Measurement xxx (xxxx) xxx

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Measuring breathability and bacterial filtration efficiency of face masks in the pandemic context: A round robin study with proficiency testing among non-accredited laboratories

Francesco Tessarolo [a,*], Giandomenico Nollo [a], Luca Benedetti [a], Fabrizia Helfer [b], Luigi Rovati [c,d], Alberto Ferrari [c,d], Giulio Marchetti [e], Federico Reverberi [e], Salvo Baglio [f], Nunzio Tuccitto [f], Stefania Stefani [f], Stefano Stracquadanio [f], Filippo Caraci [f], Antonio Terrasi [f], Alessia Tricomi [f], Mario Musumeci [f,g], Andrea Miraglia [f,g], Giacomo Cuttone [f,g], Sofia Cosentino [h], Luca Agostino Vitali [i], Dezemona Petrelli [i], Sonia Ilaria Ciancia [i], Marina Mingoia [j], Paolo Castellini [j], Serena Simoni [j], Luigi Montalto [j], Alessia Baleani [j], Paolo Chiariotti [k], Nicola Paone [j]

[a] Dipartimento di Ingegneria Industriale, Università degli studi di Trento, Italy
[b] Laboratorio di Sanità Pubblica, Azienda Provinciale per i Servizi Sanitari di Trento, Trento, Italy
[c] Dipartimento di Ingegneria Enzo Ferrari, Università di Modena e Reggio Emilia, Modena, Italy
[d] Science & Technology Park for Medicine, TPM, Mirandola, Modena, Italy
[e] SAFE srl, Laboratorio Chimico-Biologico, Mirandola, Modena, Italy
[f] AntiCovidLab - Torre Biologica - Università degli Studi di Catania, Via S. Sofia, 89, 95123 Catania, Italy
[g] INFN-Laboratori Nazionali del Sud, Catania, Italy
[h] Laboratorio di Igiene applicata, Dipartimento di Scienze Mediche e Sanità Pubblica, Università degli Studi di Cagliari, Cagliari, Italy
[i] UNICAM- U-TYM Lab Microbiologia - ex Dip. Biologia, Università di Camerino, Camerino, Italy
[j] LABC19 Centro di Ricerca e Servizio per l'Emergenza COVID-19, Università Politecnica delle Marche, Ancona, Italy
[k] Dipartimento di Meccanica, Politecnico di Milano, Milano, Italy

ARTICLE INFO

Keywords:
Face mask
Uncertainty
Differential pressure
Bacterial filtration efficiency
Round robin
COVID-19
SARS-CoV-2

ABSTRACT

Six Italian non-accredited laboratories participated to an interlaboratory study aimed at measuring Differential Pressure (DP) and Bacterial Filtration Efficiency (BFE) of three face-mask models using methods in-line with EN 14683 standard. Methodological non-conformities were annotated. Repeatability and reproducibility on quintuplicate samples were calculated according to ISO 5725-2. Sample stability was also assessed. Laboratories were ranked according to the total standard deviation over all samples and proficiency was evaluated using z-score according to ISO 13528.

Although some non-conformities were present, performances for the DP measurements were always acceptable. One laboratory had to revise the bacterial suspension preparation for the BFE test. Overall, non-accredited laboratories working during pandemic emergency performed satisfactorily.

Sample-to-sample variability impacted measurement repeatability. BFE values above 98% showed good repeatability ($\leq$1.0%) and reproducibility ($\leq$6.1%), but high BFE uncertainty was associated to community masks. Our findings suggest that relevant face-mask conformity standards should consider uncertainty of BFE and DP measurements.

1. Introduction

With the spread of the COVID-19 pandemic in 2020, face masks have become familiar to the population worldwide [1,2]. They can be distinguished into respirators, designed to protect the wearer by filtering droplets and particles down to tenths of microns, and medical masks,

* Corresponding author at: Department of Industrial Engineering & BIOtech, University of Trento, via Sommarive, 38123 Trento, Italy.
E-mail address: francesco.tessarolo@unitn.it (F. Tessarolo).

https://doi.org/10.1016/j.measurement.2021.110481
Received 6 September 2021; Received in revised form 8 November 2021; Accepted 14 November 2021
Available online 25 November 2021

designed to prevent the wearer from spreading large droplets and aerosol typically > 3 μm [3]. Although medical masks, also known as surgical masks, are designed for professional use in the medical sector, their use by the general population has been strongly suggested by national and international health agencies and mandated by governments in many countries [4,5]. Surgical masks are medical devices subject to specific requirements set by standards [6,7]. Parallel to the use of surgical masks, the so-called non-medical mask or "*community mask*" has become widespread [3], driven by the sudden and intense market demand, in a context of scarce supply and disruption of the mask value chain [8,9]. In fact, although their protection against small particles is highly variable [10–12], their use is still recommended to protect other people from the transmission of the SARS-CoV-2 virus [10,13]. Community masks are not medical devices and the requirements for surgical masks do not apply to them [3]. No specific standards define their minimal requirements and only in June 2020 CEN published a workshop agreement on "Community face coverings—Guide to minimum requirements, methods of testing and use" [14]. While the debate on the effectiveness of face masks in the community was still ongoing, the ECDC - European Centre for Disease Prevention and Control has published a recommendation stating "*Although the evidence for the use of medical face masks in the community to prevent COVID-19 is limited, face masks should be considered as a non-pharmaceutical intervention in combination with other measures as part of efforts to control the COVID-19 pandemic.*" [3]. More recently, literature provided evidences that the use of face masks significantly lowers the transmission of the virus [15,16].

This paper does not intend to enter in such a debate, it rather aims to provide an analysis, from the point of view of the experimentalist, either an engineer, a bioengineer, a microbiologist or a metrologist, about the uncertainty of the measurements of face mask performance, obtained according to the methods described in the standard EN-14683:2019 [6]. In particular, the paper addresses the question of repeatability [17] of tests within the same laboratory, and reproducibility of tests performed in different laboratories, both on surgical and on community masks, as well as the overall laboratory proficiency achievable in the pandemic emergency context, when also non-accredited laboratories provided their service in testing performance of face masks, in order to allow manufacturers to enter the market; in Italy this was allowed by special legislation for emergency [18,19].

Surgical face masks are usually made of three or more layers of non-woven fabric [20]; the typical spunbond-meltblown-spunbond (SMS) mask is made of three layers, the intermediate acting as a filter to trap particles, droplets and aerosols emitted by the source (the wearer) during breathing, speaking, coughing, sneezing. Surgical face masks are in fact intended mainly to reduce the spread of potentially infected material, not to specifically protect the person wearing them.

European Standard EN 14683:2019 + AC "*Medical face masks - Requirements and test methods*" specifies performance requirements and test methods for medical face masks [6]. Although, under the pressure of the emergency context, recent research proposed different and viable methods for the fast measurement of mask performance [15,16,21,22], in this work we considered only the existing European standards, which today are still the official reference method in the European countries. It is worth noting that, given the renewed interest on the subject of mask testing methods, recently some discussion on the methods prescribed in the standard emerged [23,24]. In particular, the European Standard EN 14683:2019 + AC indicates:

- the minimum values for Bacterial Filtration Efficiency (BFE), which is related to the filtering capacity of the mask;
- the maximum values for Differential Pressure (DP), which is related to breathability and comfort of the wearer;
- the maximum bioburden of the ready to use mask, indicating the microbial cleanliness;
- the requirements for splash resistance, which is relevant in case the user is exposed to the risk of body fluids spillage, typically in healthcare environments.

This paper will focus on the measurement of DP and BFE, which are essential for classifying the surgical masks in three different types (Type I, Type II, and Type IIR) according to the limits reported in Table 1. Masks with performances not conforming those indicated in Table 1 could be considered and sold only as community masks.

DP and BFE tests should be performed to verify the compliance of the mask model to the minimal performance requirements in terms of breathability and filtration. These tests are therefore typical conformity assessment tests. In particular, the EN-14683:2019 + AC prescribes to perform a type test [6]. Type testing consists of testing one or more samples of the product taken randomly from a production batch, which should be representative of the batch itself. It is obvious that in conformity assessment, the uncertainty of measurement plays a relevant role and the level of confidence of the final outcome of the test is affected by the measurement uncertainty. Unfortunately, no clear indications on how to evaluate uncertainty of DP and BFE measurements are available in the standard [6]. This was the main reason why this article discussed the repeatability and reproducibility of DP and BFE tests, through the analysis of data acquired from 6 laboratories, which decided to set up an inter-laboratory comparison by taking measurements on the same types of masks. The present manuscript aimed also to compare laboratories in terms of their precision (repeatability), severity, and ability to rank BFE and DP characteristics of masks consistently with their peers. The whole study was also aimed at understanding the sources of variability in test results to improve BFE and DP methods' precision. Finally, to improve laboratories performance, warning messages or the suggestion to take action on testing procedures were produced based on the assessment of the laboratory *z-scores*.

2. Methodology

2.1. Standard specifications for the BFE testing methodology

The test method and the test rig for BFE measurement are described in Annex B of the European Standard EN 14683:2019 + AC and a schematic drawing is presented in Fig. 1. The measurement method is rather complex, involving engineering and microbiological aspects, in particular the generation and control of a two-phase flow, an aerosol stream containing a known charge of *Staphylococcus aureus* strain ATCC 6538, which is forced to pass through the mask material and a multiple stage cascade Andersen-type impactor [25,26]. Briefly, an airflow throughout the whole system should be generated by a vacuum pump and maintained at a constant flow rate of 28.3 l/min using a valve and a flowmeter. A microbial loaded aerosol (Staphylococcus aureus ATCC 6538) is obtained, using a nebulizer from a bacterial suspension grown in tryptic soy broth and then diluted in peptone water up to a concentration of 5×10^5 CFU/ml (CFU stands for Colony Forming Units). A motorized syringe or a peristaltic pump should feed the nebulizer with 0.01 ml/min of bacterial suspension. The aerosol is injected in a vertical glass cylinder with a diameter of 80 mm and a length of 600 mm, and mixed with air coming from an independent filtered inlet. At the

Table 1
Requirements for BFE and DP for medical face masks according to the standard EN 14683:2019 + AC "Medical face masks - Requirements and test methods" [6].

Mask type	Bacterial Filtration Efficiency (BFE) %	Differential Pressure (DP) Pa/cm^2
I	≥ 95	< 40
II	≥ 98	< 40
II R	≥ 98	< 60

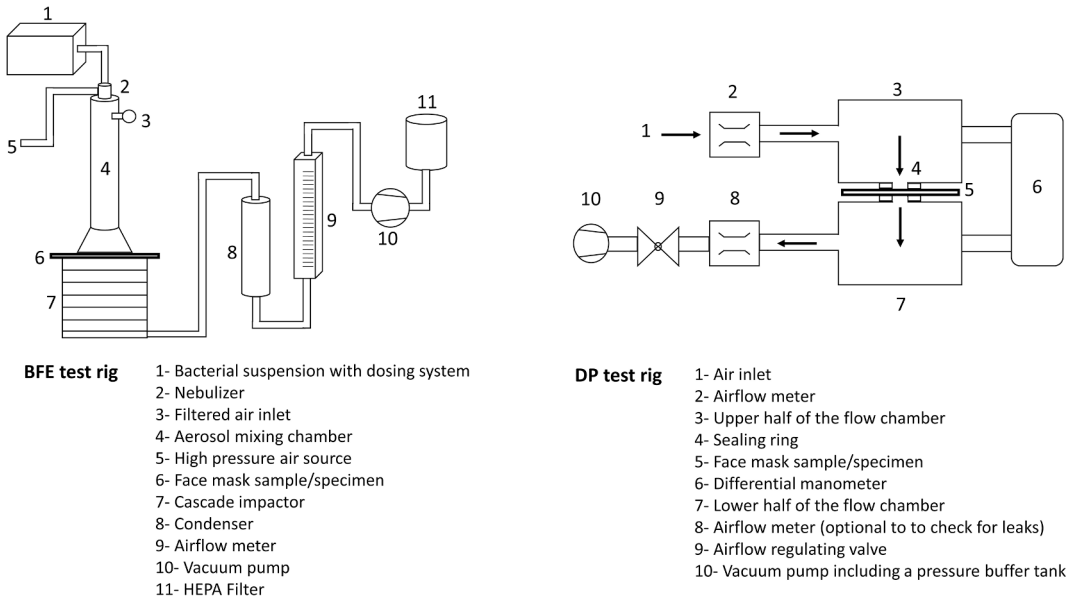

Fig. 1. Schematic drawing of the test rigs for testing Bacterial Filtration Efficiency (left) and Differential Pressure (right) according to the European Standard EN 14683:2019 + AC [6].

opposite end of the cylinder, the mixed aerosol, still high in water content, should enter an impactor consisting of six stages through which the incoming aerosol is drawn and separated into different class-sizes. Each of the six stages consists of 400 orifices and a Petri dish, containing an agar culture medium, used as impaction plates. Depending on orifices' diameter, droplets of a given class-size impact on the Petri dish and trigger the formation of the microbial colonies.

Test runs should be performed with the mask positioned at the entrance of the six stage impactor. Before and after the test runs, a positive control run should be obtained without the mask in place. The residual aerosol, exiting the impactor lower hose, should be safely condensed and filtered before reaching the vacuum pump. The standard method prescribes to run the aerosol for 1 min, followed by 1 min of airflow without aerosol generation. To check for possible contamination of the apparatus, a negative control run should be performed for each experimental session with no mask and a 2 min airflow with no aerosol generation. After the sampling period is completed, the plates should be removed from the impactor and incubated for 24 h. CFUs should be enumerated using the "positive hole" correction [27] for stages 3–6, as described by Andersen [25].

The distribution of the CFU counts in the six stages indicates the distribution of the impacting droplets divided by size classes. The Mean Particle Size (MPS) of the generated aerosol should be measured for each testing session without the mask, according to the following equation:

$$MPS = \frac{\sum_{i=1}^{6}(P_i \times C_i)}{\sum_{i=1}^{6} C_i}$$

where P_i is the particle diameter having the 50% probability of being captured by the i^{th} stage of the impactor, and C_i is the number of CFUs counted at the i^{th} stage. The Mean Particle Size of the mixed aerosol of positive control runs should be in the range from 2.7 μm to 3.3 μm and the average number of CFUs in the two positive controls should be between 1700 and 3000.

The bacterial filtration efficiency is measured indirectly, comparing results of the positive control to the results of tests when the mask acts as a filter. It is expressed as a percentage, computed according to equation below:

$$BFE(\%) = \frac{CFU_{control} - CFU_{test}}{CFU_{control}} \times 100$$

where the number of CFUs that pass through the material (CFU_{test}) of the face mask and the number of CFUs of the incoming aerosol ($CFU_{control}$) are taken into account.

According to the standard method, BFE should be measured on at least five samples and the mean value should be computed and compared to the requirements. The standard does not explicitly require an estimate of the uncertainty of the measurement and suggests checking the flowmeter calibration.

2.2. Standard specifications for the DP testing methodology

The test method and the rig for the measurement of the DP are specified in Annex C of the European Standard EN 14683:2019 + AC [6] and a schematic drawing is presented in Fig. 1. A constant airflow (8 l/min) is forced to pass through a circular section of the mask having (25 ± 1) mm diameter. The mask sample is inserted in an airflow cell, tightly pressed by two circular seals in order to prevent leakage across the mask tissue. The pressure drop across the mask sample is measured with a differential pressure transducer or two independent pressure transducers, then DP is computed as the ratio of pressure drop to test area and expressed as Pa/cm^2. The standard prescribes to perform DP measurements on at least 5 masks and, for each mask, 5 different areas should be tested. The mean DP should be computed and then compared to the requirements.

Both BFE and DP measurements should be conducted on masks conditioned at (21 ± 5) °C and (85 ± 5) % relative humidity for a minimum of 4 hrs before testing.

2.3. Participating laboratories and measurement set-ups

The following laboratories, participated to this interlaboratory round robin study:

- "AntiCovidLab - Torre Biologica", Università degli Studi di Catania and INFN-Laboratori Nazionali del Sud, Catania";
- "LABC19 Centro di Ricerca e Servizio per l'Emergenza COVID-19", Università Politecnica delle Marche"
- "LASS-TN-Covid-19 Laboratorio Associato per la verifica di Dispositivi di Protezione" Dipartimento di Ingegneria Industriale,

Università di Trento and Laboratorio di Sanità Pubblica, Azienda Provinciale per i Servizi Sanitari di Trento;
- "Safe s.r.l Laboratorio Chimico-Biologico", Mirandola;
- "Laboratorio di Igiene applicata", Dipartimento di Scienze Mediche e Sanità Pubblica, Università degli Studi di Cagliari;
- "U-TYM, Microbiology Unit", School of Pharmacy, Università di Camerino.

All these laboratories were set-up during the sanitary emergency caused by the spread of the SARS-CoV-2 virus in early 2020 [20]. The Italian government, to ensure that Italian manufacturers and importers could supply masks during the emergency, allowed non-accredited laboratories to access European standards on mask certification and to carry out tests by way of derogation from regular legislation on medical devices [28]. Indeed, none of the laboratories participating to this study was, at that time, accredited for testing face masks. Nevertheless, laboratory activity was permitted according to temporary national legislation [28], but they had to comply as much as possible to the standards [29]. However, in the emergency context of the pandemic causing difficulties in the supply chain and restrictions in transports and logistics, the equipment used for DP and BFE measurements was constructed partly by using already available components and measurement systems and partly by purchasing missing elements. All the six laboratories set up equipment for measuring DP and BFE, adhering as much as possible to the specifications of the standard EN 14683:2019 + AC. Each installation had specific characteristics, which have been detailed in a previous work [20] and will not be duplicated here.

2.4. Test samples and collected data

Three types of masks, having different performances in terms of DP and BFE were selected to be representative of both surgical and community masks (Fig. 2). Essential details of each mask are summarized in Table 2. They were expected to exhibit rather different performance in terms of DP and BFE, so to cover a range of values across different levels, as suggested by ISO 5725 standard series methodology. This approach allowed indeed to identify possible differences in the uncertainty of measurement depending on the value of the quantity being measured.

Fifteen masks of each type (same model and production lot) were selected randomly from the batches and sent to each of the participating laboratories. The amount was considered sufficient to perform measurements for DP and BFE according to standard EN 14683:2019 + AC [6]. Each laboratory was asked to report about BFE and DP measurements from 5 samples for each mask type. No additional indications to those indicated in the standard were provided to the laboratories on how to perform the tests. The laboratories were not informed on the characteristics of the masks and performed the measurements independently from each other, but in the same time frame. No requirements were set about the order the laboratories should test the different test samples. One laboratory was appointed to act as data collector.

Table 2
Main features of test samples used in the study.

	Mask model (study code)		
	RR1	**RR2**	**RR3**
Mask type (According to manufacturer claims)	Disposable face mask	Disposable surgical face mask, Type II	Community face mask
N° of layers	3	3	1
Main composition	Polypropylene	Polypropylene	Polypropylene
Layer detail	Spunbond-Meltblown-Spunbond	Spunbond-Meltblown-Spunbond	Spunbond
Mask size (cm^2)	17.5 × 19.0	17.5 × 15.5	17.5 × 19.0
Filtering area size (cm^2)	15.2 × 12.2	14.8 × 13.2	16.5 × 19.0
Fitting	Ear loops	Ear loops	Ear loops
Nose bridge	Yes	Yes	No

In addition to the five BFE and five DP measurements, laboratories were asked to report all raw data regarding colony count in each impactor stage for both test samples and controls, and all the five single DP measurements performed on each test sample. Collected data from each laboratory included also a detailed check-list of conformities and non-conformities of the BFE and DP measurement protocol with respect to EN 14683:2019 + AC specifications [6].

2.5. Data analysis and interpretative criteria

Collected data were double-checked for consistency. Information about the non-conformities of the measurement protocols was analysed and percentages of non-conformity over the considered requirements were calculated for each participating laboratory. Non-conformity percentages were also calculated for each requirement across the six laboratories.

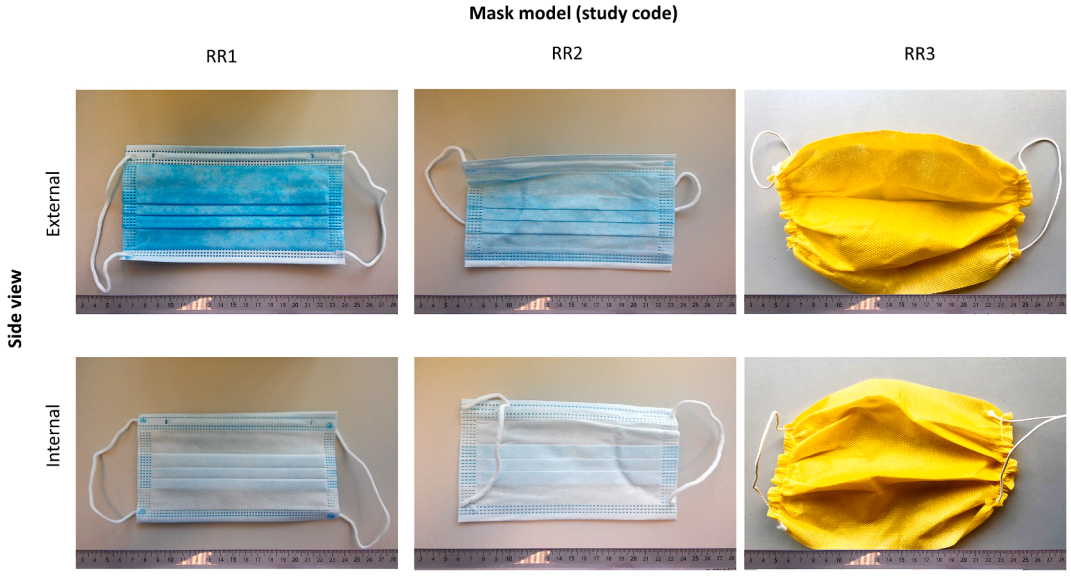

Fig. 2. The three selected mask models used as test samples in the interlaboratory study.

Results of the DP and BFE data were first analysed for the presence of outliers using Grubbs's test ($P < 0.01$) according to ISO 5725–2 specifications [30]. A check for outliers was performed on BFE and DP pooled data from all the laboratories ($N = 30$) for each of the three levels represented by the three mask types. Outliers were excluded from further analysis.

Repeatability standard deviations and reproducibility standard deviations were calculated according to ISO 5725–2 for both DP and BFE measurements for each of the three levels associated with the three mask types [30].

Participating laboratories were ranked for their performance over all tested samples in accordance with the recommendations on the round robin interlaboratory study published by The Coordinating European Council for the development of performance tests for fuels, lubricants, and other fluids [31] and the ISO 5725 series [30]. In particular, to express whether the position of the means of the sets of results from each laboratory was dissimilar to its peers, the i) "*Overall severity*" (S1) accounting for systematic differences, ii) "*Overall repeatability*" (S2)" accounting for test–retest reliability, iii) "*Severity SD*" (S3) accounting for ability to rank samples in the same order as its peers, and iv) "*Total SD*", being the geometric sum of the previous terms, were calculated for each laboratory, for both DP and BFE measurements [31]. Confidence intervals (95%) for repeatability and reproducibility were also calculated according to indications reported in the same document [31].

Overall severity is a measure of whether a laboratory produces test results that are consistently high or consistently low relative to its peers and was calculated for each laboratory according to the following formula:

$$S1_i = \sum_{j=1}^{q} \frac{\overline{y_{ij}} - \overline{m_j}}{q}$$

where, for q mask models ($q = 3$ in this study), $\overline{y}_{ij}$is the mean value for mask model j at laboratory i and m_j is the median value of $\overline{y}_{ij}$ across the p laboratories ($p = 6$ in this study).

Overall repeatability, in optimal conditions where test items variability is negligible, is an indicator for intra-laboratory variability. It was calculated according to the following formula:

$$S2_i = \sqrt{\sum_{j=1}^{q} \sum_{k=1}^{n_{ij}} \frac{\left(y_{ijk} - \overline{y_{ij}}\right)^2}{n_{ij} - 1} \Big/ q}$$

where n_{ij} is the number of tests conducted on mask model j at laboratory i..

Finally, *Severity SD* is a measure of a laboratory's ability to rank samples in the same order as its peers and is independent of the *Overall Severity* and *Overall Repeatability*. For each laboratory, *Severity SD* was calculated according to the following formula:

$$S3_i = \sqrt{var\left(d_{i1}, d_{i2}, \cdots, d_{iq}\right)/q - \frac{{S2_i}^2}{q/\sum_{j=1}^{q} 1/n_{ij}}}$$

where:

$$d_{ij} = \overline{y_{ij}} - \overline{m_j}$$

All the three components were then geometrically summed into the "*Total SD*" parameter, to provide an overall precision value for each laboratory (the higher the value the worse the precision and the lab ranking) according to the following formula:

$$Total\ SD_i = \sqrt{{S1_i}^2 + {S2_i}^2 + {S3_i}^2}$$

Considering that the BFE suffers of a "ceiling effect" (being upper limited to the value of 100%) the variability of the measure is expected to decrease approaching the value of 100%. In order to have proportional contributions from each mask model in the calculation of S1, S2, and S3 parameters, the BFE was transformed logarithmically following the formula ln*(101 - BFE)*.

Results were discussed taking into consideration possible variability within and amongst masks samples of the same type, due to manufacturing product variability. In particular, intra-sample variability was determined by computing the coefficient of variation (CV) over the five replicates collected on 5 different testing areas of DP, for each of the 5 samples of a single model and for each laboratory. Then the mean and the range values of these 30 CVs relative to each mask model were reported.

In addition to severity analysis, the proficiency of the participating laboratories was evaluated according to ISO 13528 indications [32]. The assessment included testing data normality, evaluating homogeneity and stability of test samples, and properly defining assigned values and corresponding standard uncertainties to all the three mask types.

Data normality was assessed by Shapiro-Wilk Test. Normality was checked for both BFE and DP measurements at two stages: first, for measurements collected on the same mask type for each laboratory ($N = 5$); then, for data pooled from all the participating labs ($N = 30$).

In order to address mask stability over time, two of the participating laboratories retrospectively compared data performed on two mask samples across a twelve-month period. Only for this specific evaluation, BFE measurements were repeated on the very same mask after decontamination by exposition to UV-ozone, and DP measurements were repeated on the same 25 mm diameter disks cut out from a single mask using a circular punch.

In addition, having recognized from previous testing activity [20] that discrepancies from the nominal airflow rate, required to test DP according to standard (8 l/m), have a large impact on DP measurement, one of the participating laboratories addressed in detail the relationship between airflow rate and DP measurements for the three mask models varying the flow rate from 2 to 10 l/m. Correlation between flow rate and DP measurement was checked using the Pearson's test.

Laboratory proficiency for BFE and DP measurements was finally evaluated using *z-score* in accordance with ISO 13528 [32]. Considering that the testing items were not certified by a metrology laboratory, and the assigned values and corresponding standard uncertainties were defined based on the data collected during the same interlaboratory study [33], *z-scores* associated to each single measurement were computed and the average *z-score* across the five measurements was calculated. The following criteria for interpretation of the average *z-score* results were adopted as suggested in the ISO 13528 [32]: $|z|<=2$ was considered acceptable, $2<|z|<3$ generated a warning signal for the participating lab and $|z|>=3$ was unacceptable and the participating laboratory was asked to take action. For those laboratories having an acceptable average *z-score*, the following finer criteria were also applied: i) if all single *z-scores* are smaller than 2 or equal, the Lab was notified with the valuation „*great success*"; ii) if one or more single *z-scores* are higher than 2, the lab was notified with the valuation „*success*".

3. Results

3.1. Conformity of measuring set-ups according to standard specifications

A summary of the compliance to EN 14683:2019 + AC instrumental and methodological specifications is presented in Tables 3 and 4, respectively for DP and BFE measurement protocols.

The overall compliance of laboratories was higher for DP (on average 88.3%, range 70.0–100.0%) than for BFE (on average 77.5%, range 64.7–94.1%). In particular, during BFE measurements, none of the laboratories implemented a nebulizer flow at the standard specified rate of 0.01 ml/min. This required adjusting the concentration of the *S. aureus* suspension (laboratories A, C, D, and F) and/or the exposition time to the aerosol challenge (laboratories C and E) to values different from those specified in the standard in order to guarantee the specified amount of CFUs on the positive control tests. Minor variations were

Table 3
Summary of the compliance to UNI EN 14,683 instrumental and methodological requirements for DP measurement protocol.

Item	Standard EN 14683 specification	Participating laboratory						Overall compliance with requirement (%)
		A	B	C	D	E	F	
Gas to be used	Air	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Flow generator	Vacuum pump	Yes	Yes	Yes	Yes	No	No	66.7
Sealing system	Metal ring on metal base	No	Yes	Yes	Yes	No	No	50.0
Sample compression system	Mechanical	Yes	Yes	Yes	Yes	No	No	66.7
Airflow rate	8 l/min	Yes	Yes	Yes	Yes	Yes	Yes	100.0
System for differential pressure measurement	Differential manometer or 2 separated manometers	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Test specimen	Whole mask or circular mask portions	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Test area	Circle with a diameter of (25 ± 1) mm	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Mask orientation	Internal mask side toward the incoming aerosol	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Test specimen conditioning	(21 ± 5) °C and (85 ± 5) % for at least 4 h	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Overall compliance with DP standard protocol by laboratory (%)		**90.0**	**100.0**	**100.0**	**100.0**	**70.0**	**700.0**	**88.3**

Table 4
Summary of the compliance to UNI EN 14683 instrumental and methodological requirements for BFE measurement protocol.

Item	Standard EN 14683 specification	Participating laboratory						OVERALL compliance with requirement (%)
		A	B	C	D	E	F	
Microbial strain	*Staphylococcus aureus* ATCC 6538	Yes	Yes	No	Yes	Yes	Yes	83.3
Solution used for the bacterial suspension	Peptone water	Yes	Yes	Yes	No	Yes	No	66.7
Culture medium	Tryptic soy agar	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Flow rate	28.3 l/m	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Impactor	Six stages Andersen (7.00, 4.70, 3.30, 2.10, 1.10, 0.65 μm)	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Aerosol chamber material	Glass	No	Yes	No	Yes	Yes	Yes	66.7
Aerosol chamber size	Cylinder, 600 mm long, 80 mm external diameter	Yes	Yes	No	Yes	Yes	Yes	83.3
Nebulizer flow rate	0,01 ml/min	No	No	No	No	No	No	0.0
Test specimen conditioning	(21 ± 5) °C and (85 ± 5)% for at least 4 h	No	Yes	Yes	Yes	Yes	Yes	83.3
Concentration of bacterial cell suspension	about 5 × 10^5 CFU/ml.	No	Yes	No	No	Yes	No	33.3
CFU on positive control	from 1,7 × 10^3 to 3,0 × 10^3 CFU	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Aerosol Mean Particle Size	(3,0 ± 0,3) μm	Yes	Yes	Yes	Yes	Yes	No	83.3
Flow generator	Vacuum pump	Yes	Yes	Yes	Yes	Yes	No	83.3
Exposition time to the aerosol challenge	1 min	Yes	Yes	No	Yes	No	Yes	66.7
Size of the tested area	>49 cm^2	Yes	Yes	Yes	Yes	Yes	Yes	100.0
CFU counting method	Positive hole conversion for stages 3 to 6	no	Yes	Yes	Yes	no	Yes	66.7
Mask orientation	Internal side toward the aerosol	Yes	Yes	Yes	Yes	Yes	Yes	100.0
Overall compliance with BFE standard protocol by laboratory (%)		**70.6**	**94.1**	**64.7**	**82.4**	**82.4**	**70.6**	**77.5**

adopted for the buffer solution used for the bacterial suspension (laboratories D, and F used saline solutions or a mix of saline and peptone water), the material and size of the aerosol mixing chamber (laboratory A used polymethylmethacrylate tube according to standard's dimensions, laboratory C used a polyethylene tube 360 mm long and 58 mm in diameter), the flow generator (lab F used compressed air), the bacterial strain (laboratory C used *Staphylococcus aureus* ATCC 25923), the mean particle size (laboratory F obtained a value of 2.5 ± 0.2 μm) and the CFUs counting method (laboratories A and E enumerated the colonies in all impactor stages after a short incubation time, without using the positive hole correction).

The DP instrumental and methodological requirements were mostly satisfied. Main deviations were found in the flow generator (laboratories E and F used compressed air instead of a vacuum pump), in the coupling/sealing system used to clamp the mask sample during the test (laboratories A, E, and F used a silicone elastomeric gasket instead of a metallic ring), and in the compression system (laboratories E and F used a pneumatic system instead of a mechanical one).

3.2. DP and BFE measurements precision

Laboratory testing activity was performed within 5 months from test samples delivery, in the period from June to November 2020. All the six participating laboratories provided the requested results without reporting sample loss or major inconveniences during the measurement procedures. BFE and DP data of each single laboratory were coherent to the collected raw data. The Grubbs's test identified three outliers within the DP datasets (the value 24.70 from laboratory A referred to mask RR1, and the values 12.43 and 1.47 from laboratory C referred to RR3 mask samples). No outliers were present in BFE data. These three measurements were excluded from further analysis.

Collected data are presented in Fig. 3, showing all BFE and DP measurements obtained at the six labs for each of the three mask samples. Measurements from all laboratories confirmed that major differences were present among the three mask types for both DP and BFE. High variability was found for DP measurements of RR2 masks and for BFE measurements of RR3 masks. On the opposite, low variability was present in DP measurements of RR3 masks and in BFE measurements of RR2 masks.

The inspection of the DP and BFE data broken down according to the single laboratory showed differences in both median values and variability of the five measurements collected by each laboratory for the three mask types (Fig. 4). Of note, DP measurements of laboratories E and F for mask RR2 were lower than values from laboratories B, C, and D, variability of DP measurements from laboratory C for mask RR3 were markedly higher than others, and BFE measurements from laboratory E

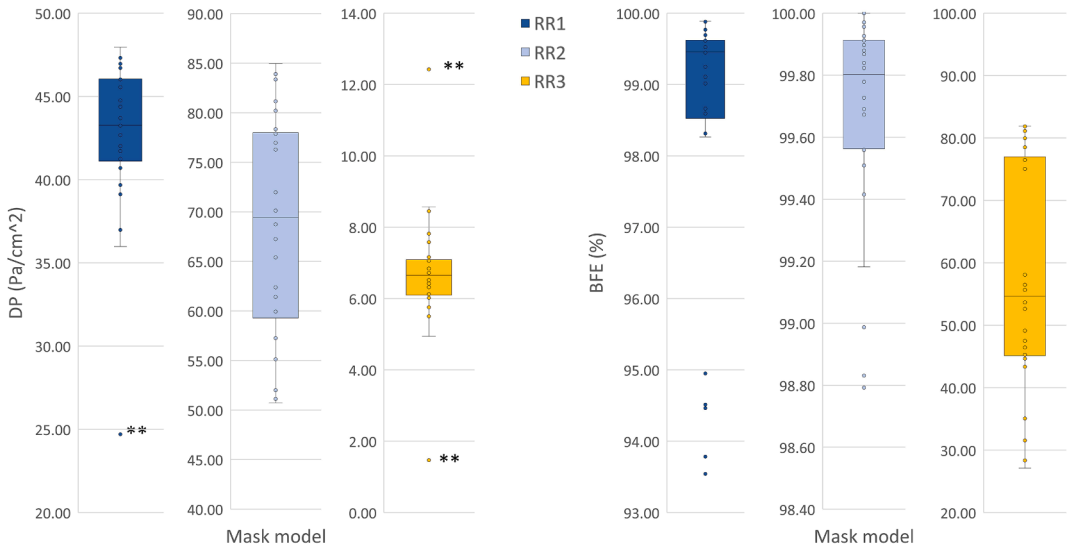

Fig. 3. DP (left) and BFE (right) data collected from the six participating laboratories grouped by mask type. Data for each mask type are presented on different y-axis ranges. Outliers are indicated by **.

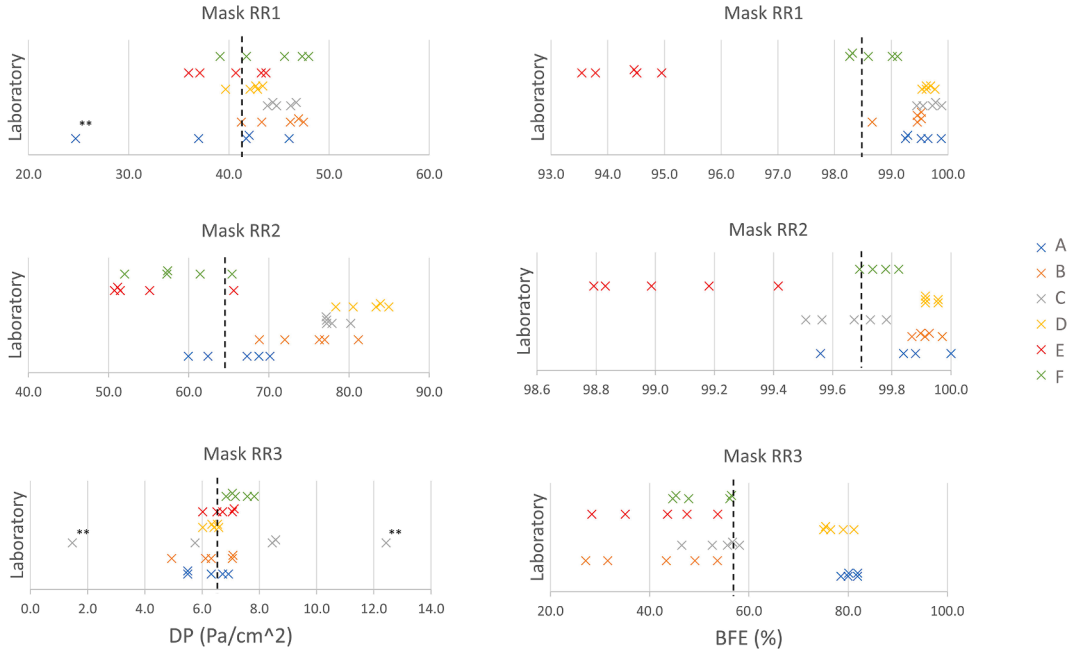

Fig. 4. DP (left) and BFE (right) measurements obtained at the six laboratories for each of the three mask samples. Mean values are indicated by the black dashed line. Outliers are indicated by **.

Table 5
Statistics of DP and BFE measurements for each of the mask model tested in the study. Data in [ ] are 95% confidence intervals. Percentages with respect to the mean values are reported in ().

Statistical descriptor	BFE test (%)			DP test (Pa/cm^2)		
	RR1	RR2	RR3	RR1	RR2	RR3
N° of laboratories	6	6	6	6	6	6
Total number of tests	30	30	30	30	30	30
Mean	98.51	99.69	57.42	43.13	69.06	6.68
Repeatability SD by level (% in respect to the mean)	0.98 [0.76–1.36] (1.0%)	0.39 [0.31–0.54] (0.4%)	19.58 [15.29–27.24] (34.1%)	8.05 [6.26–11.29] (18.7%)	12.29 [9.59–17.09] (17.8%)	2.03 [1.57–2.87] (30.4%)
Reproducibility SD (% in respect to the mean)	6.01 [3.78–14.34] (6.1%)	1.00 [0.65–2.12] (1.0%)	51.84 [33.73–110.76] (90.3%)	9.23 [7.10–13.17] (21.4%)	32.81 [21.33–70.26] (47.5%)	2.35 [1.80–3.40] (35.2%)

for both masks RR1 and RR2 were markedly lower than those reported by other laboratories. High variability among BFE values was present for the RR3 mask. The mean values of the CVs of DP measurements representative of the intra-mask variability were respectively 10.5%, 25.0% and 14.4% for the mask model RR1, RR2, and RR3 respectively. Likewise, the ranges were respectively 18.9%, 4.9% and 25.7%.

The calculated *mean* values, *repeatability standard deviation* (*rSD*), and *reproducibility standard deviations* (*RSD*) for each level (i.e. mask type) is reported in Table 5 for both DP and BFE measurements. Relative repeatability and reproducibility values were also reported as percentages with respect to the mean values. Confidence intervals (95% C.I.) for *rSD* and *RSD* are also indicated.

Relative repeatability of BFE measurements were under 1% for masks RR1 and RR2, indicating a good measurement precision within each laboratory for BFE values higher than 98%. Relative reproducibility among the participating laboratories of BFE was respectively 6.1% and 1.0% for masks RR1 and RR2. In contrast, a clear change in relative repeatability (relative *rSD* rising to about 34%) and a loss of reproducibility (relative *RSD* $>$ 90%) was found for the BFE measurements on mask RR3 having a mean BFE of 57.42%.

DP measurements showed lower, but more homogenous relative repeatability, ranging from 17.8% to 30.4% of the DP mean value, without an obvious correlation between the repeatability and the mean value of the measurement. Relative reproducibility was variable among the three tested DP levels, ranging from 21.4% of mask RR3 (mean DP value 43.13 Pa/cm^2), to 47.5% of mask RR2 (mean DP value 69.06 Pa/cm^2).

3.3. Lab ranking according to severity test

Laboratory comparison charts showing "*Total SD*" and all the three studied severity components (*S1*, *S2*, and *S3*) for the BFE and DP measurements are presented in Fig. 5. Taking into consideration that the lower the *Total SD* the better the overall performance of the laboratory among the three mask types, participating laboratories were ranked as C, B, F, D, A, and E for BFE measurements and as B, A, C, D, F and E for the DP measurements.

The inspection of the three severity components showed a minor contribution of the *Overall repeatability* (*S2*) than *S1* and *S3* in BFE measurements. On the other hand, the three severity components were almost equally present in DP measurements. Taking into consideration that *S2* is an overall estimation of the measurement repeatability (*rSD*) across the three levels (i.e. mask types), these findings are in agreement with repeatability data reported in Table 5.

3.4. Stability of test samples

Mask samples proved to be stable with time for repeated DP measures (over a six-months period). Variability of repeated DP measures performed by laboratory A on the very same 25 mm diameter samples from each of the mask models was always within or at most equivalent to the measuring instrument accuracy (flowmeter and differential manometer).

The experience of the participating laboratories indicated that one of the most relevant sources of variability was related to inadvertent variations in the flow rate used during the test. DP is indeed defined in the EN 14683:2019 + AC standard as the pressure drop per area unit (expressed in Pa/cm^2) caused by the mask at a steady airflow rate of 8 l/min over a circular specimen of 4.9 cm^2 area [6]. Data of pressure drop measured at flow rates above and below the nominal one are presented in Fig. 6 for each of the three masks tested in this study. There is a clear linear dependence of DP measurements on the flow rate, as expected for a fluid flow through a porous medium such as the layers that compose the masks under investigation. This phenomenon can be described by Darcy's law [34]:

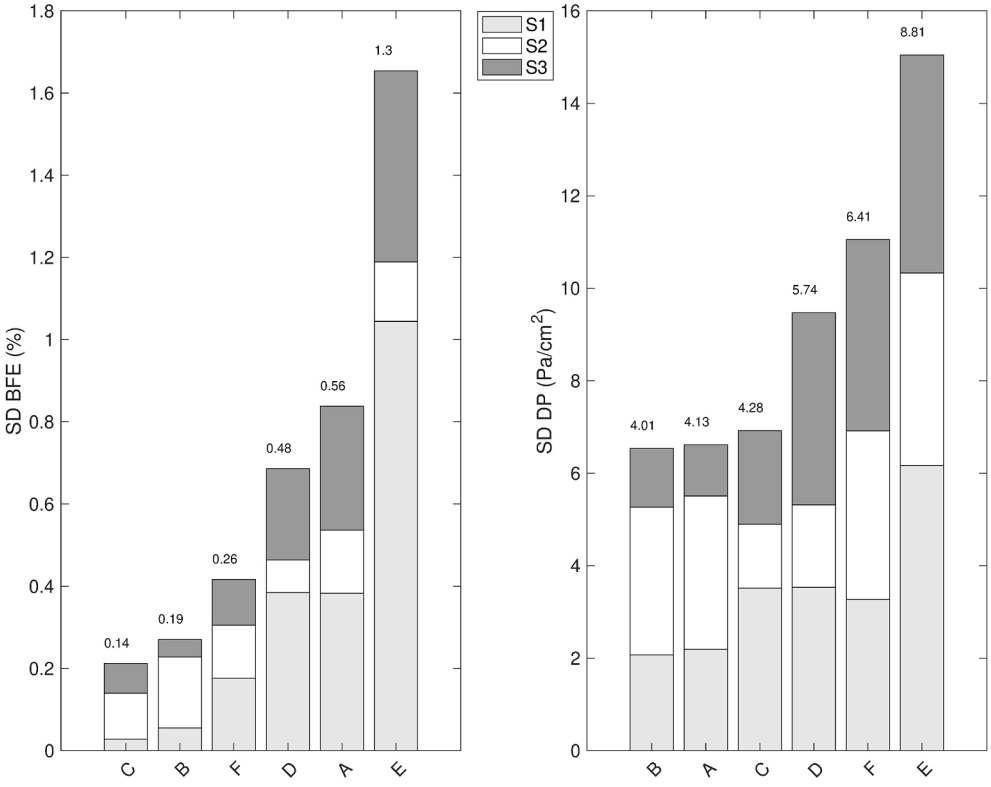

Fig. 5. Laboratory comparison charts showing "Total SD" and severity components (S1, S2, and S3) for the BFE (left) and DP (right) measurements obtained at the six labs for the three mask samples. Values reported on top of the columns indicate Total SD. Laboratories are presented according to Total SD ranking, the lower the better.

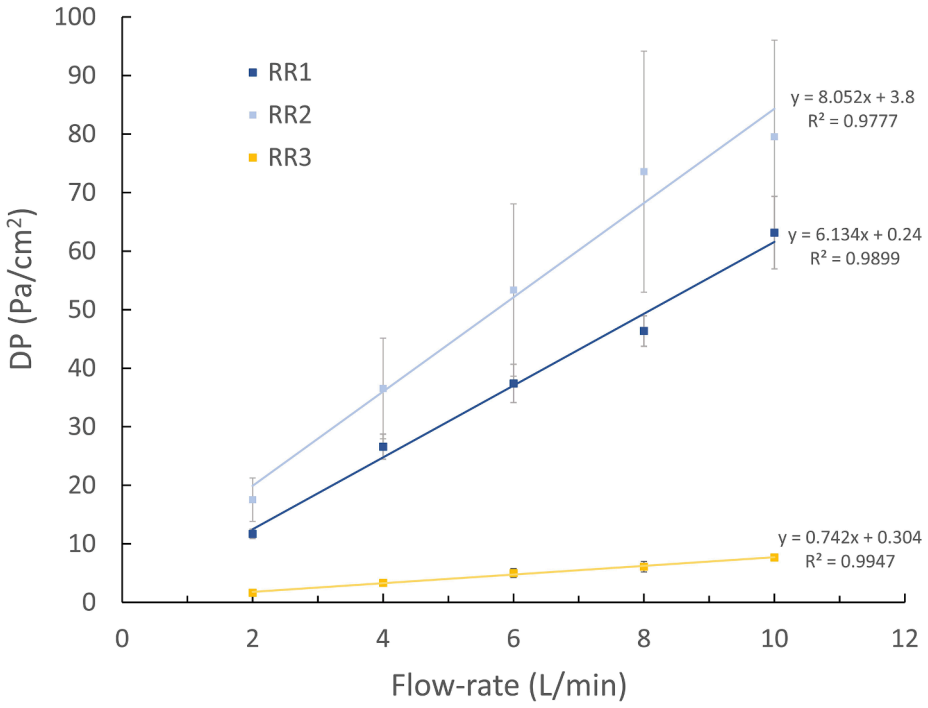

Fig. 6. Effect of airflow rate on DP. Each point represents the average out of five independent measurements. Error bars indicate the standard deviation of the mean. Main parameters derived from linear fitting of data are also reported.

$$Q = \frac{-kA}{\mu L}\Delta p$$

regulating the ease of movement of a fluid in a porous medium. Indeed, the volumetric flow rate Q of a fluid with a viscosity μ through the porous medium having a cross-sectional area A, a thickness L, and a permeability k, is proportional to applied pressure drop Δp.

Notably, masks RR1 and RR2 showed high DP values (about 70 Pa/cm^2 at 8 l·min^{-1} flow rate), while RR3 was below 10 Pa/cm^2. Measuring DP across a range of airflow, could be a valuable alternative to characterize mask breathability. This experimental approach, albeit not indicated by the standards, would be beneficial when testing new batches of filtering materials for surgical face masks in order to best match differential pressure measurement with instrument accuracy.

Differently from DP measurement, the BFE test is destructive for the sample, inducing heavy bacterial contamination of the mask, which has to be discarded as biological hazardous waste immediately after testing. Laboratory C assessed the possibility of making repeated BFE measurements on the very same sample by applying a 10 min UV-O3 (5 mW/cm^2 at a distance of around 10 cm) disinfection treatment after BFE testing

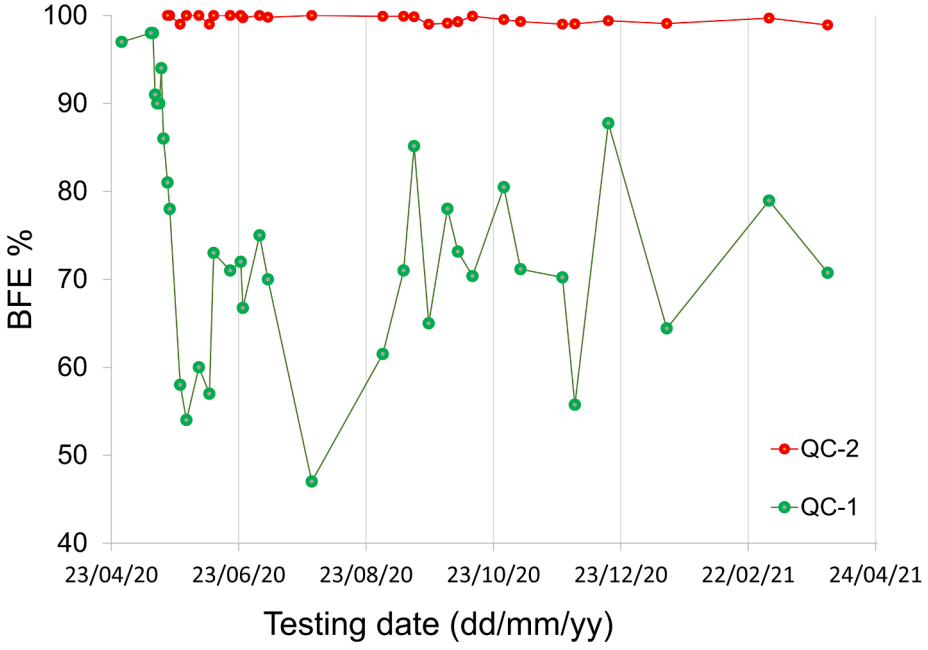

Fig. 7. Control charts for BFE stability test. Sample QC-1 suffered from a sudden drop in performance after a few measurements and disinfections were performed. The second QC-2 sample showed more stability over time.

on two samples identified as "QC-1" and "QC-2". QC-1 was a multilayer combination obtained by overlapping in sequence the three individual layers of spunbonded, meltblown and spunbonded polypropylene, and QC-2 was a Type-1 three-ply polypropylene surgical mask. Results are reported in Fig. 7. The test sample named QC-1 showed an abrupt loss of performance after the first 5 cycles of analysis and disinfection. On the contrary, QC-2 showed good stability over a year. These results showed that the disinfection process in between repeated BFE samples may affect mask properties in a different way, despite similar face mask composition and design are considered. In the light of these results, we considered the BFE measurement as a destructive process, not allowing for testing repeatability on the very same sample, but only on equivalent separate samples.

3.5. Laboratories' proficiency

Data normality tests evidenced that data were normally distributed only at the laboratory level. On the other hand, for each mask type, BFE and DP measurements pooled from all the six laboratories resulted far from being normally distributed, and typically showing multimodality, in agreement to differences already evidenced in Fig. 4. According to these findings, the *median* and the *normalized interquartile range* (*nIQR*) of the pooled measurements across all the six laboratories were adopted as the assigned value and corresponding standard uncertainty for DP and BFE measurements for each mask type in the calculation of the *z-scores* [32].

Laboratory performance in terms of averaged *z-scores* are summarized in Fig. 8. Averaged *z-score* varied across different laboratories and also according to the different masks within the same laboratory. All participating laboratories performance was acceptable for the DP measurements. Differently, laboratory E was informed about the need to take action for BFE measurement process, considering that two out of three mask models resulted in an averaged $|z|>3$. The remaining laboratories obtained acceptable *z-scores* for all the three tested mask models.

Among the acceptable *z-scores*, laboratories A, B, C, D, and F were valued with "great success" for their BFE measurements, given all single $|z|<2$. Laboratories D and F were valued with "great success" for their DP measurement.

4. Discussion

In this paper, the analysis of the uncertainty associated to DP and BFE measurements of face mask models was presented. The relevance of this subject has raised during the pandemic, because of the sudden increase in mask demand and the need to perform BFE and DP measurements at non-accredited laboratories making use of test rigs set-up in an emergency context. In Italy, non-accredited laboratories were allowed to operate during the emergency, provided they followed procedures according to the relevant standards. This paper reports about an inter-laboratory study between six of them to determine the level of confidence reachable by these emergency infrastructures. The collected data provided an insight on the uncertainty of the measurement methods prescribed by the standard.

The variability of BFE and DP measurements was assessed within each laboratory and amongst the six participating laboratories, when they all measured the same three types of masks, taken randomly from the same production batch. The implemented measuring procedures were in-line with the relevant standard, but were implemented using different equipment and operators and at different locations. A common specified timeframe was shared.

In a classical round-robin test, the same item is sent sequentially to each laboratory or certified test items are available [32]. Unfortunately, no certified samples for BFE and DP measurements were available for this study and real products should be used. Therefore, the mask samples under test were nominally of the same type, taken from the same production batch, i.e. same boxes, lots, and production dates (so called "*similar samples*" [35]), but could present individual unpredictable sample variability, related to the manufacturing process. These sample-to-sample differences (inter-samples non-homogeneity) played a role in BFE and DP measurement variability. The inspection of data referring to the DP measurements at different locations within the same mask showed also a non-negligible intra-sample non-homogeneity with mean CV values above 10% and CV ranges up to 25%.

Similar results were reported by a recent study addressing the repeatability and reproducibility of breathability measurements of surgical masks [36]. A high value of reproducibility error (19.1%, 8.0%, and 15.1% of the average value for surgical masks type I, II and IIR, respectively), led the authors to consider that different areas of the same

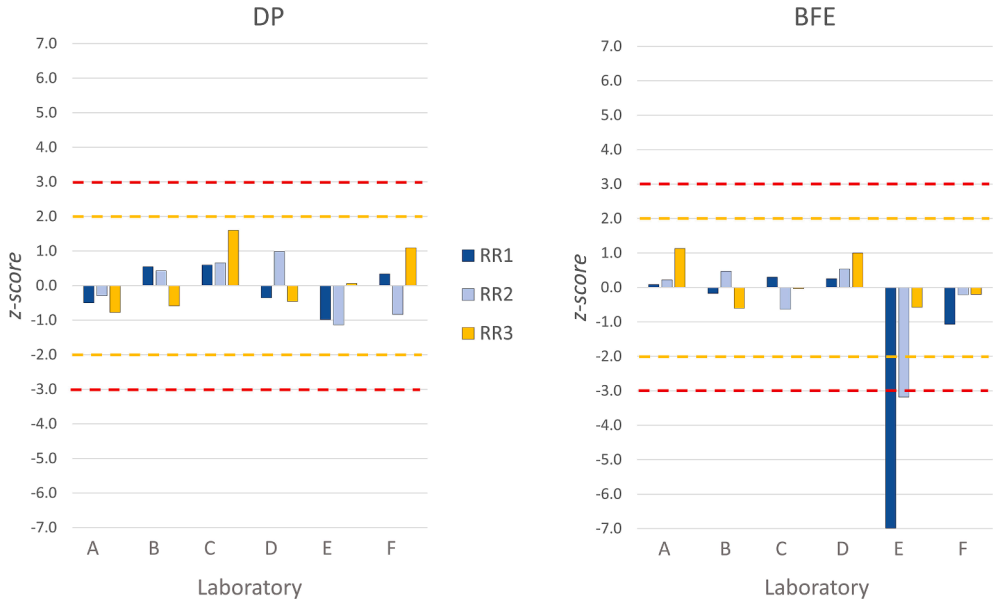

Fig. 8. Results of the proficiency test for all the participating labs for the DP (left) and BFE (right) measurements. Average z-score are indicated for each of the three mask types. Yellow and red dashed line represents thresholds for alert and intervention actions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mask were characterized by different values of breathability and thus the selection of the measurement points have an impact on the results of mask compliance [36].

The sample-related sources of variability represented a methodological limitation impacting the homogeneity of test samples, but at the same time were representative of the typical measuring conditions and sample sets tested by the laboratories dealing with DP and BFE measurements on face masks.

In addition to inter-samples and intra-sample non-homogeneities, the observed variations in the measurements of BFE and DP were generated by deviations from the standard EN14683 protocol (non-conformities listed in Tables 3 and 4), and more globally by the fact that none of the participating laboratories were accredited for these tests at the time of the study.

Unfortunately, due to the peculiar characteristics of the face masks and of the testing protocols, the specific contribution of sample non-homogeneities and methodological non-conformities to measurement variability cannot be properly discriminated in this study. However, it is worth considering that one of the participating laboratories (laboratory D) obtained the accreditation for the testing procedures according to EN 14683 standard on June 22nd, 2021, a few months after the completion of data collection for this interlaboratory study. The accreditation process provided the opportunity to focus on the applied testing method. Interestingly, no change to the test procedure or equipment was required. The laboratory was asked to re-calibrate the flowmeters and pressure sensors of the test rigs (previously calibrated by the manufacturers) at an accredited calibration national laboratory. Therefore, considering that the same operators, test site, and equipment were used for collecting data presented in this study, the results of DP and BFE test obtained at laboratory D during the round robin study should be similar to those of an accredited laboratory.

The stability of the test samples over time was also considered carefully, given that no certified reference material was available. As shown in Fig. 7, the BFE test cannot be safely repeated on the very same mask sample because of its destructive nature. The samples have to be considered a biohazard waste after the BFE test and promptly discarded. Although washing and disinfection could be tolerated by some mask models [12,34], several studies evidenced substantial changes in the DP or BFE properties after the decontamination procedure [11,37,38]. A significant decrease in the collection efficiency of the filter medium and a change in the most penetrating particle size has been shown for the so called 'electret' filters, which rely on the electrostatic charge [39]. Our data from stability tests are in agreement with previous finding, showing that decontamination procedures can affect mask material properties and can impact in a significant way the BFE performance. This aspect allowed neither to repeat BFE measurement on the same mask at a single laboratory, nor to circulate samples among laboratories. On the other hand, DP test is not destructive per se, but repeating the measurement under the very same conditions could be challenging due to difficulties in identifying the same portion of mask tested before or maintaining the correct alignment among the different mask layers. Intra sample variability further limits the reproducibility of DP measurements on the same sample.

These practical limitations on both BFE and DP measurements imposed the use of different equivalent, but distinct samples to obtain the five repeated BFE and DP measurements for each mask type at each laboratory. Therefore, part of the variability in BFE and particularly in DP measurements could be ascribed to real differences among mask samples. ISO 5725 part 1 paragraph 3.13 defines "*repeatability conditions*" as "*Conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time*" [40]. "*Repeatability*" is defined as "*measurement precision under a set of repeatability conditions of measurement*" [35]. By precision, it is meant the "*closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions*" [35]. Considering that BFE and DP tests cannot be repeated on the very same sample, we considered the five test items used for each of the measurements of this study as eligible for repeatability conditions. Based on this assumption, DP and BFE measurements were performed in quintuplicate for each of the three masks models at each of the six participating laboratories. From this perspective, the round robin study reported here, is not the classical interlaboratory comparison test and estimates of measurements repeatability and reproducibility are possibly overestimating the real figures that could be obtained in ideal conditions. Moreover, the reference values (assigned values and corresponding standard uncertainties) for DP and BFE were not known a priori. Given these limitations, overall, this exercise provides a quantitative overview of the uncertainty of these measurements and of the confidence level in the final result, which consists in the conformance assessment of the mask under test.

As mentioned above, a peculiarity of this round robin study was related to the differences in the methodological approach and in the set-up specifications among the participating laboratories, which were not fully replicated, although reaching a high level of conformity to standard specifications. These aspects had an impact on measurements reproducibility. The comparison among the results achieved in the different laboratories is an interesting example of analysis of the reproducibility achievable within the constraints and limitations imposed by the COVID-19 pandemic. Reproducibility is defined as "*measurement precision under reproducibility conditions of measurement*" [35]. Again, it has to do with precision, but is evaluated in a different context. In fact, by reproducibility conditions we should consider "*condition of measurement, … different locations, operators, measuring systems, and replicate measurements on the same or similar objects*" [35]. Therefore, reproducibility has to do with the possible differences in the measurements on similar objects obtained in different laboratories, as presented in this paper.

This study evidenced a good reproducibility of BFE measurements among the participating laboratories for BFE $>$ 98%, but a scarce reproducibility when measuring BFE around 60%. This result should be carefully considered when the BFE methodology, described in standard EN 14683:2019 + AC [6], and supposed to be used only for surgical masks with BFE $>$ 95%, is also proposed for community masks, having lower requirements in terms of filtration efficiency, typically below 80% [14].

A different reproducibility pattern was obtained for DP measurements on the three mask types, appearing not obviously dependent on the mean DP value. Possibly, the observed lower reproducibility for mask RR2 could be ascribed to sample-to-sample variability and could be lowered by increasing the number of test items.

Standard EN 14683, although recommending to test a good numerosity of samples (at least five masks), does not provide any indication about how to calculate and report, and consider the uncertainty of DP and BFE measurements. Results of this study evidenced a non-negligible variability of DP and BFE mask measurements. The sole comparison of the average DP and BFE values with the minimal performance requirements specified in the current EN 14683 standard could be improved by properly considering both the mean values and the uncertainties associated to measurements on multiple masks samples.

From the equipment perspective, uncertainty of DP measurements depends directly on the uncertainty of the pressure measurement and indirectly from the airflow rate regulation, as clearly shown by the linear relation between pressure and flow rate documented in Fig. 6. This linear relation was indeed an experimental confirmation of the Darcy's law [34] stating that the flow rate through a porous medium is proportional to pressure drop across the medium. This relation has been already proven valid for face masks and filtering materials [41,42]. A number of parameters may affect the airflow and therefore generate bias and/or random fluctuations of the pressure drop. Among the main influencing parameters, we may highlight airflow instability with time, possible air leaks in the plane of the mask sample and in air conducts

before the flowmeter, or the use of an uncalibrated flowmeter.

It is also interesting to observe that the standard EN 14683:2019 + AC allows for tolerance in the diameter of the circular sample to be tested, which is specified to be (25 ± 1 mm) [6]. Such a tolerance in the diameter size accounts for ± 8 % tolerance in the area undergoing DP testing. If the airflow rate is kept constant at 8 l/min, a possible tolerance in an area of ± 8 % determines a possible bias in the pressure drop by the same amount.

Uncertainty in BFE measurement can directly depend on errors in enumerating CFUs on the agar medium. The application of the positive hole conversion for stages 3 to 6 [25], instead of the direct enumeration of the identifiable CFUs, can also bring to non-negligible differences in raw data of positive controls. In addition, a number of relevant influencing parameters may disturb the test and therefore generate bias and/or random variability in the process of bacterial colony formation, therefore affecting the CFU counts. Among the main influencing parameters, we may highlight physical parameters, such as airflow rate (magnitude and stability), variable and unsteady aerosol flow rate, and droplet size distribution (only MPS is prescribed and controlled, but different aerosol size distributions can share the same MPS value), as well as biological parameters (bacteria viability, growth medium composition, incubation conditions) [24]. Although standard EN 14683:2019 + AC is well detailed in many methodological aspects, it neither explicitly requires monitoring of the performance of the BFE analytical process with time, nor it makes explicit reference to the need for calibration protocols or use of certified test items [6], possibly due to the same limitations in sample stability we discussed above.

The ranking of the laboratories according to severity parameters was impacted by the large values of reproducibility and repeatability of RR3 and RR2 mask type respectively for BFE and DP measurements. In particular, laboratory E suffered from scarce repeatability in characterizing BFE of RR1 and RR2 masks models, reporting an apparent bias towards lower BFE values (Fig. 3). This is reflected in the high contribution of S1 to the *Total SD* of the same lab for BFE measurements (Fig. 5). Similarly, the evaluation of z-scores identified the need to take corrective actions for laboratory E when performing measurements of BFE > 98% (Fig. 7). Corrective actions for Laboratory E included the inspection and re-check of the whole BFE apparatus. The improvement of the buffer solution composition (addition of a saline fraction to the peptone buffer) was implemented based on the experience of other participating laboratories and resulted in the reduction of BFE measurements bias, as verified after the conclusion of the RR study. Possibly, the formulation of the buffer solution to be used in preparing the microbial suspension should be optimized according to the nebulizer. Being not fully compliant to nebulizer specifications set in the standard EN 14683 (see Table 4), requires testing adequacy of the buffer solution.

5. Conclusions

This study faced the complexity of performing an interlaboratory study among non-accredited laboratories, to identify reproducibility and repeatability of measuring DP and BFE on surgical and community masks according to EN 14683 methodology. Although some non-conformities were present in the implemented methodologies and equipment, performance of laboratories for the DP measurements were always acceptable. In contrast, one laboratory had to revise part of the procedure of the BFE test to avoid significant bias in the measurements despite essential parameters, as MPS and total CFU counts on the positive controls, were in line with the standard's methodological specifications. Overall, non-accredited laboratories working during pandemic emergency performed satisfactorily, taking into consideration equipment limitations and restrictions imposed by the emergency context. The study evidenced that measurements repeatability was impacted by variability of mask samples characteristics. Although test samples were taken from the same manufacturing lot and package, both sample-to-sample variability (inter-sample variability) and non-negligible non-homogeneities of the filtering materials within a single mask sample (intra-sample variability) were present in all tested products. This intrinsic variability highlighted the need to perform DP and BFE measurements on several sample replicas, as indicated by the relevant standards, to improve measurement precision. BFE values above 98% showed good repeatability (≤1.0%) and reproducibility (≤6.1%), but high BFE uncertainty was associated to tested community masks. These findings, warn about the possibility to rely on BFE measurement according to the standard method for measuring filtration performance of community masks, usually having BFE < 80%. Overall, we advise that relevant face-mask conformity standards should consider uncertainty of BFE and DP measurements when defining criterial for minimal performance requirements.

CRediT authorship contribution statement

Francesco Tessarolo: Conceptualization, Methodology, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing. **Giandomenico Nollo:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing. **Luca Benedetti:** Methodology, Investigation, Resources, Writing – review & editing. **Fabrizia Helfer:** Methodology, Investigation, Resources, Writing – review & editing. **Luigi Rovati:** Methodology, Investigation, Resources, Writing – review & editing. **Alberto Ferrari:** Methodology, Investigation, Resources, Data curation, Writing – review & editing. **Giulio Marchetti:** Methodology, Investigation, Resources, Writing – review & editing. **Federico Reverberi:** Methodology, Investigation, Resources, Writing – review & editing. **Salvo Baglio:** Methodology, Investigation, Resources, Writing – review & editing. **Nunzio Tuccitto:** Methodology, Investigation, Resources, Writing – original draft, Writing – review & editing. **Stefania Stefani:** Methodology, Investigation, Resources, Writing – review & editing. **Stefano Stracquadanio:** Methodology, Investigation, Resources, Writing – review & editing. **Filippo Caraci:** Methodology, Investigation, Resources, Writing – review & editing. **Antonio Terrasi:** Methodology, Investigation, Resources, Writing – review & editing. **Alessia Tricomi:** Methodology, Investigation, Resources, Writing – review & editing. **Mario Musumeci:** Methodology, Investigation, Resources, Writing – review & editing. **Andrea Miraglia:** Methodology, Investigation, Resources, Writing – review & editing. **Giacomo Cuttone:** Methodology, Investigation, Resources, Writing – review & editing. **Sofia Cosentino:** Methodology, Investigation, Resources, Writing – review & editing. **Luca Agostino Vitali:** Methodology, Investigation, Resources, Writing – original draft, Writing – review & editing. **Dezemona Petrelli:** Methodology, Investigation, Resources, Writing – review & editing. **Sonia Ilaria Ciancia:** Methodology, Investigation, Resources, Writing – review & editing. **Marina Mingoia:** Methodology, Investigation, Resources, Writing – review & editing. **Paolo Castellini:** Methodology, Investigation, Resources, Writing – review & editing. **Serena Simoni:** Methodology, Investigation, Resources, Writing – review & editing. **Luigi Montalto:** Methodology, Investigation, Resources, Writing – review & editing. **Alessia Baleani:** Methodology, Investigation, Resources, Data curation, Writing – review & editing. **Paolo Chiariotti:** Methodology, Investigation, Resources, Writing – review & editing. **Nicola Paone:** Conceptualization, Methodology, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Dr.ssa Giulia Pelliciari and Dr. Mario Minichiello from the

"Laboratorio Chimico-Biologico Safe S.r.l.", and Dr. Ivan Corradi from "Laboratorio di Sanità Pubblica, Azienda Provinciale per i Servizi Sanitari di Trento" are kindly acknowledged for their technical support.

Funding

"LASS-TN-Covid-19 Laboratorio Associato per la verifica di Dispositivi di Protezione" received financial support from the internal grant "Bando interno 2020 Università di Trento Covid 19".

"LABC19 Centro di Ricerca e Servizio per l'Emergenza COVID-19" was funded in part by the Italian Ministry of Research, under article 11 FFO 2020.

"U-TYM, Microbiology Unit" was partly funded by UNICAM-FPC200014.

"AntiCovidLab - Torre Biologica", was partly funded by the Italian Ministry of Research: call FISR2020IP_02846, project title "Sviluppo di Mascherine Antidroplet in Plastica Sostenibile Ottenuta dagli Scarti del Latte", CUP E65F21001160005

References

[1] J. Yu, Y. Lu, J. Muñoz-Justicia, Analyzing spanish news frames on twitter during COVID-19—A network study of El País and El Mundo, Int. J. Environ. Res. Public Health 17 (2020) 1–12, https://doi.org/10.3390/IJERPH17155414.

[2] W. Ahmed, J. Vidal-Alaball, F. Lopez Segui, P.A. Moreno-Sánchez, A social network analysis of tweets related to masks during the COVID-19 pandemic, Int. J. Environ. Res. Public Health 17 (21) (2020) 8235, https://doi.org/10.3390/ijerph17218235.

[3] European Center for Disease prevention and Control, Using face masks in the community, ECDC-European Center for Disease prevention and Control, Technical report, (2020). https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-use-face-masks-community.pdf (accessed September 2, 2021).

[4] Emergencies Preparedness, WHO Headquarters, Mask use in the context of COVID-19, (2020). https://www.who.int/publications-detail-redirect/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-(2019-ncov)-outbreak (accessed August 30, 2021).

[5] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, B.J. Cowling, Rational use of face masks in the COVID-19 pandemic, The Lancet Respir. Med. 8 (5) (2020) 434–436, https://doi.org/10.1016/S2213-2600(20)30134-X.

[6] European Committee for Standardization, European Standard EN 14683:2019+AC Medical face masks - Requirements and test methods, 2019.

[7] International Standard Organization, ISO 10993-1:2018 Biological evaluation of medical devices — Part 1: Evaluation and testing within a risk management process, 2018.

[8] The Organisation for Economic Co-operation and Development, The face mask global value chain in the COVID-19 outbreak: Evidence and policy lessons, OECD. (n.d.). http://www.oecd.org/coronavirus/policy-responses/the-face-mask-global-value-chain-in-the-covid-19-outbreak-evidence-and-policy-lessons-a4df866d/ (accessed November 9, 2020).

[9] A. Kirubarajan, S. Khan, T. Got, M. Yau, J.M. Bryan, S.M. Friedman, Mask shortage during epidemics and pandemics: a scoping review of interventions to overcome limited supply, BMJ Open. 10 (11) (2020) e040547, https://doi.org/10.1136/bmjopen-2020-040547.

[10] F.G. Morais, V.K. Sakano, L.N.d. Lima, M.A. Franco, D.C. Reis, L.M. Zanchetta, F. Jorge, E. Landulfo, L.H. Catalani, H.M.J. Barbosa, V.M. John, P. Artaxo, Filtration efficiency of a large set of COVID-19 face masks commonly used in Brazil, Aerosol Sci. Technol. 55 (9) (2021) 1028–1041, https://doi.org/10.1080/02786826.2021.1915466.

[11] B.B. Neupane, S. Mainali, A. Sharma, B. Giri, Optical microscopic study of surface morphology and filtering efficiency of face masks, PeerJ. 7 (2019) e7142. https://doi.org/10.7717/peerj.7142.

[12] S. Sankhyan, K.N. Heinselman, P.N. Ciesielski, T. Barnes, M.E. Himmel, H. Teed, S. Patel, M.E. Vance, Filtration Performance of Layering Masks and Face Coverings and the Reusability of Cotton Masks after Repeated Washing and Drying, Aerosol Air Qual. Res. 21 (2021) 210117. https://doi.org/10.4209/aaqr.210117.

[13] C.J. Kähler, R. Hain, Fundamental protective mechanisms of face masks against droplet infections, J. Aerosol Sci. 148 (2020) 105617, https://doi.org/10.1016/j.jaerosci.2020.105617.

[14] European Committee for Standardisation, Workshop Agreement CWA 17553:2020 E, "Community face coverings - Guide to minimum requirements, methods of testing and use," (2020). https://www.cencenelec.eu/research/CWA/Documents/CWA17553_2020.pdf (accessed August 30, 2021).

[15] B. Illés, P. Gordon, Filtering efficiency measurement of respirators by laser-based particle counting method, Measurement: J. Int. Meas. Confederation 176 (2021) 109173, https://doi.org/10.1016/j.measurement.2021.109173.

[16] E.P. Fischer, M.C. Fischer, D. Grass, I. Henrion, W.S. Warren, E. Westman, Low-cost measurement of face mask efficacy for filtering expelled droplets during speech, Sci. Adv. 6 (36) (2020), https://doi.org/10.1126/sciadv.abd3083.

[17] Comitato Elettrotecnico Italiano, VIM - International Vocabulary of Metrology, (n.d.). https://www.ceinorme.it/en/normazione-en/vim-en/vim-content-en.html (accessed September 2, 2021).

[18] Gazzetta Ufficiale della Repubblica Italiana, Decreto Legge n.9 del 2 marzo 2020 (art. 34, comma 3), (2020). https://www.gazzettaufficiale.it/eli/id/2020/03/02/20G00026/sg.

[19] Gazzetta Ufficiale della Repubblica Italiana, Decreto Legge del 17 marzo 2020 n.18 (art. 15), (2020). https://www.gazzettaufficiale.it/eli/id/2020/03/17/20G00034/sg.

[20] F. Tessarolo, G. Nollo, D. Maniglio, M. Rigoni, L. Benedetti, F. Helfer, I. Corradi, L. Rovati, A. Ferrari, M. Piccini, L. Accorsi, E. Veronesi, A. Cuoghi, S. Baglio, N. Tuccitto, S. Stefani, S. Stracquadanio, F. Caraci, A. Terrasi, A. Tricomi, M. Musumeci, A. Miraglia, G. Cuttone, S. Cosentino, C. Muscas, L.A. Vitali, D. Petrelli, L. Angrisani, R. Colicchio, A. D'Anna, I. Iavicoli, G. De Falco, F. Di Natale, E. Di Maio, P. Salvatore, F. Quaglia, M. Mingoia, P. Castellini, P. Chiariotti, S. Simoni, L. Montalto, A. Baleani, N. Paone, Testing surgical face masks in an emergency context: the experience of italian laboratories during the COVID-19 pandemic crisis, Int. J. Environ. Res. Public Health. 18 (2021) 1462, https://doi.org/10.3390/ijerph18041462.

[21] P. Anfinrud, V. Stadnytskyi, C.E. Bax, A. Bax, Visualizing speech-generated oral fluid droplets with laser light scattering, N. Engl. J. Med. 382 (21) (2020) 2061–2063, https://doi.org/10.1056/NEJMc2007800.

[22] A.V. Mueller, M.J. Eden, J.M. Oakes, C. Bellini, L.A. Fernandez, Quantitative method for comparative assessment of particle removal efficiency of fabric masks as alternatives to standard surgical masks for PPE, Matter 3 (3) (2020) 950–962, https://doi.org/10.1016/j.matt.2020.07.006.

[23] P. Forouzandeh, K. O'Dowd, S.C. Pillai, Face masks and respirators in the fight against the COVID-19 pandemic: An overview of the standards and testing methods, Saf. Sci. 133 (2021) 104995, https://doi.org/10.1016/j.ssci.2020.104995.

[24] J. Pourchez, A. Peyron, Y. Montigaud, C. Laurent, E. Audoux, L. Leclerc, P. O. Verhoeven, New insights into the standard method of assessing bacterial filtration efficiency of medical face masks, Sci. Rep. 11 (1) (2021), https://doi.org/10.1038/s41598-021-85327-x.

[25] A.A. Andersen, New sampler for the collection, sizing, and enumeration of viable airborn particles, J. Bacteriol. 76 (1958) 471–484.

[26] V.A. Marple, History of impactors—the first 110 years, Aerosol Sci. Technol. 38 (3) (2004) 247–292, https://doi.org/10.1080/02786820490424347.

[27] Janet.M. Macher, Positive-hole correction of multiple-jet impactors for collecting viable microorganisms, Am. Ind. Hyg. Assoc. J. 50 (11) (1989) 561–568, https://doi.org/10.1080/15298668991375164.

[28] Gazzetta Ufficiale della Repubblica Italiana, Decreto Legge n. 34 del 19 maggio 2020, Misure urgenti in materia di salute, sostegno al lavoro e all'economia, nonche' di politiche sociali connesse all'emergenza epidemiologica da COVID-19., (2020). https://www.gazzettaufficiale.it/eli/id/2020/05/19/20G00052/sg (accessed September 2, 2021).

[29] Istituto Superiore di Sanità, Procedure per richiesta produzione mascherine - ISS, (n.d.). https://www.iss.it/procedure-per-richiesta-produzione-mascherine (accessed November 9, 2020).

[30] International Organization for Standardization, ISO 5725-2:2019 Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method, (2019).

[31] The Coordinating European Council for the development of performance tests for fuels, lubricants, and other fluids, Conduct of Precision Testing Programmes and Procedure, (2014). https://www.cectests.org/public/info_/g003/cec14_3331_003_file1.pdf (accessed July 25, 2021).

[32] International Organization for Standardization, ISO 13528:2015 Statistical methods for use in proficiency testing by interlaboratory comparison, (2015).

[33] S.F. Beckert, G.E. Fischer, Interlaboratory comparison of roughness measurement: Application of Algorithm A of ISO 13528: 2015 in determining the designated value and the standard deviation, in: 2018. https://doi.org/10.1088/1742-6596/1065/8/082007.

[34] S. Whitaker, Flow in porous media I: A theoretical derivation of Darcy's law, Transp. Porous Media 1 (1) (1986) 3–25.

[35] JCGM - Joint Committee for Guides in Metrology, (n.d.). https://www.iso.org/sites/JCGM/GUM-introduction.htm (accessed August 30, 2021).

[36] J. Taborri, B. Stocchi, G. Calabro, S. Rossi, Repeatability and reproducibility in the breathability measurement of surgical masks, in: 2021. https://doi.org/10.1109/MeMeA52024.2021.9478732.

[37] W. Hao, G. Xu, Y. Wang, Factors influencing the filtration performance of homemade face masks, J. Occup. Environ. Hyg. 18 (3) (2021) 128–138, https://doi.org/10.1080/15459624.2020.1868482.

[38] S.A. Grinshpun, M. Yermakov, M. Khodoun, Autoclave sterilization and ethanol treatment of re-used surgical masks and N95 respirators during COVID-19: impact on their performance and integrity, J. Hosp.. Infect. 105 (4) (2020) 608–614, https://doi.org/10.1016/j.jhin.2020.06.030.

[39] S.B. Martin, E.S. Moyer, Electrostatic respirator filter media: filter efficiency and most penetrating particle size effects, Appl. Occup. Environ. Hyg. 15 (8) (2000) 609–617, https://doi.org/10.1080/10473220050075617.

[40] International Organization for Standardization, ISO 5725-1:1994, Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions, 1994.
[41] L.H. Kwong, R. Wilson, S. Kumar, Y.S. Crider, Y. Reyes Sanchez, D. Rempel, A. Pillarisetti, Review of the breathability and filtration efficiency of common household materials for face masks, ACS Nano 15 (4) (2021) 5904–5924, https://doi.org/10.1021/acsnano.0c1014610.1021/acsnano.0c10146.s001.
[42] J.T. Mueller, S. Karimi, K.A. Poterack, M.T.A. Seville, S.M. Tipton, Surgical mask covering of N95 filtering facepiece respirators: The risk of increased leakage, Infection Control and Hospital Epidemiology 42 (5) (2021) 627–628, https://doi.org/10.1017/ice:2021.50.