).**	7-9 + supplementary materials

	 Figure B: Forest plot comparing the fixed effects point estimates and 95% confidence intervals the original model in the manuscript (Model 1) compared to a revised model fitted using only subgroup data from first fit test attempts (Model 3).	
Also, experiencing the first test may influence the results of the second tests as HCWs often learn how to	The fit testing procedure is independent of previous experience. It is a measure and comparator of the number of particles inside and outside the mask in order to ascertain a leak. Learning the procedure does not increase the likelihood of a pass.	

participate in fit testing during the first test. Therefore, the test data needs to be analyzed accordingly.		
For example, separate analysis of the first test and the subsequent tests or mixed effect logistic regression may be options for the statistical method.	See above for additional explanation regarding statistical methods. We have additionally included the following texts within our Methods: “The following post hoc analyses were performed to assess the possibility that healthcare workers could learn to game the fit testing process and repeated testing of the same healthcare workers using different masks could render the tests not independent of each other: First we fitted mixed effects logistic regression models with random-intercepts for healthcare workers, assuming that tests were nested within healthcare workers; Second we repeated the original fixed-effects only logistic regression modelling with a subset of our dataset, only including data from first attempt fit tests. The results of the post hoc analyses were compared with our original findings and reported within the Supplementary Material.” Also in the Results section: “To assess the possibility of non-independence between tests performed on the same healthcare worker, an additional post hoc mixed-effects model fitted with random-intercepts for healthcare workers did not materially change our findings (Supplementary Material, Figure A). Similarly, a post hoc fixed-effects only model fitted using only data from first fit test attempts also did not materially change our findings (Supplementary Material, Figure B).”	7-9
Second, in this study, RPE with various designs were used for fit testing. However, there was no description about the specific designs,	Thank you for highlighting the need for descriptions of the different RPE designs. We did not want the readers to become focused on the respirator model and create a league of best-fit by manufacturer. A variety of different masks were used as consistency in supply was a particular issue in this pandemic. However each mask used in this study was CE marked and approved according to the European Norm EN149:2001 and is a verified N99 or FFP3 mask. We will supply data on the different mask designs as a supplementary table.	Supplementary material

such as shape, size, head-band or early loop, etc. Also, the shapes and designs of re-suable RPE are largely different from disposable ones.		
Furthermore, different shapes of RPE may fit better for female HCWs or HCW of BAME background, which may confuse the readers whether HCWs did not find the better fitted RPE or the designs of RPE are not appropriate for female or non-white HCWs in general. The proportion of fit-test pass using the Design C model		

appeared to be generally lower in most HCWs whereas the design I better appeared to be well fitted for most HCWs. If HCWs with BAME backgroun d had been preferential ly tested using the generally not-well-fitted RPE, it may be difficult to conclude that there are gender and ethnicity biases in RPE.		
For these reasons, the following is advised. 1) The detailed description of designs and filtering function of RPE used should be provided	This will be provided as a supplementary table	Supplemen ntary material

2) In addition to the overall analysis, it is advised to analyzes on the gender and ethnicity differences among HCWs who were tested using the same design.	The masks are all designed to a standard that uses measurements that are not representative of the current demographic of the healthcare workforce and the purpose of this study is to highlight this issue so further research can focus on updating the current measurements and design. In theory all the masks are designed to fit the workforce. However they do not fit as there is a discrepancy between the perceived and actual demographic of the workforce at a regulatory level. Identifying a mask that fits the demographic most appropriately is not what we wish to do as this may dilute the message that changes need to be made at a design and regulatory level.	
3) other factors which may influence the fit testing were not investigated nor adjusted for. Prior experience in respirator use or working experience of HCWs may influence the fit test results. So, demographics and work-related characteristics of HCWs should be provided and be adjusted	The fit testing procedure is objective and independent of previous experience. It is a measure and comparator of the number of particles inside and outside the mask in order to ascertain a leak. Learning the procedure does not increase the likelihood of a pass. Moreover, both fit-testing and use of respirators were not standardised institutional practice before the pandemic, and therefore the data collected are likely to represent de novo practitioners with little or no experience in respirators. This has now been added to the discussion: "Finally, previous experience with fit-testing was not accounted for, although quantitative fit-testing is an objective and independent of experience, and the use of respirators was generally poor prior to the pandemic so we assumed a homogeneous lack of experience in our cohort."	14

for.		