The influence of gender and ethnicity on facemasks and respiratory protective equipment fit: a systematic review and meta-analysis

Jagrati Chopra,1 Nkemjika Abiakam,2 Hansung Kim,3 Cheryl Metcalf,2 Peter Worsley,2 Ying Cheong1,4

ABSTRACT

Introduction Black, Asian and minority ethnic (BAME) people are disproportionately affected by COVID-19. Respiratory protective equipment (RPE) has conventionally been developed for a predominantly white male population that does not represent the healthcare workforce. The literature was reviewed to determine the protection offered to female and BAME users.

Methods Five databases were searched. Eligible studies related to respirator fit in the context of anthropometrics, gender and ethnicity. Meta-analysis was performed for gender-based anthropometric differences. A priori protocol registration was not performed.

Results 32 studies were included and anthropometric data was extracted from 15 studies. Meta-analysis revealed 14 anthropometric measurements were significantly smaller for females. Mean differences ranged from 0.37 mm to 22.05 mm. Gender-based anthropometric differences did not always translate to lower fit factor scores, with 12 studies reporting worse performance and fit for females and 10 reporting no gender effect. No studies provided disaggregate anthropometric data by ethnic group. Pass rates (PR) were low or moderate among selective BAME people. 18 studies reported RPE fit. Three comparative studies showed lower PR among BAME groups. Pass rates varied with model and design. Most studies did not pre-specify inclusion/exclusion criteria. Small sample size and lack of justification or power calculations was a concern. Significant heterogeneity in study design limited comparisons, particularly relating to respirator selection or availability and defining study outcomes relating to RPE performance.

Conclusion The literature reports on largely Caucasian or single ethnic populations, and BAME people remain under-represented, limiting comparisons between ethnic groups. Facial anthropometrics vary between gender and likely between ethnicity, which may contribute to lower PR among females and ethnic minorities, particularly Asians. There is a need for studies including a broader spectrum of ethnicities and for consideration of female and BAME users during RPE development.

Key questions

What is already known?

► For respirators to provide respiratory protection they must fit the user well, and this is determined by ‘fit-testing’—a process of trialling successive facemasks until one is identified which provides a good seal, and thus the user passes ‘fit-testing’.

► Pass rates (PR) are the proportion of participants that pass ‘fit-testing’ and are successful at identifying a facemask that fits.

► Several factors may affect respirator fit and performance: anthropometric influences are relatively well studied and described but the association of gender and ethnicity is disputed.

► The current COVID-19 pandemic is disproportionate-ly affecting black, Asian and minority ethnicities (BAME) healthcare workers most at risk and appropriately fitting respiratory protective equipment (RPE) is paramount.

INTRODUCTION

There is growing evidence that black, Asian and minority ethnic (BAME) people are disproportionately affected by SARS-CoV-2 (COVID-19).1–5 Indeed, data from the UK-based Office for National Statistics demonstrates COVID-19 related death rates in BAME communities are four times higher compared with those of white ethnicity.6 BAME people comprise only 14% of the population in the UK, yet account for 34% of COVID-19-related admissions to intensive care and 35% of deaths.7–8 Similar trends are seen internationally.9–11 BAME people comprise a large proportion of workers in essential services, including healthcare, and their over-representation among patients affected by COVID-19 is a growing concern. Among National Health Service (NHS) staff, 63% of COVID-related deaths are of BAME people even though they represent only 20%
of the NHS workforce. The effect is likely multifactorial, and addressing these ethnic inequalities requires efforts in various aspects, including effective personal protection equipment (PPE) in the workplace.

Respiratory protective equipment (RPE) is vital in the prevention of nosocomial viral transmission. Systematic reviews and meta-analyses demonstrate the use of masks can reduce the risk of respiratory virus infection by 80%, suggesting mask use offers significant protection against transmission of respiratory viruses such as influenza, SARS and COVID-19. In the context of COVID-19, mask use has been shown to reduce the risk of infection by nearly 70% among healthcare workers, highlighting the importance of RPE in the current pandemic.

European and American safety regulatory bodies such as the Occupational Safety and Health Administration (OSHA) or Health and Safety Executive mandate RPE must meet certification requirements, such as those developed by the National Institute for Occupational Safety and Health (NIOSH), International Organization for Standardization (ISO) or British Standards Institution (BSI). Certification requires respirators to be fit-tested on participants from a respirator fit test panel (RFTP) comprising subjects with facial sizes representative of the user population. Historically, sizing and respirator certification has been based on the Los Alamos National Laboratory (LANL) standardised adult head shape panels, developed in the 1960s using a US Air Force (USAF) Anthropometry Survey of predominantly white male military personnel. The bivariate RFTP referenced for half-mask respirators uses two facial measurements—face length and lip length (figure 1). With evolving population demographics such as changing body shape and increasing female and BAME representation, the USAF data is no longer reflective of the current American workers. Therefore, NIOSH created a novel anthropometric database. This has been used to update the bivariate panel to include face length and face width as well as identify 10 facial dimensions (FD) most relevant to respirator fit, which defines the principal component analysis model. In the UK, BSI standards have been based on the 50th percentile of four dimensions (face length, face width, face depth and mouth width) of the adult white male face shape (figure 1). More recent panels have included a more ethnically diverse sample group.

Fit testing is used to determine if the facial fit of a respirator is free of significant inward leak. Both qualitative fit test (QlFT) and quantitative fit test (QNFT) are recommended. QLFT uses one’s olfactory or taste response to an aerosolised solution. QNFT measures the ratio of external aerosol concentration to internal aerosol concentration, to produce a fit factor (FF) score. Definitions and standards have evolved over time, but currently OSHA recommends a QNFT FF score of 100 affords the user adequate protection and is equivalent to a successful QLFT. Suboptimal fit compromises respiratory protection and can be damaging to underlying skin.

Figure 1  Anthropometric measurements.
The relationship between FD and RPE shape determines RPE fit. FD vary significantly between genders, ethnicities and with age, as well on an individual basis. These may influence RPE fit and there is already some, although mixed, evidence that RPE protection varies with gender-based differences in facial dimension. Certainly, studies of BAME cohorts have yielded particularly low success rates of fit-testing, and similar trends are seen among healthcare workers. These findings may be important in respirator design and manufacturing processes. While newer RFTPs may be more diverse, they are not necessarily representative of healthcare workers (HCWs) or BAME people. There is growing concern that RPE in current use is inadequate at protecting female staff and those from at-risk BAME communities. The objectives of this systematic review were (1) to compare the anthropometric measurements of users across gender and ethnic groups and (2) to assess the effects of FD, gender and ethnicity on RPE fit and effectiveness as measured by fit-test FF scores, fit-test pass rates (PR) or inward leak.

METHODS
The systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The PRISMA checklist is available in online supplemental appendix 1. A protocol for the review was defined, including inclusion and exclusion criteria but a priori protocol registration was not performed.

Search strategy
A literature search was conducted using Embase and Medline via Ovid, PubMed, Scopus and Web of Science in April 2021. The search strategy (online supplemental appendix 2) included key terms relating to respirators, face masks or PPE, respirator fit, FD or facial anthropometrics and race or ethnicity. Gender anthropometrics and differences between sexes were found to be discussed in most studies, therefore gender search terms were not applied as these restricted search results. Reference lists of included papers were also screened. Only human studies, reported in English were included and no time restrictions were applied.

Study selection and eligibility
Two authors independently screened the search results for relevance based on title and abstract, and unrelated studies were excluded. Subsequently, both authors reviewed full texts to identify studies meeting the inclusion criteria: human studies of any age/gender/ethnicity, assessing half or quarter size filtering facepiece respirators meeting N95/PPF3 standards. Studies pertaining to full-facepiece masks were excluded as these likely relate to different FD. Both disposable or reusable RPE was accepted regardless of brand, design, models and sizes. Studies relating to qualitative or quantitative fit-testing were eligible. Outcomes related to fit-test FF scores, fit-test PR or inward leak in the context of anthropometrics, gender and/or ethnicity. No restriction for setting were applied nor to participant characteristics such as occupation, ethnicity, race, gender or age. Studies not assessing the effect of at least one of, anthropometrics, gender or ethnicity, were excluded. Non-English language studies were excluded. Findings were compared and differences were addressed by re-review and discussion until a consensus was reached.

Outcomes
The outcomes of this review were to compare the anthropometric measurements of users across gender and ethnic groups and assess the effect of FD, gender and ethnicity on RPE fit and effectiveness as measured by fit-test FF scores, fit-test PR or inward leakage.

Data extraction
An initial data extraction pro-forma was piloted on a small number of records, modified as required and confirmed. Data extracted related to study characteristics and outcomes, including study design, study population, participant characteristics (age, gender distribution, race distribution), method of FD measurement, anthropometrics data, RPE type, fit-testing protocol, and outcome measures of differences in anthropometrics and in RPE fit. For meta-analysis, we intended to collate data on anthropometric measurements for gender and ethnic groups as well as disaggregated group FF scores and PR.

Analysis
For systematic review, variables including FD, gender and ethnicity were organised into tables and described qualitatively. Association of variables FD, gender and ethnicity with RPE fit were summarised. Limitations and implications for this review are discussed.

Facial measurement means and associated SD were extracted where possible and a meta-analysis was performed for gender-based anthropometrics. Standardised methodologies for anthropometric measurements were employed by included studies and therefore sufficiently similar for meta-analysis. A random-effects meta-analysis was performed using RevMan. Statistical heterogeneity was assessed by the measure of I². For facial measurements where I² indicated substantial heterogeneity (>50%), study methods were reviewed for possible explanations. Studies were assessed for clinical and methodological heterogeneity to identify any outlying studies conflicting with the remaining studies across the 14 anthropometrics. Sensitivity analysis was conducted to determine whether the gender-based differences in anthropometrics are robust. Attempts were made to identify studies contributing to heterogeneity for exclusion. Anthropometrics were suspected to differ between ethnicities, therefore results were reviewed to identify groups of studies with conflicting results based on ethnicity for subgroup analysis.
Disaggregated anthropometric data was not available to allow for ethnicity-based FD comparisons. Due to heterogeneity in study design, outcome measures and reporting, meta-analysis could not be conducted for RPE performance.

**Risk of bias assessment**

The National Heart, Lung and Blood Institute (NHRBI) study quality assessment tools for observational cohort and cross-sectional studies has previously been adapted to assess the quality of studies in the context of anthropometric measurements between gender groups. The NHRBI tool was similarly modified and applied to the studies included in this systematic review based on available guidance from the NHRBI tool.

**Patient and public involvement**

This research does not directly include patient or public involvement. The aims and questions are informed by national and international experiences of female and BAME HCWs in using RPE during the ongoing pandemic.

**RESULTS**

**Literature search results**

Search of the five databases yielded 796 records, with 544 remaining after excluding duplicates (figure 2). Of these, 401 studies were excluded based on title alone and 100 studies based on abstract. These were either unrelated to RPE or pertained to mask-design, methods of fit-testing and other predictors such as facial hair.

---

**Figure 2** Preferred Reporting Items for Systematic reviews and Meta-Analyses flow diagram detailing study selection. RFTP, respirator fit test panel; RPE, respiratory protective equipment.
and temporal changes. Review articles and conference papers were also excluded. Full texts were reviewed for the remaining 43 records and a further 12 articles were excluded. Further detail of reasons for exclusion are shown in online supplemental appendix 3. One additional study was included from screening of references. Therefore 32 articles were identified as eligible for inclusion. Publication year ranged from 1982 to 2021, and all publications were in English. Most studies were published in non-medical journals, largely relating to occupational, industrial or environmental hygiene, ergonomics or health and safety fields. Finally, 15 studies reported anthropometric measurements for meta-analysis.

### Study characteristics

Study characteristics are presented in table 1. The 32 included studies yielded a total of 10658 participants, of which 33% were male and 60% female, with 8% being unreported. Four studies included a Caucasian population, five studied a Korean population, two studied a Chinese population, three studied an Iranian population, one studied a Taiwanese population and one studied a Latino migrant workers population. Eight studies had populations of mixed ethnicity, with the predominant group being Caucasian or black/African. Ethnicity was not reported for eight studies, which were based in Australia, France, Spain, UK and the USA. The distribution of participant ethnicities is shown in online supplemental appendix 4. Participants included HCWs, university students and staff or civilian workers from surrounding communities.

Between 1 and 21 FD were measured by 26 studies in varying combinations and one study recorded overall face shape and size. The most frequently reported anthropometrics are shown in the table which references standardised measurements from the US air force anthropometric report. Fit-testing protocols were in accordance with regulations relevant at the time of study, including ANSI and OSHA standards and in most studies involved quantitative measurement of FF using a PortaCount Plus. Six studies performed qualitative fit-testing and two assessed inward leak. The variety of RPE brands, models and sizes used and fit-testing methods are reported in table 1.

### Systematic review and meta-analysis findings

Study results were compared qualitatively. Comparisons of anthropometrics between gender and ethnicity groups are shown in table 2. Anthropometric data was available for meta-analysis from 15 studies. Mean differences and 95% CIs for 14 standardised anthropometric measurements are shown in table 3, with complete data and forest plots available in online supplemental appendix 5. A random-effects meta-analysis revealed all 14 anthropometric measurements were significantly smaller for females (p<0.05). Differences ranged from 0.37 mm for the smallest measurement (nasal root breadth) to 22.05 mm for the longest measurement (bitragion-menton arc). Heterogeneity was substantial (I²>50%) for nine FD. Gender effect was in the opposite direction in one study, with greater face length and face width for females. Sensitivity analysis with exclusion of this study increased the mean difference between genders minimally and improved I² by 10% for face length and 6% for face width. No specific study was identified to contribute substantially to heterogeneity across all 14 measurements. Therefore, no further studies were excluded for sensitivity analysis. Separation of studies by ethnicity did not improve I² substantially but significantly reduced participant population, therefore subgroup analysis was not performed. Data for anthropometrics of ethnic groups were not available to meta-analyse. Effects of anthropometrics, gender and ethnicity on RPE fit are summarised in figure 2, with complete data per study available in online supplemental appendix 6. Disaggregated data for FF scores and/or PR were not available and heterogeneity in study design and reporting hampered direct comparison of RPE fit outcomes between studies.

### Qualitative synthesis

**FD differ with gender**

Gender-based anthropometrics were compared by 15 studies (table 2). Overall, 13 studies demonstrated gender differences, with smaller average female FD for most measurements. Female measurements were reported to range between 91.5% and 92.5% of the comparable male measurements although with significant overlap of 20%–50%. Some studies reported no gender differences for nasal root breadth, nose length, nose protrusion and lip width, lower face length and one reported greater smiling lip length for females. Meta-analysis demonstrated that all anthropometric measurements were significantly smaller for females than males. Differences in nasal root breadth were minimal but still statistically significant.

**FD differ with ethnicity**

Ethnicity data was collected by six studies, of which two studies reported anthropometric data and between-group differences. An American study with participants from three ethnic groups found significant differences in all facial measurements, except face length. Interestingly, facial measurements were comparable to early studies comprising a largely Caucasian male population. A South African study including four ethnic groups also reported variation between ethnicities. Asian females had significantly smaller facial measurements and black males had greater nasal root breadth measurements as compared with their white counterparts. An Australian survey collected information on overall facial shape and size/shape rather than anthropometric measurements and reported facial characteristics were strongly associated with racial group. The three remaining studies...
<table>
<thead>
<tr>
<th>Study, country</th>
<th>No. of participants (% male)</th>
<th>Age range (years), (mean*)</th>
<th>Population</th>
<th>Population ethnicity† (%)</th>
<th>RPE type Number of brands/models/sizes (RPE tested per user)***</th>
<th>No. of facial dimensions</th>
<th>RPE fit measure (guidelines/standards) and outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liau et al, USA</td>
<td>190 (100%)</td>
<td>N/A</td>
<td>Laboratory employees</td>
<td>Caucasian</td>
<td>Reusable half mask 4 brands, 10 sizes total</td>
<td>7</td>
<td>QNFT Protection factor</td>
</tr>
<tr>
<td>Gross and Horstman, USA</td>
<td>121 (50%)</td>
<td>(37.5)</td>
<td>Community members</td>
<td>N/A</td>
<td>Reusable half masks 3 brands (2), 3 sizes each</td>
<td>10‡</td>
<td>QNFT§ (ANSI) Fit factor, pass rates</td>
</tr>
<tr>
<td>Oestenstad et al, USA</td>
<td>73 (53%)</td>
<td>21–50 (30.6)</td>
<td>University student, staff and faculty</td>
<td>White (68%) Black (12%) Asian (12%) Other† (7%)‡</td>
<td>Reusable half mask 1 brand, 3 sizes</td>
<td>12</td>
<td>QNFT (ANSI) Leak shape, size and distribution†† Fit factor</td>
</tr>
<tr>
<td>Oestenstad and Perkins, USA</td>
<td>68 (56%)</td>
<td>21–50 (30.4)</td>
<td>University students and staff</td>
<td>White (69%) Black (13%) Asian (10%) Hispanic (4%) Asian Indian (3%)‡</td>
<td>Reusable half mask 1 brand, 3 sizes</td>
<td>12‡</td>
<td>QNFT++++ (ANSI) Fit factor</td>
</tr>
<tr>
<td>Brazile et al, USA</td>
<td>186 (49%)</td>
<td>N/A</td>
<td>Community members, university students</td>
<td>White (35%) African (31%) Mexican (33%)</td>
<td>Reusable half mask 1 brand, 3 sizes</td>
<td>14‡</td>
<td>QNFT§ (ANSI) Fit factor, pass rates</td>
</tr>
<tr>
<td>Han, South Korea</td>
<td>778 (52%)</td>
<td>20–55</td>
<td>Industrial workers, university students</td>
<td>Korean</td>
<td>Reusable quarter mask 3 brands (3)</td>
<td>2</td>
<td>QNFT+++ (ANSI) Fit factor, pass rates</td>
</tr>
<tr>
<td>Han and Choi, South Korea</td>
<td>150 (75%)</td>
<td>20–55</td>
<td>Community members, university students</td>
<td>Korean</td>
<td>Reusable half mask 3 brands (3), 1 size (M)</td>
<td>10‡</td>
<td>QNFT+++ (ANSI/OSHA) Fit factor, pass rates</td>
</tr>
<tr>
<td>Kim et al, South Korea</td>
<td>110 (64%)</td>
<td>N/A</td>
<td>University students</td>
<td>Korean</td>
<td>Reusable quarter mask 3 brands (3), 1 size (M)</td>
<td>12‡</td>
<td>QNFT+++ (ANSI/OSHA) Fit factor, pass rates</td>
</tr>
<tr>
<td>Zhuang et al, USA</td>
<td>32 (45%)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Disposable N95 half mask 18 models (18), 1–3 sizes</td>
<td>12‡</td>
<td>QNFT SWPF</td>
</tr>
<tr>
<td>Oestenstad et al, USA</td>
<td>41 (51%)</td>
<td>20–55 (30)</td>
<td>Institute student and staff</td>
<td>Caucasian</td>
<td>Reusable half mask 3 brands (3), &gt;1 size</td>
<td>12‡</td>
<td>QNFT Fit factor</td>
</tr>
<tr>
<td>McMahon et al, Canada</td>
<td>1295 (24%)</td>
<td>19–71</td>
<td>Healthcare workers</td>
<td>N/A</td>
<td>Disposable N95 half masks 3 brands §§, 6 models</td>
<td>N/A</td>
<td>QLFT Pass rates</td>
</tr>
<tr>
<td>Zhuang et al, USA</td>
<td>30 (43%)</td>
<td>N/A</td>
<td>Community members</td>
<td>While (90%) Black (33%) Asian (66%)</td>
<td>Disposable and reusable half masks 4 models (4), 3 sizes</td>
<td>3</td>
<td>QNFT Fit factor, pass rates</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Study, country</th>
<th>No. of participants (% male)</th>
<th>Age range (years), (mean*)</th>
<th>Population</th>
<th>Population ethnicity† (%)</th>
<th>RPE type Number of brands/models/sizes (RPE tested per user)***</th>
<th>No. of facial dimensions</th>
<th>RPE fit measure (guidelines/standards) and outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter et al, Australia⁵⁸</td>
<td>50 (N/A)</td>
<td>N/A</td>
<td>Healthcare workers</td>
<td>N/A</td>
<td>Disposable N95 half masks 2 brands, 3 models (3)</td>
<td>1</td>
<td>QLFT</td>
</tr>
<tr>
<td>Wilkinson et al, Australia⁵⁹</td>
<td>5024¶¶ (21%)</td>
<td>Mode age group: 41–50</td>
<td>Healthcare workers</td>
<td>Aboriginals (0.9%) White (88.9%) East-Asian (5.7%) South/ Central-Asian (3.5%) Other*** (1.0%)</td>
<td>Disposable P2/N95 half masks 3 brands§§, 1–2 sizes</td>
<td>N/A (overall face shape/size data collected)</td>
<td>QNFT‡‡ Pass rates</td>
</tr>
<tr>
<td>Oestenstad and Bartolucci, USA⁶⁰</td>
<td>41 (51%)</td>
<td>20–55 (3)</td>
<td>University students and staff</td>
<td>Caucasian</td>
<td>Reusable half masks 3 brands (3)</td>
<td>12‡</td>
<td>Leak size, shape and distribution ††</td>
</tr>
<tr>
<td>Spies et al, South Africa⁶¹</td>
<td>29 (48%)</td>
<td>N/A</td>
<td>Research institute employees</td>
<td>African (45%) European (41%) Coloured††† (7%) Asian (7%)</td>
<td>Disposable P2 half mask 1 model, 1 size (M)</td>
<td>4††† ‡</td>
<td>QNFT‡‡ (OSHA) Fit factor, pass rates</td>
</tr>
<tr>
<td>Ciotti et al, France⁶²</td>
<td>50 (N/A)</td>
<td>N/A</td>
<td>Healthcare workers</td>
<td>N/A</td>
<td>Disposable PPF2 half masks 9 models (2-3)</td>
<td>N/A</td>
<td>QNFT‡‡ Fit factor, pass rates</td>
</tr>
<tr>
<td>Earle-Richardson et al, USA⁶²</td>
<td>56 (88%)</td>
<td>15–81 (33.2)</td>
<td>Farmworkers</td>
<td>Latino</td>
<td>Disposable N95 and reusable half masks 4 brands, 7 models§§</td>
<td>N/A</td>
<td>QLFT (OSHA) Pass rates</td>
</tr>
<tr>
<td>Yu et al, China⁶³</td>
<td>50 (52%)</td>
<td>Mean age (SD) 21.5 (2.2)</td>
<td>N/A</td>
<td>Chinese</td>
<td>Disposable N95 half masks 4 brands, 10 models (10), 2 sizes</td>
<td>21‡</td>
<td>QNFT‡‡ (OSHA) Fit factor, pass rates</td>
</tr>
<tr>
<td>Bergman et al, USA⁶⁴</td>
<td>229 (N/A)</td>
<td>N/A</td>
<td>General population</td>
<td>N/A</td>
<td>Disposable N95 half masks 7 models§§, 1–2 sizes</td>
<td>13</td>
<td>QNFT‡‡ (OSHA) Fit factor, inward leak</td>
</tr>
<tr>
<td>Kim et al, South Korea⁶⁵</td>
<td>49 (67%)</td>
<td>Mean age (SD) 23.0 (3.8)</td>
<td>Healthcare workers</td>
<td>Korean</td>
<td>Disposable N95 half masks 1 brand, 2 models, 3 sizes</td>
<td>7‡</td>
<td>QNFT‡‡ (OSHA) SWPF, fit factor, pass rates</td>
</tr>
</tbody>
</table>

Table 1 Continued
## Table 1

<table>
<thead>
<tr>
<th>Study, country</th>
<th>No. of participants (% male)</th>
<th>Age range (years), (mean*)</th>
<th>Population</th>
<th>RPE type Number of brands/models/sizes (RPE tested per user)***</th>
<th>No. of facial dimensions</th>
<th>RPE fit measure (guidelines/standards) and outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin and Chen, Taiwan⁶⁵</td>
<td>206 (49%)</td>
<td>21–30</td>
<td>Community members, university students</td>
<td>Disposable N95 half masks 3 models (3), 1 size</td>
<td>19‡</td>
<td>QLFT (OSHA)</td>
</tr>
<tr>
<td>Manganyi et al, South Africa⁷⁸</td>
<td>562 (33%)</td>
<td>Mode age group: 19–30 years</td>
<td>Laboratory employees</td>
<td>Disposable N95/FFP2 half masks &gt;2 brands (1) or 2 sizes (S, M)</td>
<td>4</td>
<td>QNFT‡‡ (OSHA) Fit factor, pass rates</td>
</tr>
<tr>
<td>Honarbakhsh et al, Iran⁶⁷</td>
<td>95 (33.5%)</td>
<td>N/A</td>
<td>Healthcare workers</td>
<td>Disposable N95 half masks 3 models, 1 size</td>
<td>2‡</td>
<td>QLFT (OSHA) Fit factor, pass rates</td>
</tr>
<tr>
<td>Huh et al, South Korea⁶⁸</td>
<td>211 (51%)</td>
<td>Median 26 IQR 23–31</td>
<td>Military hospital volunteers</td>
<td>Disposable N95 half masks 3 brands, 4 models (4), 1–3 sizes</td>
<td>2</td>
<td>QNFT‡‡ (OSHA) Fit factor, pass rates</td>
</tr>
<tr>
<td>Foereland et al, Norway⁶⁹</td>
<td>127 (88%)</td>
<td>18–65 (37)</td>
<td>Smelting industry workers</td>
<td>Disposable P3 half masks 4 brands, 14 models (≥5), 1 or 3 sizes</td>
<td>N/A</td>
<td>QNFT‡‡ (OSHA) Fit factor, pass rates</td>
</tr>
<tr>
<td>Winski et al, UK⁷⁰</td>
<td>262 (90.5%)</td>
<td>N/A</td>
<td>General population</td>
<td>Disposable PPF3 half mask 1 model</td>
<td>3</td>
<td>QNFT‡‡ (BSIF) Fit factor, pass rates</td>
</tr>
<tr>
<td>Fakherpour et al, Iran⁷¹</td>
<td>62 (40%)</td>
<td>Mean age (SD) 23.45 (4.66)</td>
<td>University students</td>
<td>Disposable N95/PPF2/FFP3 half masks 4 brands (4)</td>
<td>2‡</td>
<td>QLFT Fit factor, pass rates</td>
</tr>
<tr>
<td>Zhang et al, China⁷²</td>
<td>85 (36%)</td>
<td>Mean age (SD) 27 (4.4)</td>
<td>University students</td>
<td>Disposable N95/FFP3 half masks 4 models (4), 1 size</td>
<td>8</td>
<td>QNFT‡‡ Fit factor, pass rates</td>
</tr>
<tr>
<td>De-Yñigo-Mojado et al, Spain⁷³</td>
<td>74 (50%)</td>
<td>Mean age (SD) 34.31 (7.13)</td>
<td>Healthcare workers</td>
<td>Disposable FFP3 half masks 2 brands, 3 models</td>
<td>4‡</td>
<td>QNFT‡‡ Fit factor, pass rates</td>
</tr>
<tr>
<td>Fakherpour et al, Iran⁷⁴</td>
<td>37 (32%)</td>
<td>Mean age (SD) 24.6 (4.2)</td>
<td>University volunteers</td>
<td>Disposable N95/FFP2 half masks 15 brands, 20 models (20)</td>
<td>2</td>
<td>QNFT‡‡ Fit factor, pass rates</td>
</tr>
</tbody>
</table>

Continued
were unable to compare anthropometrics between ethnic groups due to small sample sizes.

Studies also drew comparisons between their cohorts and those of previous studies (table 2). Studies of various Asian populations reported significantly different FD compared with Caucasian cohorts, with generally smaller and wider faces. Korean participants had wider face width and nose breadth, narrower nasal root breadth and lip width. Chinese and Iranian participants had wider face width and shorter face length and Taiwanese participants had overall smaller faces. FD of males from an ethnically mixed South African cohort were also smaller and wider than for Caucasians. Several studies showed skewed distribution of participant FD compared with the American panel FD such that significant proportions of their cohorts lie outside RFTPs.

Gender effects on RPE fit

Gender-based differences in anthropometrics have not consistently translated to a difference in FF (table 4). Of 24 studies comparing PR and/or FF scores between genders, 13 studies demonstrated significant gender effects. Of these, 11 studies reported higher fit-test failure rates and/or lower FF scores among females. Factors such as facial stubble which hamper RPE performance may reduce fit for males such that PR appear similar between genders, but comparison of only clean-shaven males yielded higher PR than for females. Gender was also reported to account for a higher proportion of variability in FF scores in analysis of variance. Association of FD and leak sites was mostly attributed to gender. Two studies did not compare PR but did demonstrate an association of gender-based FD with leak distribution and greater predictability of FF using gender specific models.

In comparison, 11 studies reported no gender effects, with similar PR, no effect on FF score or no effect of gender on leak distribution/shape/sizes. One study reported mixed results with higher PR among males for two of three RPE models but comparable PR overall across all RPE models. A further study reported higher PR among female users. The variable effects of gender on RPE fit may be the result of differences in methodology. Study design was variable, with some studies assessing one model in multiple sizes, multiple models...
### Table 2  Comparison of anthropometrics between gender and ethnicity groups

<table>
<thead>
<tr>
<th>Studies, country</th>
<th>Between genders</th>
<th>Between ethnicities</th>
<th>To other populations/panels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross and Horstman, USA&lt;sup&gt;19&lt;/sup&gt;</td>
<td>♂ smaller dimensions for 11/12 FD</td>
<td>N/A</td>
<td>Comparable to US Air Force population</td>
</tr>
<tr>
<td>Oestenstad et al, USA&lt;sup&gt;50&lt;/sup&gt;</td>
<td>N/A</td>
<td>No comparison made due to small sample size</td>
<td>Skewed distribution relative to LANL panel</td>
</tr>
<tr>
<td>Oestenstad and Perkins, USA&lt;sup&gt;27&lt;/sup&gt;</td>
<td>♂ smaller dimensions</td>
<td>No comparison made due to small sample size</td>
<td>Comparable to previous studies and US Air Force population</td>
</tr>
<tr>
<td>Brazile et al, USA&lt;sup&gt;51&lt;/sup&gt;</td>
<td>♂ smaller dimensions for 12/14 FD except binocular and NRB</td>
<td>Significant difference between ethnic groups except for FL</td>
<td>Comparable to previous studies and US Air Force population</td>
</tr>
<tr>
<td>Han, South Korea&lt;sup&gt;52&lt;/sup&gt;</td>
<td>♂ smaller dimensions separately screwed distribution of FD but with significant overlap</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Han and Choi, South Korea&lt;sup&gt;29&lt;/sup&gt;</td>
<td>♂ smaller dimensions for all 10 FD</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Kim et al, South Korea&lt;sup&gt;53&lt;/sup&gt;</td>
<td>♂ smaller dimensions for 11/12 FD except for NRB</td>
<td>N/A</td>
<td>Comparable to Korean cohorts. Different (smaller and wider faces) to American cohorts</td>
</tr>
<tr>
<td>Zhuang et al, USA&lt;sup&gt;54&lt;/sup&gt;</td>
<td>♂ smaller dimension for 9/12 measurements except LW, NRB, NP</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Oestenstad et al, USA&lt;sup&gt;65&lt;/sup&gt;</td>
<td>♂ smaller dimensions for 10/12 FD except LFL and NL</td>
<td>N/A</td>
<td>Comparable to previous studies</td>
</tr>
<tr>
<td>Wilkinson et al, Australia&lt;sup&gt;59&lt;/sup&gt;</td>
<td>N/A</td>
<td>Facial characteristics were strongly associated with racial group</td>
<td>N/A</td>
</tr>
<tr>
<td>Spies et al, South Africa&lt;sup&gt;61&lt;/sup&gt;</td>
<td>♂ smaller and narrower dimensions</td>
<td>Comparison not made</td>
<td>Screwed distribution relative to LANL panel. Mean FD comparable to Korean and American cohorts but male FD different (smaller and wider) from American cohort</td>
</tr>
<tr>
<td>Yu et al, China&lt;sup&gt;63&lt;/sup&gt;</td>
<td>♂ smaller dimensions</td>
<td>N/A</td>
<td>Comparable to Chinese cohort. Different (smaller and wider) to American cohorts</td>
</tr>
<tr>
<td>Kim et al, South Korea&lt;sup&gt;65&lt;/sup&gt;</td>
<td>♂ smaller LW only</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Manganyi et al, South Africa&lt;sup&gt;28&lt;/sup&gt;</td>
<td>♂ smaller dimensions</td>
<td>Asian♂: smaller dimensions African ♀: greater NRB</td>
<td>N/A</td>
</tr>
<tr>
<td>Lin and Chen, Taiwan&lt;sup&gt;66&lt;/sup&gt;</td>
<td>♂ smaller dimensions separately screwed distribution of FD</td>
<td>N/A</td>
<td>Screwed distribution relative to NIOSH panel. Difference to American cohorts (smaller)</td>
</tr>
<tr>
<td>Honarbashkh et al, Iran&lt;sup&gt;67&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>Significant proportion outside RFTP. Different to South African, Korean and American cohorts (smaller FL and FW)</td>
</tr>
</tbody>
</table>

Continued
Studies, country

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Studies</th>
<th>Male participants (n)</th>
<th>Female participants (n)</th>
<th>Mean difference (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biectoorbitale breadth</td>
<td>4</td>
<td>260</td>
<td>150</td>
<td>9.26 (7.54 to 10.97)</td>
</tr>
<tr>
<td>Bizygomatic breadth (face width)</td>
<td>15</td>
<td>1503</td>
<td>742</td>
<td>7.54 (6.80 to 8.27)</td>
</tr>
<tr>
<td>Bigonial breadth (jaw width)</td>
<td>8</td>
<td>834</td>
<td>468</td>
<td>6.75 (5.81 to 7.69)</td>
</tr>
<tr>
<td>Menton-nasion length (face length)</td>
<td>15</td>
<td>1503</td>
<td>742</td>
<td>7.82 (7.13 to 8.50)</td>
</tr>
<tr>
<td>Menton-subnasale length (lower face length)</td>
<td>9</td>
<td>727</td>
<td>409</td>
<td>5.26 (4.54 to 5.97)</td>
</tr>
<tr>
<td>Subnasale-nasion length (nose length)</td>
<td>9</td>
<td>973</td>
<td>541</td>
<td>3.64 (3.16 to 4.13)</td>
</tr>
<tr>
<td>Bicocular breadth</td>
<td>4</td>
<td>260</td>
<td>150</td>
<td>3.87 (3.00 to 4.74)</td>
</tr>
<tr>
<td>Nasal root breadth</td>
<td>8</td>
<td>734</td>
<td>382</td>
<td>0.37 (0.12 to 0.61)</td>
</tr>
<tr>
<td>Nose width</td>
<td>12</td>
<td>1083</td>
<td>585</td>
<td>3.42 (3.06 to 3.78)</td>
</tr>
<tr>
<td>Lip length/width</td>
<td>13</td>
<td>1157</td>
<td>622</td>
<td>2.82 (2.36 to 3.28)</td>
</tr>
<tr>
<td>Bitragion-menton arc</td>
<td>9</td>
<td>884</td>
<td>494</td>
<td>22.05 (20.15 to 23.95)</td>
</tr>
<tr>
<td>Bitragion-subnasale arc</td>
<td>10</td>
<td>933</td>
<td>510</td>
<td>18.43 (16.70 to 20.16)</td>
</tr>
<tr>
<td>Nose protrusion</td>
<td>6</td>
<td>745</td>
<td>405</td>
<td>2.03 (1.65 to 2.40)</td>
</tr>
<tr>
<td>Interpupillary distance</td>
<td>3</td>
<td>288</td>
<td>141</td>
<td>2.70 (2.02 to 3.39)</td>
</tr>
</tbody>
</table>
### Table 4  Summary of findings: association of variables facial dimensions, gender and ethnicity with RPE fit

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Studies assessing outcome</th>
<th>Summary of findings (n=number of studies)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pass rates†</strong></td>
<td>26</td>
<td>High (&gt;75%) overall user PR*: n=9. Moderate (50%–74%) overall user PR: n=4. Low (&lt;50%) overall PR: n=4. Variable PR between gender and/or ethnic groups: n=9. Overall low or low-moderate PR in studies of non-white cohorts (n=12).</td>
</tr>
<tr>
<td><strong>Association between facial dimensions (FD) and fit</strong></td>
<td>25 n=14</td>
<td>FD association with fit: FW (n=7), FL (n=6), NRB (n=4), JW (n=4), LFL (n=3), NL (n=3), NP (n=3), NW (n=2), LW (n=2), BIOC (n=2), BECTO (n=2), BTMA (n=2). Facial size and shape categories associated with FF (n=6). Extremes in FW and FL or narrower faces associated with fit (n=2).</td>
</tr>
<tr>
<td><strong>Association between gender and fit‡</strong></td>
<td>24 n=12</td>
<td>PR males&gt;females (n=8). PR females&gt;males (n=2). Gender differences in PR varied with model (n=1). Association of facial dimensions and leak site attributed to gender. Male gender is independent predictor for fit.</td>
</tr>
<tr>
<td><strong>Association between ethnicity and fit†</strong></td>
<td>5 n=2</td>
<td>Race specific models improve fit predictability. Lower PR for Asians, highest for Caucasians.</td>
</tr>
<tr>
<td><strong>Association between mask factors and fit§</strong></td>
<td>20 n=18</td>
<td>Variability in FF based on brand. Significant difference in FF/PR between brands (n=12). Influence of RPE on fit within facial size categories (n=2) or shape (n=1). RPE is determinant or predictor of fit (n=2).</td>
</tr>
</tbody>
</table>

*Overall user pass rates—percentage of participants successfully fit-testing on at least one RPE model.
†PR are reported as either (1) PR of users, as a percentage of participants who passed fit-testing on at least one respirator or (2) PR for RPE groups, as a percentage of participants who passed fit-testing for the respirator being tested.
‡RPE fit as measured by respective studies, including fit/protection factor (FF), simulated workplace protection factor (SWPF), inward leakage (IL), fit-testing PR (PR).
§Mask factors are reported as any differences in FF or PR relating to mask factors such design, model, brand, shape or size.
BECTO, biectoorbitale breadth; BIOC, biocular breadth; BTMA, bitragion-menton arc; FL, face length; FW, face width; JW, jaw width; LFL, lower face length/menton-subnasale length; LW, lip width; NL, nose length/subnasale-nasion length; NRB, nasal root breadth; NW, nose width; PR, pass rates; RPE, respiratory protective equipment.

Their mixed cohorts of predominately BAME participants, using single model/size RPE and multiple brands/sizes RPE, respectively. In particular, the lowest PR were seen in Asian females. The largest study, an Australian survey, similarly reported the highest failure rates were among Asian HCWs and the highest PR were among white HCWs. Of studies assessing BAME cohorts, ten have reported particularly low PR with significant variability between RPE models. Studies of solely Chinese or Korean cohorts report low PR when assessing subgroups for gender and certain mask type. While some masks were associated with PR between 60% and 87%, others were successful for only 10%–30% of users.
and Iranian studies even found some masks were ineffective for all of their participants. Additional studies demonstrate low PR among Hispanic workers, suggesting current RPE may be inadequate, however, the ethnic distribution of these populations was not reported.

Mask factors affect RPE performance
A total of 20 studies compared FF and/or PR between different RPE brands and models; 17 studies demonstrated RPE performance differs significantly based on design. One study reported FF score varied with RPE brand for females only, with no correlation in the male group. A study assessing 18 RPE models however demonstrated the number of models and sizes available is associated with FF, rather than the RPE design itself.

Risk of bias within studies
Quality assessment is presented in table 5. The majority of studies fail to meet criteria three as inclusion and exclusion criteria were not prespecified. The majority of studies also do not provide sample size justifications or power calculations. However, many are still able to meet criteria four as they report on variance or effect estimates, as detailed by the NHLBI assessment tool. Of note, several studies do not meet criteria five as anthropometric data were not collected.

DISCUSSION
Our review demonstrates significant gender-based variance in standardised anthropometric measurements, with significantly smaller female FD for all measurements. Comparing Asian and black/African groups to Caucasians shows differences in facial geometry such as overall face size and nose measurements. With regard to RPE performance, female and BAME participants have generally low FF scores and/or fit-test PR. However, only a limited number of studies included BAME people in RPE fit-testing. Given the limited number of comparative studies available and heterogeneity in study design, we cannot be conclusive in our evaluation of RPE performance in gender or ethnic groups and their associations with specific anthropometric parameters.

BSI recognises anatomical and structural differences between genders. Our review shows that facial measurements included in RFTPs, namely face length, face width and lip width, are smaller for females. This is consistent with a large gender-based anthropometric study. In the context of fit-testing; most studies collected data limited to FD included in the LANL and NIOSH bivariate RFTPs. A limited number of studies collected additional facial measurements, such as nose dimensions, and showed that these features are relevant to RPE fit. Hence, the inclusion of these additional dimensions and their correlation to RPE performance would be valuable in future studies.

ISO has reported differences in facial characteristics between Caucasian, Sub-Saharan and European facial types. Comparisons between Caucasian and black participants demonstrate that the latter have greater protrusion of lips, greater head depth, and shorter, wider, shallower noses. Hispanic workers have significantly larger facial features for 14 measurements than Caucasians, with shorter nose protrusion and head length. Asian participants have statistically different dimensions as compared with Caucasians for 16 anthropometric values. However, only a limited number of studies comparatively evaluate the impact of ethnicity on RPE performance.

Furthermore, disaggregated comparisons are lacking for ethnicities outside predominant American groups (Caucasian, black, Hispanic). Often studies categorise participants as ‘Other’ which includes a diverse group of Central, South and East Asians, even though there are significant anthropometric differences between these groups based on ancestry. Our review also includes studies using American RFTPs as benchmarks, which show significant proportions of Chinese, Korean and Iranian participants’ facial measurements lie outside the distribution of American RFTPs. Additionally, individuals from Asian and black ethnic groups continue to be under-represented in RFTPs. There appears to be an urgent need to use fit-test panels that account for ethnicity-specific differences.

Gender-based anthropometric differences are associated with RPE performance in about half of our studies, the majority of which demonstrate that female participants have significantly lower RPE performance, need a variety of mask models for successful fit and are more likely to fail fit-testing altogether. The heterogeneity in results is likely related to study design, of which RPE availability and the assortment of models on offer are particularly relevant. First, many studies do not make gender-based comparisons of RPE performance for individual mask models, comparing overall fit-testing success between genders instead. This is based on successful fit-testing with at least one respirator, which fails to account for the higher fit-testing failure rates for individual RPE models among females, therefore reducing gender-based differences in RPE performance. Second, provision of one model in limited sizes or RPE designed as ‘one-size-fits-all’ fails to cater to smaller FD. Increasing RPE choice improves user success rates and reduces gender-based fit-testing differences. For example, a study demonstrated that inclusion of two additional models accounts for a 20% improvement in female PR. Certainly, several studies included here recommend a variety of RPE should be made available to ensure successful fit-testing.

In practice, implementing a comprehensive fit-testing programme is a financial and logistical challenge. The variety of RPE in different healthcare environments is variable and procurement dependent. It may not be feasible to test HCWs on all available RPE given the time-consuming nature of fit-testing.
Studies report mixed results for ethnicity-based differences in RPE performance. Small comparative studies have demonstrated lower PR for black and Asian females, but with no effect of ethnicity on FF scores. These studies were likely underpowered to recognise subgroup differences. Studies of Asian populations have consistently yielded higher rates of fit-test failure among Chinese, Koreans, Taiwanese and Iranians, further emphasising the need to consider FD of their population in RPE design. Therefore, RPE currently available does not provide comparable protection across ethnicities, likely disadvantaging those from minority groups. This implies, RPE design may be failing to accommodate for heterogeneity in facial features.

Table 5  Assessment for bias using modified National Heart, Lung and Blood Institute (NHRBI) study quality assessment tools

<table>
<thead>
<tr>
<th>Studies</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Liu et al</td>
<td>◐</td>
</tr>
<tr>
<td>Gross and Horstman</td>
<td>○</td>
</tr>
<tr>
<td>Oestenstad et al</td>
<td>○</td>
</tr>
<tr>
<td>Oestenstad and Perkins</td>
<td>○</td>
</tr>
<tr>
<td>Brazile et al</td>
<td>○</td>
</tr>
<tr>
<td>Han</td>
<td>○</td>
</tr>
<tr>
<td>Han and Choi</td>
<td>○</td>
</tr>
<tr>
<td>Kim et al</td>
<td>○</td>
</tr>
<tr>
<td>Zhuang et al</td>
<td>○</td>
</tr>
<tr>
<td>Oestenstad et al</td>
<td>○</td>
</tr>
<tr>
<td>McMahon et al</td>
<td>○</td>
</tr>
<tr>
<td>Zhuang et al</td>
<td>○</td>
</tr>
<tr>
<td>Winter et al</td>
<td>○</td>
</tr>
<tr>
<td>Wilkinson et al</td>
<td>○</td>
</tr>
<tr>
<td>Oestenstad and Bartoletti</td>
<td>○</td>
</tr>
<tr>
<td>Spies et al</td>
<td>○</td>
</tr>
<tr>
<td>Ciotti et al</td>
<td>○</td>
</tr>
<tr>
<td>Earle-Richardson et al</td>
<td>○</td>
</tr>
<tr>
<td>Yu et al</td>
<td>○</td>
</tr>
<tr>
<td>Bergman et al</td>
<td>○</td>
</tr>
<tr>
<td>Kim et al</td>
<td>○</td>
</tr>
<tr>
<td>Lin and Chen</td>
<td>○</td>
</tr>
<tr>
<td>Mangani et al</td>
<td>○</td>
</tr>
<tr>
<td>Honarbashlou et al</td>
<td>○</td>
</tr>
<tr>
<td>Huh et al</td>
<td>○</td>
</tr>
<tr>
<td>Foereland et al</td>
<td>○</td>
</tr>
<tr>
<td>Winski et al</td>
<td>○</td>
</tr>
<tr>
<td>Falkeporour et al</td>
<td>○</td>
</tr>
<tr>
<td>Zhang et al</td>
<td>○</td>
</tr>
<tr>
<td>De-Yrigo-Mojado et al</td>
<td>○</td>
</tr>
<tr>
<td>Falkeporour et al</td>
<td>○</td>
</tr>
<tr>
<td>Williams et al</td>
<td>○</td>
</tr>
</tbody>
</table>

**Criteria met:** ◐ criteria not met; ○ criteria not met.

---

*Table 5: Assessment for bias using modified National Heart, Lung and Blood Institute (NHRBI) study quality assessment tools.*

---

1. Did the authors address all the criteria? (1) were arms and objectives clearly stated? (2) was the study population clearly specified and defined? (3) was inclusion/exclusion criteria for being in the study prespecified and applied uniformly to all participants? (4) was a sample size justified? (5) were power description, variance or effect estimates provided? (6) were methods of anthropometric measurement clearly described, valid, reliable and implemented consistently across all study participants? (7) were the independent variables clearly defined, valid, reliable and implemented consistently across all study participants? (8) is it clear what was used for analysis or to determine statistical significance estimated? (9) results—were basic data adequately described? (10) were limitations of study discussed?

Where indicated as ‘criteria not met’ for criteria, (5) inclusion/exclusion criteria have not been specified. Where indicated as ‘criteria not met’ for criteria, (4) no sample size justification or power calculation has been reported, nor assessment of variance or effect size. Most studies did not report sample size justification or power calculation, but criteria were deemed to be satisfied if variance or effect estimate provided.

Anthropometric measurements made from photographs of participants using established landmarks for five of seven facial dimensions. Protection factor scores required to pass not reported. Correlation analysis performed for only a white male subset of the study population.

*Correlation analysis for facial dimensions and respiratory protective equipment (RPE) fit reported as having been performed but results were not provided as no significant correlations made.*

The study was underpowered to assess for race.

Facial measurements not entirely in keeping with standard anthropometric landmarks and measurements, as judged by included figure. Physical examination and pulmonary function performed but inclusion/exclusion criteria not stated.

Study population not specified. Some participants did not test all respirator models and were substituted by others with similar face size categories. Once a successful fit test was obtained other models were not tested. The order of masks tested was applied consistently.

Results of effect of gender, age and occupation reported only briefly.

Data on facial categories collected rather than anthropometric measurements. Respirator for testing was selected by the tester based on observed facial characteristics rather than measured facial dimensions and Los Alamos National Laboratory categories. Once a successful fit test was obtained other models were not tested. Healthcare workers who failed fit testing were not tracked and if returned for second fit-testing sessions were treated as independent events.

Two facial measurements were collected only on a small proportion of participants. SD provided but no between group comparisons available. Correlation analysis was not performed between the facial dimensions and fit factor.

Once a successful fit test was obtained other models were not tested. Estimate of variance and/or effect size were irrelevant for arms of study to determine if RPE fit of respirator size relates to respirator fit test panel facial size categories.

Participants that were not clean shaven were initially included in the analysis which likely skewed results given known effect of facial hair on RPE performance.

Factors such as facial hair presence was not recorded, and could influence the difference in fit factors between genders.

Anthropometric data not collected. Ambiguous categorisation on ethnicity of participants as South-East Asian and non-Asian.
Strengths and limitations
This is the first systematic review and meta-analysis of the influence of gender and ethnicity on RPE, to the best of our knowledge. Our search strategy and eligibility criteria were broad and have captured a large number of relevant studies. However, we were limited to English-based databases. We excluded studies in Chinese as we were unable to gain access to the data. This is a significant limitation considering the focus of our review and inclusion of non-English studies may affect results significantly.

Inherent associations exist between gender and FD as well as multicollinearity between FD, although these associations were not always clearly accounted for or reported by studies. Meta-analysis showed significant heterogeneity existed for nine FD. Of these measurements, those with small magnitude of effect (ie, smaller differences in measurements) such as nasal root breadth (MD 0.37 mm), nose length (MD 3.64 mm), nose protrusion (MD 2.03 mm) and lip width (MD 2.82 mm) may be less relevant or irrelevant to gender-based differences in anthropometrics. By extension, they may be less relevant to RPE fit.

There was significant disparity in study design and methodology in gender-based studies. Assessment of study design confirmed anthropometrics were collected by standardised methods. Only one study reported conflicting results, with FD greater for females. Exclusion of this study did not sufficiently improved heterogeneity. BAME people have different FD to Caucasians, and it was suspected that heterogeneity may be result of participant diversity. However, subgroup analysis based on ethnicity was not possible as studies measured varying combinations of FD and ethnicity-based grouping reduced sample size such that meta-analysis would not provide meaningful conclusions. Risk of bias assessment demonstrated most studies failed to meet criteria three, relating to use of prespecified inclusion and exclusion criteria. This may contribute to heterogeneity observed in meta-analysis of anthropometrics and differences in conclusions regarding gender-based differences in RPE performance. Several studies failed to account for their sample size through justification, power calculation or estimate of variance/effect. These risks studies being underpowered to detect differences in RPE performance between gender and/or ethnic groups, and may account for the conflicting results. Limited number of studies included ethnically diverse participants with all relevant anthropometrics. Hence, we cannot be conclusive in our evaluation of RPE performance on gender or ethnic groups and their associations with specific anthropometric parameters.

Future research
Successfully fit-testing HCWs is particularly important in the current climate. Future studies addressing the disparity in RPE fit will require a review of how respirators are designed and tested, including use of a relevant fit-test panel. Studies should aim to include a diverse group of participants inclusive of BAME people to better inform future mask design and fit testing performance. Studies should include the provision of a variety of mask models, brands and sizes, denoting modifications made during the donning process, and the fit-test PR for all mask models tested rather than using an overall success rate. Longitudinal studies on how facial anthropometrics influence fit, but also user comfort and adverse outcomes thereafter would be useful to inform mask designs. The future clearly lies in personalising fit-testing with modern technology. For example, three-dimensional facial model-capture may be used to assess fit in order to reduce time and costs of fit-testing as well as expedite identification of HCWs who need alternative RPE.

CONCLUSION
Anthropometric data is key in the design and testing of respirators, and user demographics reflected in respiratory fit test panels may influence the level of protection respirators provide. Facial measurements vary significantly between gender and ethnicity. Our meta-analysis demonstrates women have significantly smaller facial measurements for 14 standardised measurements compared with men. The literature suggests significant differences in
anthropometrics between ethnicities, however, minority
groups continue to be under-represented in compar-
itive studies and race-based differences could not be
established in our study. The effect of differences in
facial anthropometrics on respirator fit and effective-
ness is less clear. Over half of studies reporting gender-
based comparisons in RPE performance report signif-
icantly lower PR among females. Three studies report
lower PR among Asian or black participants. However,
these PR differences are inconsistently associated with
absolute FF scores. FD across ethnic minorities may fall
outside the parameters of current RFTP requirements and impact RPE
performance. Therefore, RFTPIs need to be expanded to
capture the distribution of anthropometric data from all
ethnicities and RPE development needs to reflect a more
diverse group of users.

Twitter Peter Worsley @PeteWors and Ying Cheong @Ycheong1

Contributors
The review was designed by JC, CM, PW and YC. The search was
conducted by JC under supervision from YC. Records were screened by JC and NA
independently. Data were extracted and meta-analysis conducted by JC. JC led
the writing of the manuscript for publication with significant contributions from all
authors. All authors approved the final article and are guarantors of the study. The
Corresponding author attests that all listed authors meet authorship criteria and
that no others meeting the criteria have been omitted.

Funding
This work was supported by the Wessex Academic Health Sciences
Network and Southampton Faculty of Medicine Enterprise board. NA was supported
by the European Union’s Horizon 2020 research and innovation programme under
the Marie Skłodowska-Curie grant agreement No. 811 965 (Project STINTS - Skin
Tissue Integrity under Shear). PW had funding support from the Engineering
and Physical Research Council grant ‘A BioEngineering approach for the SAFE
design and fitting of Respiratory Protective Equipment (BE-SAFE RPE)’ (EP/V045563/1). JC
is an NIH Academic Fellow.

Competing interests
None declared.

Patient consent for publication
Not applicable.

Provenance and peer review
Not commissioned; externally peer reviewed.

Data availability statement
All data relevant to the study are included in the
article or uploaded as supplementary information.

Supplemental material
This content has been supplied by the author(s). It has not
been vetted by BMJ Publishing Group Limited (BMJ) and may not have
been peer-reviewed. Any opinions or recommendations discussed are solely those of
the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and
responsibility arising from any reliance placed on the content. Where the content
includes any translated material, BMJ does not warrant the accuracy and reliability
of the translations (including but not limited to local regulations, clinical guidelines,
terminology, drug names and drug dosages), and is not responsible for any error
and/or omissions arising from translation and adaptation or otherwise.

Open access
This is an open access article distributed in accordance with the
Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits
authors to copy, redistribute, remix, transform and build upon this work for any
purpose, provided the original work is properly cited, a link to the licence is given,
and indication of whether changes were made. See: https://creativecommons.org/
licences/by/4.0/.

REFERENCES
1 Aldridge RW, Lewer D, Katikireddi SV, et al. Black, Asian and
minority ethnic groups in England are at increased risk of death from
COVID-19: indirect standardisation of NHS mortality data. Wellcome
2 Platt L, Warwick R. Are some ethnic groups more vulnerable
2020;584:430–6.
5 PHE. Beyond the data: understanding the impact of COVID-19 on
6 Office for National Statistics. Coronavirus (COVID-19) related deaths
7 ONS. 2011 census analysis: ethnicity and religion of the non-UK
born population in England and Wales. Office for National Statistics:
ONS, 2011.
8 ICNARC. ICNARC report on COVID-19 in critical care 07 September
9 Johns A, Buford T. Early data shows African Americans have
contracted, died of coronavirus at an Alarming rate. Medscape:
10 Ramos E, Zamudio M, Chicago I. 70 per cent of COVID-19 deaths are
11 Eligon J, Burch A, Searcey D. Black Americans face Alarming rates
12 THF. Black and minority ethnic workers make up a disproportionately
large share of key worker sectors in London. The Health Foundation,
2020.
13 Cook T, Kursumovic E, Lennane S. Exclusive: deaths of NHS staff
article
14 Workforce and Facilities ND. NHS workforce statistics. Health and
Social Care Information Centre. 2019.
respiratory virus transmission: a systematic review and meta-
analysis. Travel Med Infect Dis 2020;36:101751.
16 Li Y, Liang M, Gao L, et al. Face masks to prevent transmission
17 OSHA. 1910.134. Respiratory protection. Subpart I: Personal
protective equipment, Occupational Safety and Health Standards.
U.S. Department of Labor, Occupational Safety and Health
Administration, 2016.
18 ISO. 16900–1:2019 - Respiratory protective devices — Methods
of test and test equipment — Part 1: Determination of inward leakage
19 NIOSH. RCT-APR-STP-0068: total inward leakage test for Half-
mask Air-purifying particulate respirators. 42 cfr part 84. Subpart K:
20 Hack AL, McConville JT. Respirator protection factors: Part I -
development of an anthropometric test panel. Am Ind Hyg Assoc J
21 Zhuang Z, Guan J, Hsiao H. Evaluating the representativeness of
the LANL respirator fit test panels for the current US. civilian
workers. . The International Society for Respiratory Protection,
2004: 21; 83–93.
22 Zhuang Z, Bradtmiller B, Shaffer RE. New respirator fit test panels
representing the current U.S. civilian workforce. J Occup Environ
23 BSI. Respiratory protective devices - Filtering half masks to protect
against particles - Requirements, testing, marking EN 149:2001+
24 OSHA. 1910.134 App A. Fit testing procedures (mandatory), Subpart
I: personal protective equipment, Occupational Safety and Health
Standards. U.S. Department of Labor, Occupational Safety and
Health Administration, 2016.
25 Worsley PR, Spratt F, Bader DL. COVID19: challenging tissue
viability in both patients and clinicians. J Tissue Viability
differences among gender, ethnicity, and age groups. Ann Occup
27 Oesterndorf RK, Perkins LL. An assessment of critical
anthropometric dimensions for predicting the fit of a half-mask
28 Manganyi J, Wilson KS, Rees D. Quantitative respirator
fit, face sizes, and determinants of fit in South African
diagnostic laboratory respirator users. Ann Work Expo Health
29 Han D-H, Choi K-L. Facial dimensions and predictors of fit for
masks for healthcare workers, in France. Med Mal Infect
31 Feinmann J. When the masks don’t fit: the British collaboration
making PPE designed for individuals. BMJ 2020;m4585.


97 Dean E. COVID-19: how to avoid skin damage while wearing PPE. *Nursing standard*, 2020.