COVID-19 and mental disorders in Healthcare Personnel: A Novel Framework to Develop Personas from an Online Survey

Emanuele Tauro[1], Alessandra Gorini[2,3], Chiara Caglio[3], Paola Gabanelli[4] , Enrico Gianluca Caiani[1,5]

1 Electronics, Information and Bioengineering Dpt., Politecnico di Milano, 20133 Milano, Italy;

2 IRCCS Istituti Clinici Scientifici Maugeri di Milano, 20138 Milan, Italy

3 Department of Oncology and Hemato-Oncology, Università degli Studi di Milano

4Psychology Unit, Istituti Clinici Scientifici Maugeri, IRCCS, Pavia, Italy

5 Consiglio Nazionale delle Ricerche, Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni, 20133 Milan, Italy

Corresponding author:

Prof. Enrico G Caiani

Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria

Piazza L. Da Vinci 32, 20133 Milano, Italy

+39.3387163426

enrico.caiani@polimi.it

Abstract

Background: In this paper we propose a novel framework for the definition of Personas for healthcare workers based on an online survey, with the aim of highlighting different levels of risk of developing mental disorders induced by COVID-19 and tailor psychological support interventions.

Methods: Data were gathered from Italian healthcare workers between April and May 2020. Information about socio-demographic characteristics, current lifestyle, occupational, COVID-19 infection, and psychological indexes (Maslach Burnout Inventory, Impact of Event Scale and Patient Health Questionnaire) was collected. Respondents were divided in four subgroups based on their health profession: physicians (P), nurses (N), other medical professionals (OMP) and technical-administrative (TA). For each sub-group, collected variables (46) were reduced using Principal Component Analysis and clustered by means of k-medoids clustering. Statistical analysis was then applied to define which variables were able to differentiate among the k clusters, leading to the generation of a Persona card (i.e., a template with textual and graphical information) for each of the obtained clusters.

Results: From the 538 respondents (153 P, 175 N, 176 OMP, 344 TA), the highest stress level, workload impact and risk of mental disorders were found in the N subgroup. Two clusters were identified for P, three clusters for N, two for OMP and one for TA.

Conclusions: The proposed framework was able to stratify different risk levels of possible development of mental health issues in healthcare workers due to COVID-19. This approach could represent the first step towards the development of mobile health tools to tailor psychological interventions in pandemic situations.

Keywords: E-health; Personas; Burnout syndrome; COVID-19.

1. Introduction

Since the beginning of the COVID-19 pandemic, healthcare workers worldwide have been under heavy work-related conditions that may negatively impact their psychological wellbeing. The rapid and unexpected virus spread, the high risk of contagion, the need of reorganizing their working activity and the huge increase in workload are just some of the significant variables that have contributed to the onset of moderate to severe psychological disorders, including stress, anxiety and depression in physicians, nurses and other healthcare providers already in the immediate wake of the viral pandemic [1]–[6]. [1]

Besides these contextual and organizational factors, different studies have also highlighted the role of specific sociodemographic and psychological characteristics as predisposing factors for the early onset of distress and emotional burden in these specific categories of workers. In particular, it has been noticed that being a female nurse, having fewer years of working experience, adopting maladaptive coping strategies and having a high fear of being infected are all factors that increased the risk of developing mental disorders during the initial phases of the pandemic spread [6], [7].

As largely discussed in previous literature, prolonged distress and related psychological symptoms can affect cognitive and technical performance of workers [8] other than triggering pre-existing mental health disturbances [9] or resulting in severe psychological illnesses, such as post-traumatic stress disorder (PTSD) and burnout [10], [11]. Starting from these findings, it is absolutely important not only to develop early psychological interventions which include psychological assessment, support, and services for healthcare workers within the health emergencies [2], [12]–[14], but also to customize such interventions as a priority, in order to meet the different needs of the different users' categories [2], [15].

A possible modern approach towards the customization of supporting interventions is based on the creation of Personas, where a Persona represents the generic participant in a specific cluster, and it is able to represent hypothetical archetypes of the actual users in that cluster. [16] Personas are defined through their "goals", namely their main needs and requirements, and are developed from individual data directly collected from real users,

---

[1] **Abbreviations**: IES: Impact of Event-Scale, MBI: Maslach Burnout Inventory, N: Nurses, OMP: Other Medical Professionals, P: Physicians, PAM: Partitioning Around Medoids, PCA: Principal Component Analysis, PHQ-4: Patient Health Questionnaire-4, PTSD: Post-Traumatic Stress Disorder, TA: Technical Administrative.

properly analysed in order to group them, forming clusters of subjects that share similar characteristics and represent the same archetype of user [17].

Originally created for marketing purposes, applications of Personas in the healthcare settings are starting to be explored and created using several methods, having the potential to be a useful tool for designing empowering personalized digital health solutions [17]–[19]. In fact, as one-to-one personalization in the context of patient-centered approach is practically impossible using digital health solutions, referring to the Persona as representative of patients with the same main characteristics (according to the goal for which the Persona was created) allows a one-to-N customization (i.e., different engagement design, different level of medical attention relevant to the corresponding risk stratification, etc.), focusing on the common features within each cluster. The intrinsic nature of Personas requires high interpretability of the underlying relationship among input data to develop realistic and usable representation of archetypes of real users. This suggests the utilization of methods for their development where a clear understanding of the statistical relationship among the variables of interest is preserved. In this perspective where the focus is not on person-centric estimates, but on generic group level characteristics, specific approaches need to be explored.

Accordingly, the aim of this paper was to propose a novel framework for the creation of Personas, applied to results of an on-line survey dedicated to healthcare professionals working during COVID-19, as the first step for designing a digital solution towards personalized assessment and prevention of mental health conditions. The proposed framework includes a specific quantitative data processing strategy to compute relevant features and define those variables able to characterize different Personas in the context of risk stratification.

1.1. Background

Considering the current literature on Persona development in the healthcare domain [17-25], it is possible to notice a lack of a “gold standard” method in the creation of Personas, which is also reflected by the variety in the target population, in the data collection protocols, in the persona creation methods and in the key variables utilized.

Regarding the target population, previous studies on Persona development in the healthcare domain were mainly focused on patients [17]–[22], or on a wider audience including journalists, researchers, caregivers and others [23]. Only one study was focused on healthcare workers, and more specifically intensive care unit nurses, but with a very specific usability goal to investigate their preferences for patient monitoring display prototypes [24]. In this perspective, a gap is thus present not addressing Persona development for healthcare professionals considered as potential patients.

Data collection in previous studies was performed using different strategies, including both quantitative (surveys) and qualitative (focus groups, semi-structured interviews) information, while also combining it from different sources (surveys, health records, data log) [17]–[25]. These approaches, hence valid, are complex, costly and time-expensive, with qualitative data requiring specific interpretation, thus highlighting the need for a more straightforward and quantitative data collection strategy.

These differences are also reflected in the methods used for Personas creation, ranging from more qualitative approaches either through open coding [17] or use of pro-forma [20], to more quantitative and precise algorithms such as hierarchical clustering [18], K-means clustering [19] or K-medoids clustering with Partitioning Around Medoids (PAM) algorithm based upon Gower distances [22]. Interestingly, in all previous studies the problem of dimensionality that often comes up in large datasets was not addressed, thus limiting the generalizability of results generated from high dimensional data to the overall population.

Finally, the key variables of interest also varied from study to study [17-25], changing the goal of the developed Personas and their context of usage, with demographic variables as the most commonly included, while psychological variables were utilized only in few studies [18,19,22]. Moreover, Personas were never created with the goal of addressing healthcare workers' mental health, in particular during a pandemic event.

2. Methods

2.1. Framework Definition

The list of the steps constituting the proposed framework for Personas creation is presented in the following table, considering that, according to its goals, each step needs to be adapted to the specific application:

136 **Table 1.** Steps in the proposed framework for Personas creation, with the associated
137 general and application-specific descriptions.

Step	General Description	Specific implementation
Survey definition	The expected goals of the Personas will need to be defined, together with the associated questions and relevant additional information	The goal corresponded to the mental health of the individual, represented by psychological indexes
Data collection	Choose the best modality according to the type and quantity of data that would need to be collected, the speed of data collection (and the time variant phenomena which could modify the results), the desired level of realism of obtained Personas	Single web-survey to increase the speed and the amount of collected data, at expenses of the realism of the Personas. Including semi-structured interviews conducted on a small batch of respondents could have been used to collect also qualitative data.
Data pre-processing	Perform data transformation (i.e., one-hot encoding) to encode nominal variables, and then apply the most proper dimensionality reduction method (i.e., Principal Component Analysis (PCA), Factor analysis of mixed data (FAMD), Multiple factor analysis (MFA), Multiple correspondence analysis (MCA), Categorical Principal Components Analysis (CATPCA)) according to the mix of observed variables, to select a number of features to reduce dataset dimensionality, and to enhance clustering results in the next step.	This represents a specific novelty proposed in our application. In our implementation, the number of features resulting from the PCA was chosen as cumulatively explaining at least 75% of the total variance.
Data clustering	Define the optimal number of clusters to be obtained and perform clustering on the PCA features using the k-medoids method applying the most proper algorithm based on data numerosity (Partitioning Around Medoids – PAM or Clustering LARge Applications - CLARA [26])	Evaluation of both the sum of within-cluster distances and the average silhouette value heuristics (plus input of the domain expert in case of uncertainty) was used to define the optimal number of clusters for each professional

		group, followed by PAM.
Statistical analysis	For each variable, define the proper statistical test and apply it to test null hypothesis of no difference among clusters. Variables for which null hypothesis is discarded represent specific characteristics that define the Personas, to be highlighted in Personas description.	Comparisons were performed separately among each professional group.
Personification	In defining the Persona cards, a graphical template is designed based on the goals set and results of statistical analysis	Results in a form of traffic light-based colored bars and related values were implemented, together with textual description. Availability of semi-structured interviews and qualitative data would have allowed to increase empathy and realism [27], [28].

The proposed framework has been inspired by the 10-step one proposed by Holden et al. [18], combined with further adaptation to the specific context of application. In the following, the implementation of each step will be described in detail.

2.2. Survey definition and data collection

Data have been collected by means of different questions, including validated psychological questionnaires, sociodemographic and working-related items, selected in collaboration with a team of domain experts in psychology at ICS Maugeri, Pavia and IRCCS Centro Cardiologico Monzino, Milan. These questions were disseminated by means of the online Qualtrics® platform to the healthcare workers of these institutions, localized in the Lombardy region, Italy, from the last week of April to the end of May 2020. This period corresponded to the end of the first wave of the pandemic, whose peak in Lombardy was registered on March 22, 2020 in terms of daily hospitalizations (1230) and on March 28, 2020 in terms of daily number of deaths (equal to 542) [29] followed by a lift of the mobility restrictions starting from May 18, 2020. [30]

The survey was composed of five different blocks, as shown in Table 3 and described in the following paragraphs.

Table 2. *Description of each block of questions composing the online survey, based on the focus of the information collected and the relevant number of questions.*

Block of questions	**Focus and number of questions *(n)***
Socio-demographic characteristics and current lifestyle	Common socio-demographic and current lifestyle: 9 questions.
Occupational: generic	Working characteristics of respondents: *5* questions.
Occupational: COVID-19 related	Work-related variables during the pandemic: *16* questions.
COVID-19 infection	Ascertained / Supposed positivity to COVID-19: 2 questions.
Psychological Indexes	Different psychological questionnaires, validated or not: • Impact of Event Scale – Revised (IES-R): *22* questions[a] • Patient Health Questionnaire (PHQ-4): *4* questions[a] • Maslach Burnout Inventory (MBI)- Emotional Exhaustion subscale: *5* questions[a] • Perceived COVID-19 fear for self / for family: *4* questions • Stress: *2* questions

[a] Questionnaire validated by scientific literature.

With the first block of questions, we collected information about age, gender and marital status of the respondent, as well as about the presence of close family members living in the same house (i.e. children and/or elderly people). The presence of chronic diseases and the implementation of protective strategies taken at home (i.e., use of personal protective equipment in the house, isolation in a separate room or in a different house) were also investigated.

The second block investigated working seniority, professional status, and specialization of the respondent.

The third block was focused to understand how much the respondent perceived the impact of COVID-19 on his/her working environment since the beginning of the pandemic. It also included 9 specific questions (on a 0-100 scale) whose answers were averaged to obtain the *workload impact* index.

With the fourth block, the respondent was asked if he/she was tested positive with the virus and, if not, if he/she thinks to have contracted it, even without having performed a swab test for confirmation.

In the final block, a psychological evaluation was conducted using both *ad hoc* and validated questionnaires. Four questions (on a 0 – 100 scale) were used to assess the perceived risk and probability for the respondent and/or his/her family members to contract the virus, and the relevant associated fear, respectively defining the "COVID-19 risk for self" (2 questions) and "COVID-19 risk for family" (2 questions) indexes. Then, two questions on stress perception and work-related personal satisfaction (on a 0 – 100 scale) were used to define a Stress index.

In order to evaluate the risk of developing burnout in the long term, as the Maslach Burnout Inventory is a lengthy questionnaire usually administered some months after the acute episode, only its exhaustion subscale was used, which is validated by the literature to be used separately [31]. The Impact of Event Scale – Revised *(IES-R)* [32] is a validated questionnaire used to assess the response to a traumatic event, also allowing the evaluation of the potential insurgence of Post-traumatic Stress Disorder (PTSD). Finally, the Patient Health Questionnaire – 4 *(PHQ-4)* [33], a validated tool to detect anxiety and depression [34], was administered.

The online survey was designed as a compromise between the entirety of the evaluation and the need to keep it concise as to lessen the impact on the free time of the health care personnel in order to complete it during the COVID-19 emergency, thus resulting in a total of 94 questions that required, on average, less than fifteen minutes to be concluded.

The study was approved by the Ethical Committees of the Istituti Clinici Scientifici Maugeri (approval number 2411, 26 March 2020) and IRCCS Centro Cardiologico Monzino (approval number 1238, 17 April 2020). The respondents gave their explicit electronic consent to data treatment and usage, in accordance with the rules defined by General Data Protection Regulation (GDPR), with obtained data anonymized by removing possible identifiable personal data such as the Internet Protocol (IP).

2.3. Data pre-processing

All the data analyses were performed using the MATLAB® software (The MathLab, Natick, MN, USA) with its Statistics and Machine Learning Toolbox, and the R language (The R Foundation, Vienna, Austria).

Records corresponding to uncompleted submitted surveys (i.e. with less than the 98% of the required items filled in) were removed. Answers resulting from the selection of option “Other” were removed due to their low information content. Empty fields deriving from logical branches were converted into numerical values to be used in further analysis, while multi-answer questions (i.e, children age, in case of multiple children) were split into dummy binary variables.

The single scores obtained from the validated psychological questionnaires were summarized into total scores, as suggested by the corresponding validation studies [31]–[35]. Based on the respondent’s profession, the records from unlicensed assistive personnel, psychologists, physiotherapists, speech therapists and other medical categories were grouped together into “other medical professionals” category. In this way, respondents were divided in a total of four groups: physicians (P), nurses (N), other medical professionals (OMP), and technical administrative staff (TA). The following analysis aiming at the definition of Personas was then performed separately for these four groups.

2.4. Data analysis

At the end of the previous pre-processing step, the collected information included a total of 46 variables. To further reduce this number, methods of dimensionality reduction have to be applied, with the final choice varying depending on the characteristics of the collected dataset.

Principal Component Analysis (PCA) [36] could be used when the vast majority of the variables in the dataset are quantitative or ordinal in nature. Accordingly, all the nominal variables must be one-hot-encoded, to ensure that they are not treated as quantitative variables. When the dataset is entirely or mostly categorical, Multiple Correspondence Analysis (MCA) [37] can be used to perform dimensionality reduction. Finally, when the dataset includes both quantitative and nominal categorical variables, Factor Analysis of Mixed Data (FAMD) [38] could also be applied. FAMD performs a combination of PCA and MCA, using the former for quantitative and ordinal variables and the latter for nominal variables. Other available methods include Categorical PCA (CATPCA) [39] if the data is mostly categorical, or Multiple Factor Analysis (MFA) [40], for categorical or numerical features.

In this study PCA was applied, thus highlighting 7 features (i.e., as linear combinations of the original variables) cumulatively explaining at least 75% of the total variance [36] for P, N and OMP, and 5 features for TA. The variance threshold value was defined by a trial-and-error process by considering the results obtained through the next steps of the analysis.

In order to obtain clusters of records based on the resulting features from the PCA, K-medoids clustering was applied through the Partitioning Around Medoids (PAM) algorithm. [26] A characteristic of this method is the use of medoids (i.e., actual points in the dataset) as the center of mass for each cluster.

A requirement of the K-medoids clustering is that the number of clusters *K* must be decided *a priori*. As no standard rules to take such decision are available, heuristic methods need to be applied. In this study a combination of two heuristics was used to assess the optimal number of clusters: 1) the evaluation of the sum of within-cluster distances (i.e., the Euclidean square distance between each point of a cluster and its medoid) [41] for K in a range between 1 to 10:

$$tot\ within\ cluster\ dist = \sum_{k=1}^{k} W(C_k) = \sum_{k=1}^{k} \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

For this monotonically decreasing heuristic, the higher is this value, the more disperse are the points in each of the corresponding K clusters.

2) the average silhouette value S for K in a range between 2 to 10, defined as the mean of the silhouette value for each point x_i [42]:

$$S = \frac{1}{|C_i|} \sum_{x_i \in C_i} s(x_i) = \frac{1}{|C_i|} \sum_{x_i \in C_i} \frac{b(x_i) - a(x_i)}{\max\{a(x_i),\ b(x_i)\}}$$

with $b(x_i)$ defined as the minimum of the average of distances from point x_i to each point in all the clusters C_k except its own (i.e., C_i), and the C_k with the minimum average distance to x_i is defined as the neighboring cluster:

$$b(x_i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{x_j \in C_k} d(x_i, x_j)$$

while $a(x_i)$ is defined as the mean distance from point x_i to each point in its own cluster C_i.

$$a(x_i) = \frac{1}{|C_i| - 1} \sum_{x_j \in C_i,\ x_j \neq x_i} d(x_i, x_j)$$

This heuristic results in a value between -1 (sample very close to the neighboring cluster) and +1 (sample very far from the neighboring cluster). If $a(x_i)$ is smaller than $b(x_i)$ the silhouette $s(x_i)$ is closer to +1, meaning that the distance x_i from the neighboring cluster is larger than the one from its own cluster. On the other hand, if $a(x_i)$ is larger than $b(x_i)$ the silhouette $s(x_i)$ gets closer to -1, implying that point x_i is closer to the neighboring cluster than to the one to which it has been assigned. A value of 0 means that the point x_i is on the border between two clusters.

Using these heuristics and plotting the corresponding results as a function of K, the optimal number of clusters corresponds to an "elbow" or to a "peak", respectively for the former and the latter. In case two different K were found by the two heuristics, a final decision between the two was taken considering the input of the domain expert (i.e., the psychologists).

Once clustered, the resulting data were converted back into the original 46 variables to proceed with statistical analysis.

In Figure 1 a flowchart of the analysis process is presented, showing at each step the amount of variables or features used in the dataset. For the purpose of the shown example the physicians' dataset is used.

2.5. Statistical Analysis

For binary and nominal attributes, ratios or proportions, contingency tables with Fisher and Chi square test were applied, while for the other variables Mann-Whitney U [43], or Kruskal-Wallis [44] followed by multiple Mann-Whitney U tests between groups with Bonferroni correction, were applied respectively for K=2 or $K > 2$. For all tests, statistical significance was set to $p < 0.05$.

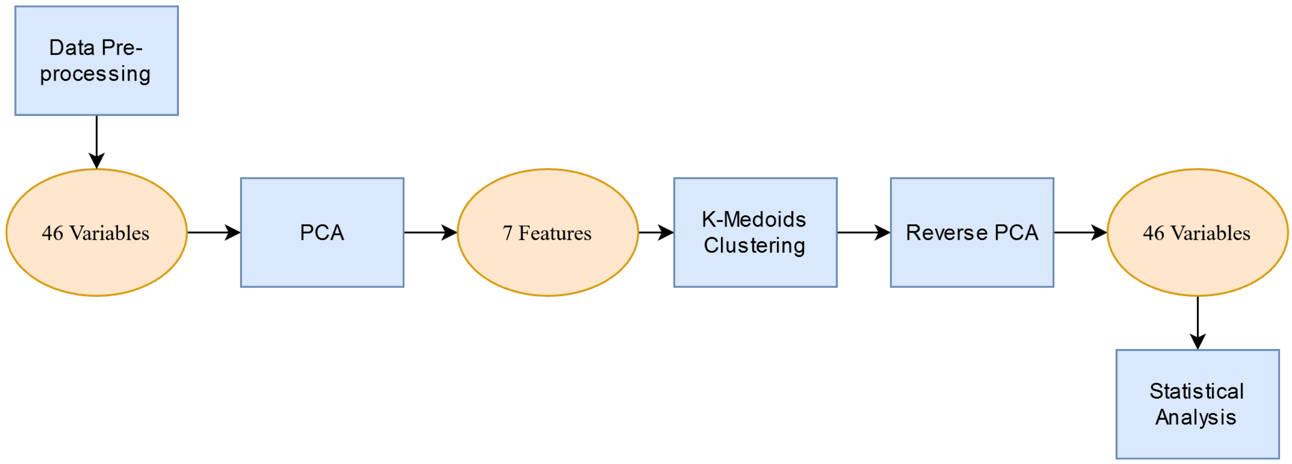

Figure 1. Flowchart representing the proposed data processing, applied as example to the physicians' dataset. The different processes are shown in blue boxes, while the number of resulting variables or features in the dataset is shown in orange ellipses.

As a first step, differences in the recorded 46 variables among the four professional groups were evaluated to highlight the possible impact of COVID-19 on the different healthcare categories. Afterwards, once the final clustering was performed by the PAM algorithm within each professional group, proper statistical analysis was applied to define which variables out of the original 46 were able to differentiate among the different clusters.

2.6. Personification

For each cluster within the corresponding professional group, a "persona card" was created. The "persona card" is a template filled with information associated to those specific attributes that makes the Persona easily accessible, while also providing a realistic representation of the end-user that such Persona is supposed to represent [45]. This template was created starting from those variables that differentiated the clusters, thus assigning a characteristic trait to the Persona based on the relevant median value for each attribute. In addition, a randomly chosen name and a non-existing face [46], together with an age randomly chosen in the 25^{th} – 75^{th} percentile range of the corresponding variable, were given to each Persona. Finally, as the focus of our analysis was on mental disorders eventually developed during the COVID-19 emergency, the main identifiable characteristics referring to each specific Persona were represented by the scores obtained in the different psychological questionnaires. To allow a fast interpretation and utilization, these indexes were then represented in a graphical form together with the Persona description. In particular, bar length and color were coded accordingly to the values referred to in Table 4, in which the scales were empirically stratified into three levels or according to validated cut-off values, as in the IES [32] and the PHQ-4 [47]. For each index, a green bar describes a safe range of values, a yellow bar highlights a range potentially dangerous for health, while a red bar identifies an extremely dangerous score.

Table 3. Risk scale based on the values of the psychological indexes, in which three levels have been defined and color-coded to be utilized in Persona cards.

Index Value	Workload Impact	Stress	MBI	IES	PHQ-4	Burnout Knowledge
Low (green)	0 - 34	0 - 34	0 - 8	0 - 23	0 - 3	4 - 5.9
Medium (yellow)	35 - 65	35 - 65	8.1 - 13.5	24 - 32	3.1 - 4.9	3 - 3.9, 6 +
High (red)	66 +	66 +	13.6 +	33+	5 +	< 3

[a] MBI = Maslach Burnout Inventory. IES = Impact of Event Scale-Revised. PHQ-4 = Patient Health Questionnaire-4.

3. Results

A total of 570 respondents started filling in the online survey between April 27th and May 31st 2020. Due to their uncompleted submitted surveys or missing privacy data usage consent, 32 respondents were removed, thus resulting in 538 completed surveys among which a prevalence of women *(n=361,* 67.1%*)*, with a median (25th; 75th percentile) age equal to 45 (37; 52) years was observed, and a remaining male component (*n=177,* 32.9%*)* with a median age of 45 (35; 55) years. Considering the distribution of the respondents by professions, 28.4% *(153)* were P, 32.6% *(175)* were N, 32.7% (176) were classified as OMP, and 6.3% (34) were TA staff.

In Table 5 the attributes that resulted statistically different between the four professional groups are reported: N group included more women than men compared to P and OMP, where P were older than N and OMP. In general, the N group was more afraid to be infected and more worried about the risk for their family members to be infected than P and OMP. The N group was also the one showing the highest perceived impact of COVID-19 on the workload. Accordingly, the stress level, the risk of burn-out (as reported by MBI) and PTSD (as reported by IES), as well as the risk of anxiety and depression (as reported by PHQ-4), resulted higher in the N than in the P and the OMP groups.

Table 4. Subset of variables (out of the original 46) showing statistical significant differences between the four professional groups, reported as median (25^{th};75^{th}) for continuous variables, % for binary variables, and mode for nominal variables.

	Physicians *(n =153)*	**Nurses** *(n = 175)*	**Other Medical** *(n = 176)*	**Tech-Admin** *(n = 34)*	***P* value**
Sex	65M 88F	36M 139F *	62M 114F #	14M 20F	< 0.001
Age	48 (40.75; 58)	45 (34; 50) *	43 (32.5; 53) *	45.5 (35; 51)	< 0.001
Lives With	spouse + children (46%)	spouse (45%) *	spouse (43%)	spouse (50%)#	< 0.001
COVID-19 fear for family	50 (2.5; 67)	65 (10; 83)	55 (11; 75)	50 (0; 75)	0.046
COVID-19 fear for self	60 (49; 75)	70 (50; 80) *	60 (50; 75) #	65 (50; 83)	0.010
Ward	other (39%)	other (37%)	other (40%)	other (85%) * # &	0.002
Does shifts	yes (54%)	yes (79%)	no (72%) #	no (74%) * #	< 0.001
Workload impact	58 (47; 67)	65 (53; 77) *	54 (41; 70) #	55 (34; 64) #	< 0.001
Stress	60 (50; 71.5)	70 (55; 84) *	60 (49; 74.5) #	62.5 (51; 76)	< 0.001
MBI	8 (6; 13)	12 (8; 18) *	8 (5; 12) #	6 (3; 9) #	< 0.001
IES	18 (10; 30)	28 (17; 43) *	20 (10; 33.5) #	23 (16; 33)	< 0.001
PHQ-4	3 (1; 5)	4 (2; 7) *	3 (2; 5) #	3 (2; 6)	< 0.001

[a] MBI, Maslach Burnout Inventory; IES, Impact of Event Scale – Revised; PHQ-4, Patient Health Questionnaire-4.

*: p<.05 vs Physicians; #: p<.05 vs Nurses; &: p<.05 Other medical vs Tech Admin

Considering the results of the PCA analysis conducted separately for each professional group, Figure 2 shows the percentage weights attributed to the questions for each of the blocks described in Table 3: for all professions, the “Occupational: COVID-19 related” questions were the ones that resulted in the highest combined weight, followed by the questions relevant to the psychological indexes. Lifestyle questions had a low impact when compared to the two previous categories. For all professions the “Occupational: generic” and the “COVID-19 infection impact” questions were the ones with the lowest impact overall (<1% of the total). An almost identical pattern was found for all professional groups.

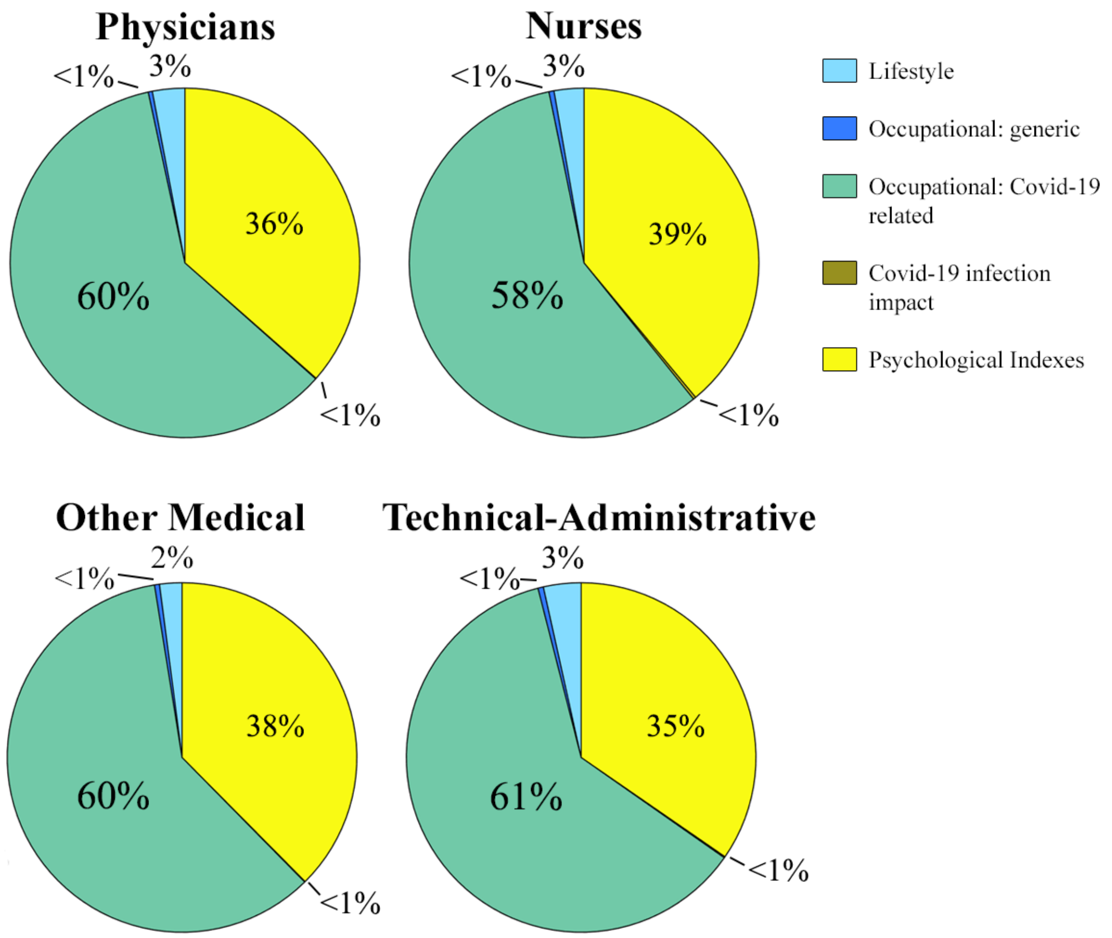

Figure 2. Percentage weight of the original 46 variables, grouped by the question blocks as defined in Table 3, in the resulting components explaining >=75% of the total variance from PCA analysis, applied separately for each professional group (see text for more details).

3.1. Data clustering of physicians' responses

The age and gender distribution of the 153 surveys originated from the physicians showed 65 men (42.5%) of median age 53 (40.75; 59) years and 88 women (57.5%) of median age 46.5 (40.5; 56) years. The optimal number of clusters, based on the previously defined rules, was identified as *K*=2 (see Supplementary Material 1). Consequently, the physicians' surveys were subdivided into two clusters of 66 and 87 respondents, respectively.

Gender distribution was not different between these two clusters (chi-square statistic, p-value = 0.328), as well as age distribution: cluster 1 median age resulted in 49 *(41; 59)* years while cluster 2 was 46 *(40; 57)* years (p=0.642). Cluster 1 was composed of 31 men (47.7% of the total 65 men physicians) and 35 women (39.8% of the total 88 female physicians). Cluster 2 included 34 men (52.3% of the total) and 53 women (60.2% of the total). Physicians in cluster 1 suffer from chronic pathologies and usually live alone, with no need to adopt protective measures at home. In cluster 2, they are less prone to suffer from chronic pathologies and live with their spouse and children using personal protective equipment at home; however, they are afraid of the possibility for their family members to be infected, and they are also more afraid than those in cluster 1 to get sick themselves. The pandemic had a lower impact on the physicians' workload (mainly not working in shifts) of cluster 1, than in cluster 2, where workload was highly impacted (usually working in shifts). Psychological indexes in cluster 1 showed lower risk of developing burnout *(MBI =7),* PTSD, *(IES = 12.5),* anxiety and depression *(PHQ-4 = 2),* while cluster 2 showed a higher risk (but still moderately low) of developing burnout *(MBI = 9),* PTSD *(IES = 21)*, anxiety and depression *(PHQ-4 = 3).* The corresponding table reporting all the statistically different attributes and related distributions can be found in Supplementary Materials 2

3.2. Data clustering of nurses' responses

The gender and age distribution of the 175 surveys completed by the nurses showed 20.6% (36) of men with a median age of 39 (32.5; 45.5) years and 57.5% (88) of women with a median age of 45 (34; 50.75) years. In this case, the previously described heuristics gave slightly discordant results (*K=3* for the total sum of within-cluster distances*, and K=2 for* the average silhouette). However, as the values of average silhouette for *K*=2 (0.312) and for *K*=3 (0.299) were similar, by evaluation of domain experts the decision to consider three clusters of 67, 38 and 70 respondents respectively was taken (see Supplementary Material 1). The gender and age distributions were not different among the three clusters (chi-square statistic, sex: p = 0.582; age: p = 0.074): cluster 1 was composed by 12 males (33.3% of the total 36 men nurses N) and 55 females (39.6% of the 139 female nurses) with a median age of 46 (37.5; 52) years; cluster 2 was the less numerous with 10 males (27.8%) and 28 females (20.1%) with a median age of 40 (30; 47) years, while cluster 3 included 14 males (38.9%) and 60 females (40.3%) with a median age of 44.5 (34; 50) years.

In cluster 1 and 3, nurses were married, had 2 children and used personal protective equipment at home, while in cluster 2 nurses were engaged and lived alone with no children, and consequently did not have the need to use protective measures at home. Cluster 1 was characterized by the highest fear among the three groups for the possibility that both nurse and his/her family members could become infected, with the highest probability (76%) of having COVID-19 cases in the ward, and the highest impact of the pandemic on workload index (77). Nurses in cluster 2 had an intermediate impact on the workload index (69), and also a high probability of having COVID-19 cases in their ward (76%). Conversely, nurses in cluster 3 had the lowest impact on the workload index (51), and the lowest probability (60%) of having COVID-19 cases in the ward. Psychological indexes in cluster 1 show the highest risk of developing both burnout *(MBI = 16)* and PTSD *(IES = 38),* and higher scores for anxiety and depression (*PHQ-4 = 5),* compared to the other two clusters. In cluster 2, these indexes are still high, with a medium risk of developing burnout *(MBI = 12)*, while also being highly susceptible to develop PTSD *(IES = 38)*, anxiety and depression (*PHQ-4 = 4)*, but less than in cluster 1. Finally, in cluster 3 there is a lower risk of developing burnout *(MBI = 8),* PTSD *(IES = 20)*, anxiety and depression *(PHQ-4 = 3)* when compared to cluster 1. The corresponding table reporting all the statistically different attributes and related distributions can be found in Supplementary Materials 3.

3.3. Data clustering of other medical professionals' responses

The age and gender distribution of the 176 surveys originated from the OMP showed 35.2% (62) of men with a median age of 45.5 (31; 53) and 64.8% (114) of women with a median age of 43 (33; 51). The optimal number of clusters was identified as K=2 (see Supplementary Material 1). Consequently, the OMP' surveys were subdivided into two clusters of 109 and 67 respondents respectively. Gender distribution was not different between the two clusters (chi-square statistic, p-value = 0.398), as well as age distribution, with cluster 1 composed by 41 males (66.1% of the total number of OMP) and 68 females (59.6%), with a median age of 44 *(31; 53)* years, and cluster 2 composed of 21 males (33.9%) and 46 females (40.4%) with median age of 43 *(33; 51)* years (p=0.576).

In cluster 1, OMP live with their spouse and have no children, while in cluster 2 they live with their spouse and one child, and consequently, they were more afraid for themselves and their family members to become infected, with a consequent larger use of personal protective equipment at home compared to cluster 1. The OMP in cluster 2 had their workload index more impacted (75) by the pandemic than in cluster 1 (45), with only 37% of professionals with work shifts. Psychological indexes in cluster 1 show a lower risk of developing burnout *(MBI = 6),* PTSD (*IES = 16)*, anxiety and depression *(PHQ-4 = 2)* compared to cluster 2 (MBI = 11; IES = 29; PHQ-4 = 4). The corresponding table reporting all the statistically different attributes and related distributions can be found in Supplementary Materials 4.

3.4. Data clustering of Technical Administrative staff's responses

The gender and age distribution of the 34 surveys originated from the TA staff showed 41.2% (14) of men with a median age of 45.5 (40; 49), and 58.8% (20) of women with a median age of 43.5 (32; 51). As the number of subjects in this group was extremely low, further division of the respondents would create extremely small clusters with weak validity. Consequently, the TA group of respondents was kept as a single cluster. Relevant attributes of this group can be found in Table 5, where they are compared to those of the other healthcare professionals.

3.5. Persona cards

Following the personification process, as no significant difference was found between male and females among the subgroups of the healthcare workers, a male/female Persona card was created for each cluster and profession (2 for P, 3 for N, 2 for OMP, and 1 for TA), including two names and photos of the opposite sex, sharing age and background defined as described in the Methods section, with the scores of the psychological indexes translated into colored bars to allow immediate visual identification of the associated level or risk.

Figure 3 shows the Persona cards for the physicians resulting from clusters 1 and 2, respectively. The top one (cluster 1) shows a lower combined risk profile, with workload and stress in the medium range, and MBI, IES and PHQ-4 in the low range. The bottom one is similar, except for the level of stress and the workload impact closer to high, as well as higher values for MBI, IES and PHQ-4. In the description of cluster 1, hypertension was chosen to report the presence of a chronic illness in this cluster, as it constitutes one of the most common chronic pathologies worldwide [48].

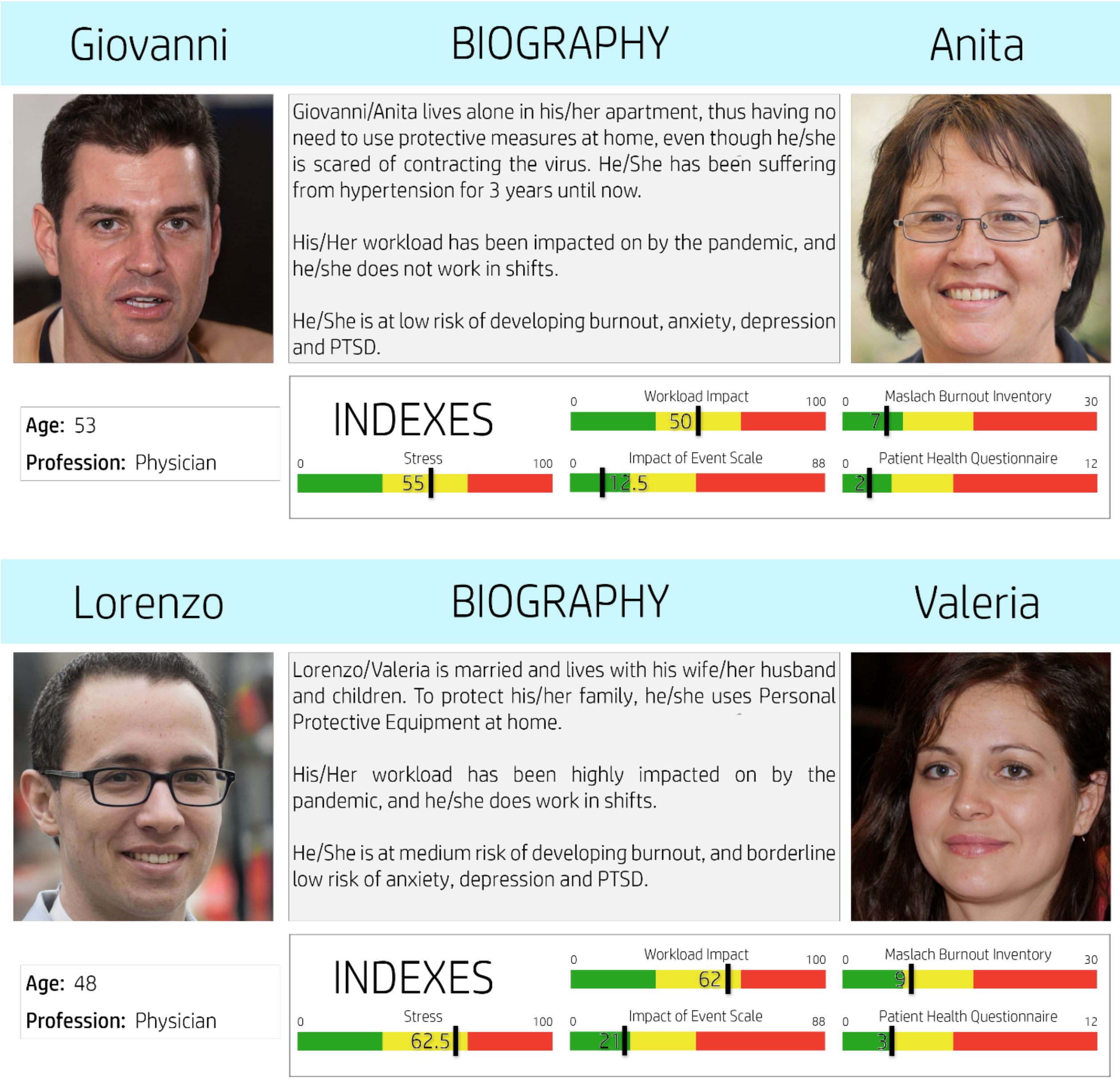

Figure 3. Persona cards resulting from clustering applied to the physicians' group. The first one (Giovanni and Anita) represents cluster 1, while the second (Lorenzo and Valeria) represents cluster 2.

In Figure 4 the Persona cards for the nurses group resulting from the three obtained clusters are shown. The top one (cluster 1) shows the highest combined risk profile (also among all the 8 Personas), characterized by high workload and stress levels, and high scores for MBI, IES and PHQ-4. The second one (cluster 2) is still associated to high workload and stress levels, but with the MBI, IES and PHQ-4 scores in the middle range. The last one (cluster 3) has a profile characterized with workload and stress in the medium range, and the MBI, IES and PHQ-4 scores in the upper level of the low range.

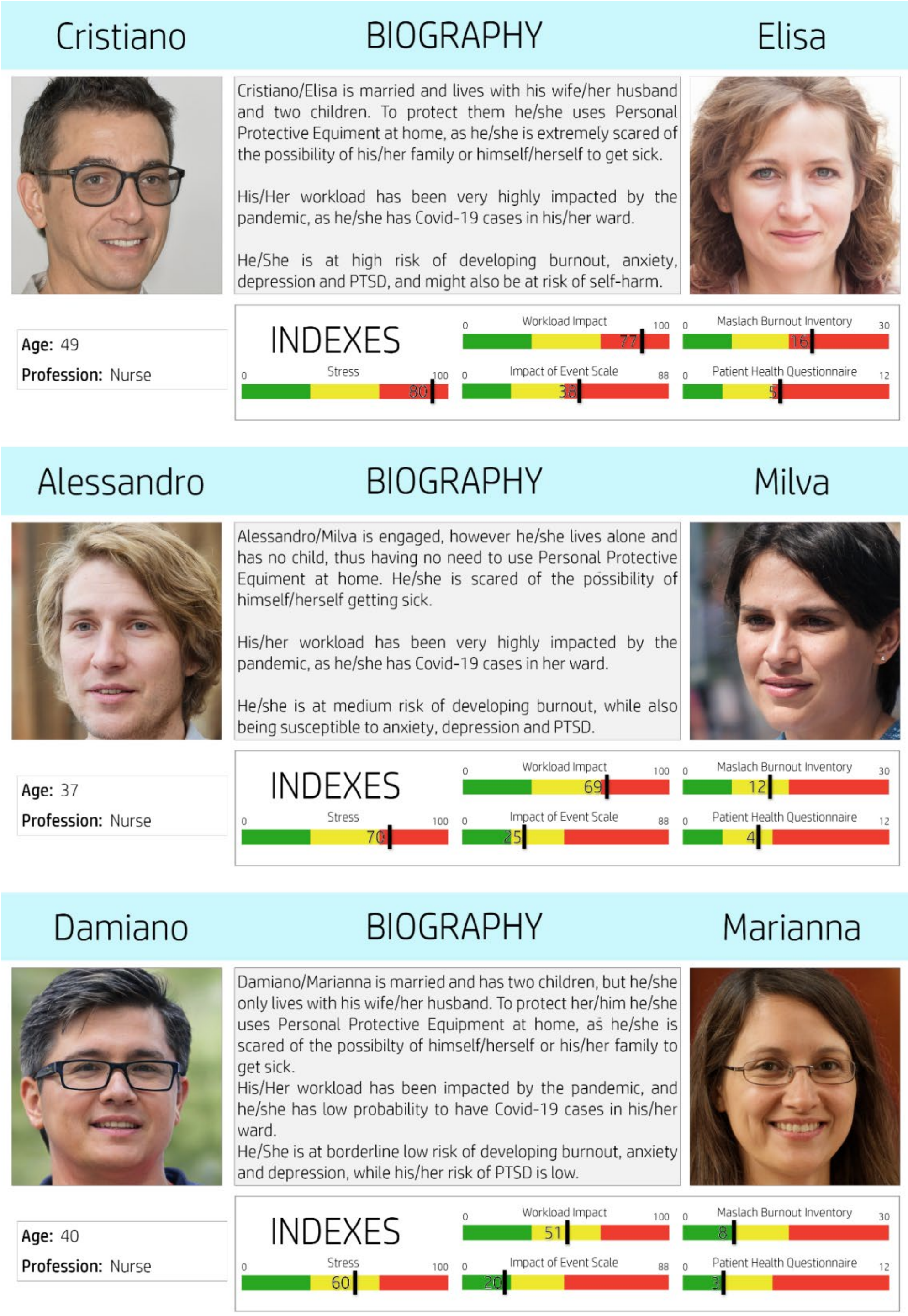

Figure 4. Persona cards resulting from clustering applied to the nurses' group. The top one (Cristiano and Elisa) represents cluster 1, characterized by the highest risk; the middle one (Alessandro and Milva) represents cluster 2, while the lower one (Damiano and Marianna) represents cluster 3.

Figure 5 shows the Persona cards for the two clusters obtained from the group of the OMP. The first one (cluster 1) shows workload and stress in the medium range, and the MBI, IES and PHQ-4 scores in the low range. The second one (cluster 2) has a combined higher risk profile, with workload and stress in the high range values, as well as the MBI, IES and PHQ-4 scores in the middle range.

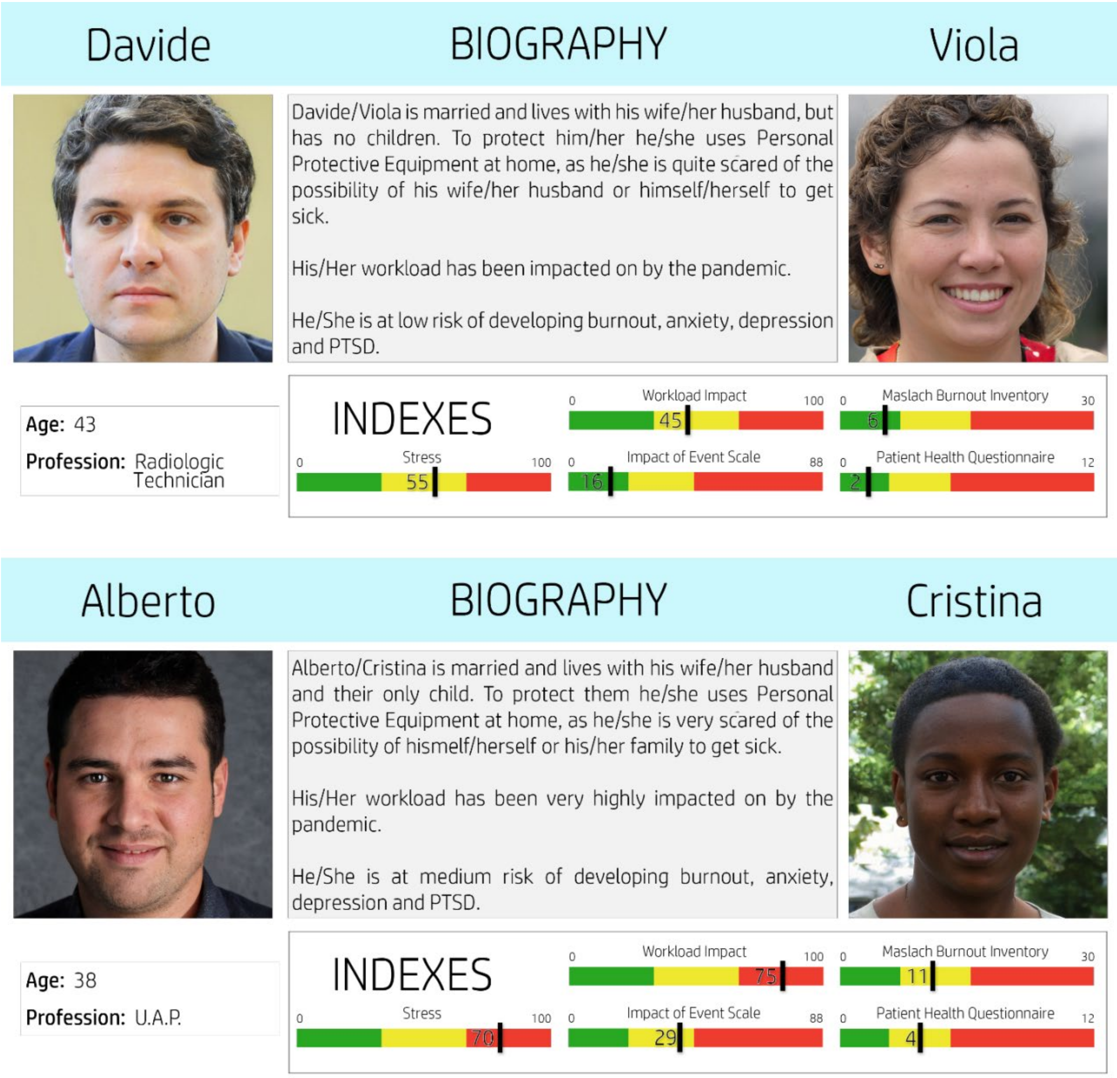

Figure 5. Persona cards resulting from clustering applied to the other medical professionals' group, with the upper one (Davide and Viola) representing cluster 1, and the lower one (Alberto and Cristina) representing cluster 2.

In Figure 6 the Persona card for technical-administrative group is shown. It is characterized by workload and stress in the medium scale, and the MBI, IES and PHQ-4 scores in the low range, showing a low risk of developing burnout, anxiety, depression, and PTSD.

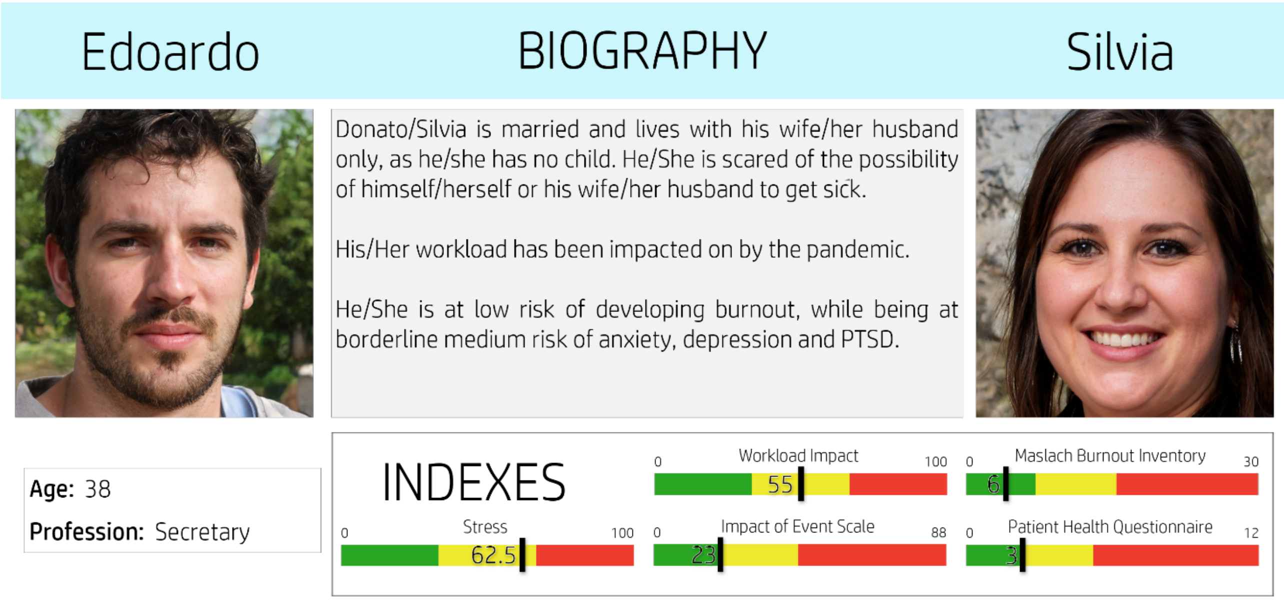

Figure 6. Persona card representing the technical-administrative group (Silvia and Edoardo).

4. Discussion

In this study, a novel framework was proposed and applied to create Personas for different categories of healthcare workers, with the purpose to perform risk stratification relevant to the development of mental disorders induced by a sudden stressful condition such as that induced by the COVID-19 pandemic.

From a methodological point of view, the presented framework presents four main novelties when compared to other studies in the field of developing Personas for healthcare:

1) it only makes use of an online survey to gather data, thus greatly reducing the time and money requirements to collect the needed information that, being only quantitative in its nature, makes it also easier to perform the presented analysis;

2) this is the first time that dimensionality reduction methods, hence not novel, were proposed and applied to reduce the complexity of input data, thus enhancing the performance of k-medoids clustering using the PAM algorithm;

3) a combination of the methods of average silhouette and total within sum of square distances were used to define the optimal number of clusters (and thus Personas) to be obtained;

4) color-coded bars in Persona cards were used to represent the psychological indexes and their potential level of risk, allowing immediate visualization and faster interpretation of the characteristics of each Persona.

Different from the previous methodological approaches that used both quantitative (surveys) and qualitative (focus groups, semi-structured interviews) variables to create Personas, while also combining information from different sources (surveys, health records, data logs), our innovative approach was based on data collection using a self-administered online survey, including questions about sociodemographic characteristics, lifestyle habits, occupational condition, and the impact of COVID-19 on personal feelings and the psychological status of the responders. To our knowledge, this is the first time that a similar approach has been applied for the Personas' creation. It has the benefit to reach a more widespread gathering of data among the target population, resulting in a larger number of respondents in a short time and reduced costs, while avoiding direct contact between the interviewer and the respondents. This aspect makes it applicable in contexts of high risk of contagion, with ubiquitous and time uncorrelated possibility to complete the online survey. Moreover, the total number of questions included was reduced to minimally impact the respondents' professional and personal obligations, with the quantitative data collected allowing a faster implementation of the methods for the subsequent analysis.

Concerning the target population, our study is the first that is specifically focused on different healthcare professionals and their mental health condition for Persona creation. Regarding the observed variables, also in previous studies some psychological indexes were included [18], [22], but only focusing on specific samples of patients.

Another important methodological improvement in respect to previous studies consisted in the application of dimensionality reduction methods to the original variables to reduce the dataset dimensionality to a range between 5 and 7 features, thus simplifying the following clustering operation. Among the available methods of dimensionality reduction, PCA was chosen as it resulted in higher average silhouette values when compared to the other methods in all the four professional groups. Performing PCA corresponds to a primitive form of noise reduction [53] lowering the weight of the variables with a lower variance in the dataset, and thus giving them less importance when performing the clustering operation. Similarly to what applied in a recent study [25], k-medoids clustering was used: this approach has been shown to generate better performance at the cost of higher complexity when compared to k-means clustering [54]. The applied PAM algorithm, despite requiring some computational effort [55], did not require more than a few seconds for analysis, thus showing its applicability for the number of variables and respondents in the considered task, making full usage of the strengths of dimensionality reduction techniques.

Importantly, once clusters were defined on the principal components, the following statistical analysis performed on the original 46 variables among the created clusters allowed to highlight those minimal sets of variables able to discriminate among the obtained clusters for each healthcare professional category. The appropriateness of this method is indirectly confirmed by the fact that the obtained results show that among the four analyzed health professional groups, the nurses were the ones characterized by the highest risk of developing mental health issues relevant to the COVID-19 pandemic, in agreement with previous studies [48]–[50]. In general, the obtained Personas were deemed by the psychologist experts in our team, which have worked at close contact together with health workers, as appropriate and coherent with the existing literature on mental health issues deriving from the pandemic event [6].

To define the optimal number of clusters a combination of the average silhouette and total within sum of square distances were used, together with help from domain experts in case of a tie between the two methods. To our knowledge, this is the first time that both methods were used in deciding the optimal number of clusters to develop Personas in the field of healthcare. In this way, a total of eight clusters (two for P, three for N, two for OMP, and 1 for TA) were created, corresponding to different levels of risk of developing burnout, anxiety, depression and PTSD in response to the first wave of the COVID-19 pandemic in Italy. Of the four identified professional groups, the nurses included one cluster associated with the highest overall risk of developing mental health issues, and the created Personas (Elisa/Cristiano, Milva/Alessandro and Marianna/Damiano) were shown different reactions to the pandemic event associated to distinct risk levels, and to the perceived impact of workload and family situation. In fact, a higher score in the workload impact, associated to the presence of COVID-19 patients in the ward, and to the fact of living with other family members, corresponded to higher values in the psychological indexes (MBI, PHQ-4, IES). The fact of living alone (i.e. Milva/Alessandro's Persona) contributed to lower risk levels, despite high values of workload impact. Not having direct contact with COVID-19 patients (i.e., Marianna/Damiano's Persona) generated lower impact on workload and stress.

The use of psychometric tools (i.e. questionnaires related to mental health) as well as the proposed graphical representations of the most important indexes in the Persona cards as colored bars with related values, allows a user-friendly and easy identification of the relevant characteristics [45] and different risk profiles for immediate understanding of healthcare professionals, with the advantage of potentially bringing them closer to the design process, as well as supporting designers in a better comprehension of the medical-related problem.

As regards to the generalizability of the proposed framework, it could also be applied to different goals for Persona's creation in the context of healthcare, where an on-line survey could be opportunely produced and disseminated to reach potential target users (i.e., patients with a specific chronic disease within hospital reach). Based on the collected data, proper methods for dimensionality reduction and relevant statistics could be applied, to determine the corresponding Personas descriptions in accordance to the defined goals.

Our current findings relevant to the risk of developing mental health issues following the COVID-19 pandemic could be generalizable to other situations of healthcare workers operating in the context of scenarios with high risk of contracting other communicable diseases, such as Ebola or Severe Acute Respiratory Syndrome Coronavirus 1 (SARS-CoV-1) outbreaks.

Compared to other methods such as machine learning, the proposed approach can be suitable when a gold standard label is apriori missing or not assignable from the examined subjects, thus preventing supervised machine learning methods to be applied to solve classification problems. As well, compared to unsupervised learning approaches, Personas do not represent a prediction, which is the main goal of these machine learning algorithms [51], but the description of the main characteristics of clusters of subjects, with more transparency on how they are computed and higher explicability of the results. In fact, in developing Personas, all the collected data are used to identify the main characteristics of the analyzed population, without any distinction between training and testing data. In machine learning algorithms the accuracy of prediction prevails over the interpretability of the statistical relationship found in the training data; on the other hand, in the development and characterization of Personas, the understanding of the underlying relationship between attributes is key to achieve realistic and useful results. In this view, the proposed framework could be applied to other contexts in which these factors could be relevant.

4.1 Clinical implications

The current COVID-19 pandemic in its first development phases has shown that healthcare workers, nurses and physicians in particular, were significantly exposed to increased workloads, stress, and the lack of protective personal measures. All these factors could increase the risk of developing short- and long-term mental health problems as a consequence of physical and mental distress experienced during the emergency [6], [14], [52], together with a lack of opportunities for psychological assessment and support. On the other end, when this support is available, it was often not easily accessible as it referred to a specific time and place, thus interfering both with professional obligations and personal life [53]–[55].

In this perspective, the possibility to have a mobile health application capable of providing both the monitoring of healthcare workers' mental health status and direct access to a tailored ubiquitous support, adapted to the user's personal and situational characteristics, could represent a useful solution for healthcare workers during long-lasting emergency situations.

The proposed methodology represents the first necessary step to reach this aim, by which Personas characterized by different risks of developing mental health issues, for each healthcare profession (i.e., P, N, OMP and TA), were created. The potential usage of such Personas could be twofold:

1) following further analysis based on feature selection protocols to better elucidate which attributes are more capable to differentiate for the risk of developing mental health issues among the clusters of a certain professional group, the survey size could be reduced to a minimal set of questions to be subministered by a psychologist to achieve a fast assessment of their risk level during a pandemic event, based on which further attention could be dedicated to those subjects with the highest risk factors.

2) the possibility of providing the healthcare professional (in particular for nurses or physicians) with a self-monitoring tool capable to provide the new respondent with the corresponding Persona could increase his/her awareness about the possible risk situation and trigger the need to search for psychological assessment and support.

4.2. Limitations

The majority of the psychological questionnaires included in the web survey were validated by literature. However in some cases, to reduce the number of questions (such as the 2-item stress scale introduced instead then the validated 10-item Perceived Stress Scale [56] or to explore ad hoc aspects relevant to the pandemic scenario (i.e., the 9-item Workload Impact index), not validated questionnaires were used.

The utilized dataset had an uneven distribution both in gender and profession. In particular, females were two thirds of the whole dataset, thus potentially skewing the obtained results. Furthermore, the TA group had very few respondents that prevented performing clustering on it.

If from one side the choice of using exclusively quantitative data facilitated data collection and clustering, the absence of qualitative data deriving from semi-structured interviews and focus groups may limit the realism of the obtained Personas. Without qualitative data and techniques such as Empathy Maps, it is not possible to add quotes to Persona cards or fit more nuanced information into their narrative to increase the empathy felt by designers in their usage. [56] The risk of this approach could be that the obtained Personas would result as a caricature and unrealistic, which increases their engagement at the beginning but lowers their effectiveness over time. [28]

An additional limitation of this study is that, apart from the general approval of meaningful Personas obtained by domain experts in our team, a more in-depth validation was not performed as part of this study. However, we are currently investigating this aspect with a longitudinal follow up in a subgroup of respondents who gave their written consent during the previous online survey, by evaluating the effective insurgence of mental health issues one year later and correlating results with the previously assigned Personas.

Finally, the possible applicability of our Personas to different international contexts, as well as to other emergencies different from epidemics or pandemics, was beyond the scope of our work. Further studies are needed to evaluate results for cross-cultural international Personas [27].

4.3. Conclusions

The proposed framework for Personas creation was applied to the problem of risk stratification of development of mental health issues in healthcare workers in Italy due to the COVID-19 pandemic. From the analysis of quantitative data obtained through an online survey, after opportunely dimensionality reduction followed by k-medoids clustering, several clusters representing Personas with a different associated risk within each health professional group were created, and described using Persona cards, in which also colored bars and related values were used. This graphical representation has the potential to bring healthcare professionals closer to the design process and supports designers to understand better the medical-related part of the solution they will design, as a first step for interdisciplinary cooperation.

This approach constitutes the first step towards the development of mobile health tools capable of providing both monitoring of the current mental health status and access to psychological support customized to the user, representing a possible solution to allow ubiquitous assistance at any time, also avoiding face-to-face interviews, to the healthcare workers in emergency situations, such as epidemic or pandemic events.

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The PhD fellowship of ET is supported by Istituto Nazionale Previdenza Sociale (INPS), Italy.

Bibliography

[1] D. Wang *et al.*, “Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China,” *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 11, pp. 1061–1069, Mar. 2020, doi: 10.1001/jama.2020.1585.

[2] Y. T. Xiang *et al.*, “Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed,” *The Lancet Psychiatry*, vol. 7, no. 3, pp. 228–229, 2020, doi: 10.1016/S2215-0366(20)30046-8.

[3] L. Kang *et al.*, “Impact on mental health and perceptions of psychological care among medical and nursing staff in Wuhan during the 2019 novel coronavirus disease outbreak: A cross-sectional study,” *Brain. Behav. Immun.*, vol. 87, no. March, pp. 11–17, 2020, doi: 10.1016/j.bbi.2020.03.028.

[4] V. A. Naushad *et al.*, “A Systematic Review of the Impact of Disaster on the Mental Health of Medical Responders,” *Prehosp. Disaster Med.*, vol. 34, no. 6, pp. 632–643, Dec. 2019, doi: 10.1017/S1049023X19004874.

[5] N. W. S. Chew *et al.*, “A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak.,” *Brain. Behav. Immun.*, vol. 88, no. January, pp. 559–565, 2020, doi: 10.1016/j.bbi.2020.04.049.

[6] A. Gorini *et al.*, “Mental health and risk perception among Italian healthcare workers during the second month of the Covid-19 pandemic,” *Arch. Psychiatr. Nurs.*, vol. 34, no. 6, pp. 537–544, 2020, doi: 10.1016/j.apnu.2020.10.007.

[7] Z. Zhu *et al.*, “COVID-19 in Wuhan: Immediate psychological impact on 5062 Health Workers,” *medRxiv*, no. 1095, 2020, doi: 10.1101/2020.02.20.20025338.

[8] Z. Ashtari, Y. Farhady, and M. R. Khodaee, “Relationship between job burnout and work performance in a sample of Iranian mental health staff,” *African J. Psychiatry (South Africa)*, vol. 12, no. 1, pp. 71–74, Feb. 2009, doi: 10.4314/ajpsy.v12i1.30282.

[9] B. M. Pauly, C. Varcoe, and J. Storch, “Framing the issues: Moral distress in health care,” *HEC Forum*, vol. 24, no. 1, pp. 1–11, 2012, doi: 10.1007/s10730-012-9176-y.

[10] A. Pollock *et al.*, “Interventions to support the resilience and mental health of frontline health and social care professionals during and after a disease outbreak, epidemic or pandemic: a mixed methods systematic review,” *Cochrane database*

Syst. Rev., vol. 11, p. CD013779, Nov. 2020, doi: 10.1002/14651858.CD013779.

[11] F. C. T. da Silva and M. L. R. Neto, “Psychological effects caused by the COVID-19 pandemic in health professionals: A systematic review with meta-analysis,” *Prog. Neuro-Psychopharmacology Biol. Psychiatry*, vol. 104, no. August 2020, p. 110062, 2021.

[12] X. Hu and W. Huang, “Protecting the psychological well-being of healthcare workers affected by the COVID-19 outbreak: Perspectives from China,” *Nurs. Heal. Sci.*, vol. 22, no. 3, pp. 837–838, 2020, doi: 10.1111/nhs.12727.

[13] N. Vindegaard and M. E. Benros, “COVID-19 pandemic and mental health consequences: Systematic review of the current evidence,” *Brain. Behav. Immun.*, vol. 89, pp. 531–542, Oct. 2020, doi: 10.1016/j.bbi.2020.05.048.

[14] S. Pappa, V. Ntella, T. Giannakas, V. G. Giannakoulis, E. Papoutsi, and P. Katsaounou, “Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis,” *Brain. Behav. Immun.*, vol. 88, no. May, pp. 901–907, 2020, doi: 10.1016/j.bbi.2020.05.026.

[15] Q. Chen *et al.*, “Mental health care for medical staff in China during the COVID-19 outbreak,” *The Lancet Psychiatry*, vol. 7, no. 4, pp. e15–e16, 2020, doi: 10.1016/S2215-0366(20)30078-X.

[16] A. Cooper, “Design for Pleasure,” in *The inmates are running the asylum Why High-Tech Products Drive Us Crazy and How to Restore the Sanity*, 2004, pp. 123–149.

[17] C. LeRouge, J. Ma, S. Sneha, and K. Tolle, “User profiles and personas in the design and development of consumer health technologies,” *Int. J. Med. Inform.*, vol. 82, no. 11, pp. e251–e268, 2013, doi: 10.1016/j.ijmedinf.2011.03.006.

[18] R. J. Holden, A. Kulanthaivel, S. Purkayastha, K. M. Goggins, and S. Kirpalani, “Know thy eHealth user: Development of biopsychosocial personas from a study of older adults with heart failure,” *Int. J. Med. Inform.*, vol. 108, no. October, pp. 158–167, 2017, doi: 10.1016/j.ijmedinf.2017.10.006.

[19] S. Vosbergen *et al.*, “Using personas to tailor educational messages to the preferences of coronary heart disease patients,” *J. Biomed. Inform.*, vol. 53, pp. 100–112, 2015, doi: 10.1016/j.jbi.2014.09.004.

[20] V. Haldane *et al.*, “User preferences and persona design for an mhealth intervention to support adherence to cardiovascular disease medication in singapore: A multi-method study,” *JMIR mHealth uHealth*, vol. 7, no. 5, 2019, doi: 10.2196/10465.

[21] O. Bhattacharyya, K. Mossman, L. Gustafsson, and E. C. Schneider, “Using human-centered design to build a digital health advisor for patients with complex needs: Persona and prototype development,” *J. Med. Internet Res.*, vol. 21, no. 5, pp. 1–12, 2019, doi: 10.2196/10318.

[22] I. Klooster, J. Wentzel, F. Sieverink, G. Linssen, R. Wesselink, and L. Gemert-Pijnen, “Personas for Perfectly Tailored eHealth Technologies: Segmenting Heart Failure Patients using the Persona Approach Twente (Preprint).” 2020, doi: 10.2196/preprints.24172.

[23] N. Pattisapu, M. Gupta, P. Kumaraguru, and V. Varma, “A distant supervision based approach to medical persona classification,” *J. Biomed. Inform.*, vol. 94, no. April, p. 103205, 2019, doi: 10.1016/j.jbi.2019.103205.

[24] S. Koch, A. Sheeren, and N. Staggers, “Using personas and prototypes to define nurses’ requirements for a novel patient monitoring display,” *Stud. Health Technol. Inform.*, vol. 146, pp. 69–73, 2009, doi: 10.3233/978-1-60750-024-7-69.

[25] K. Schäfer *et al.*, “Survey-based personas for a target-group-specific consideration of elderly end users of information and communication systems in the German health-care sector,” *Int. J. Med. Inform.*, vol. 132, no. July, p. 103924, 2019, doi: 10.1016/j.ijmedinf.2019.07.003.

[26] J. E. Gentle, L. Kaufman, and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis.*, vol. 47, no. 2. John Wiley & Sons, 1991.

[27] L. Nielsen, K. S. Nielsen, J. Stage, and J. Billestrup, “Going Global with Personas,” in *Human-Computer Interaction -- INTERACT 2013*, 2013, pp. 350–357.

[28] J. Grudin and J. Pruitt, “Personas, participatory design and product development: An infrastructure for Engagement,” *Proc. PDC*, 2002.

[29] Regione Lombardia, “Dashboard Covid-19,” 2020. https://www.regione.lombardia.it/wps/portal/istituzionale/HP/servizi-e-informazioni/cittadini/salute-e-prevenzione/coronavirus/dashboard-covid19 (accessed Nov. 30, 2020).

[30] Consiglio dei Ministri, “Decreto del Presidente del Consiglio dei Ministri 17 maggio 2020.” 2020, [Online]. Available: https://www.gazzettaufficiale.it/eli/id/2020/05/17/20A02717/sg.

[31] C. Maslach and S. E. Jackson, “The measurement of experienced burnout,” *J. Organ. Behav.*, vol. 2, no. 2, pp. 99–113, 1981, doi: 10.1002/job.4030020205.

[32] D. S. Weiss, “The Impact of Event Scale: Revised,” in *Cross-Cultural Assessment of Psychological Trauma and PTSD*, J. P. Wilson and C. S. Tang, Eds. Boston, MA: Springer US, 2007, pp. 219–238.

[33] K. Kroenke, R. L. Spitzer, J. B. W. Williams, and B. Löwe, “An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ–4,” *Psychosomatics*, vol. 50, no. 6, pp. 613–621, 2009, doi: 10.1016/s0033-3182(09)70864-3.

[34] L. F. Kerper *et al.*, “Screening for depression, anxiety, and general psychological distress in pre-operative surgical patients: A psychometric analysis of the Patient Health Questionnaire 4 (PHQ-4),” *Clin. Heal. Promot. - Res. Best Pract. patients, Staff community*, vol. 4, no. April 2014, pp. 5–14, 2014, doi: 10.29102/clinhp.14002.

[35] N. Schutte, S. Toppinen, R. Kalimo, and W. B. Schaufeli, “The factorial validity of the Maslach Burnout Inventory-General Survey (MBI-GS) across occupational groups and nations,” *J. Occup. Organ. Psychol.*, vol. 73, no. 1, pp. 53–66, 2000, doi: 10.1348/096317900166877.

[36] R. Bro and A. K. Smilde, “Principal component analysis,” *Anal. Methods*, vol. 6, no. 9, pp. 2812–2831, 2014, doi: 10.1039/c3ay41907j.

[37] R. M. V. R. Almeida, A. F. C. Infantosi, J. H. R. Suassuna, and J. C. G. D. Costa, “Multiple correspondence analysis in predictive logistic modelling: Application to a living-donor kidney transplantation data,” *Comput. Methods Programs Biomed.*, vol. 95, no. 2, pp. 116–128, 2009, doi: 10.1016/j.cmpb.2009.02.003.

[38] J. Pagès, “Factorial Analysis of Mixed Data,” in *Multiple Factor Analysis by Example Using R*, Chapman and Hall/CRC, 2014, pp. 67–78.

[39] M. Linting and A. van der Kooij, “Nonlinear principal components analysis with CATPCA: a tutorial,” *J. Pers. Assess.*, vol. 94, no. 1, pp. 12–25, 2012.

[40] L. L. Thurstone, “Multiple factor analysis.,” *Psychol. Rev.*, vol. 38, no. 5, p. 406,

1931.

[41] University of Cincinnaty, “K-means Cluster Analysis,” 2018. https://uc-r.github.io/kmeans_clustering#fn:kauf (accessed Aug. 30, 2020).

[42] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[43] J. H. McDonald, “Handbook of Biological Statistics - Paired t–test,” *Sparky House Publ.*, pp. 180–185, 2014, [Online]. Available: http://www.biostathandbook.com/pairedttest.html.

[44] W. H. Kruskal and W. A. Wallis, “Use of Ranks in One-Criterion Variance Analysis,” *J. Am. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952, doi: 10.1080/01621459.1952.10483441.

[45] J. Pruitt and T. Adlin, *Persona Conception and Gestation*. 2006.

[46] T. Karras and NVIDIA, “This Person Does Not Exist,” 2019. https://thispersondoesnotexist.com/ (accessed Sep. 10, 2020).

[47] B. Löwe *et al.*, “A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population,” *J. Affect. Disord.*, vol. 122, no. 1–2, pp. 86–95, Apr. 2010, doi: 10.1016/j.jad.2009.06.019.

[48] J. Lai *et al.*, “Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019,” *JAMA Netw. open*, vol. 3, no. 3, p. e203976, 2020, doi: 10.1001/jamanetworkopen.2020.3976.

[49] S. Kisely, N. Warren, L. McMahon, C. Dalais, I. Henry, and D. Siskind, “Occurrence, prevention, and management of the psychological effects of emerging virus outbreaks on healthcare workers: rapid review and meta-analysis,” *BMJ*, vol. 369, p. m1642, 2020, doi: 10.1136/bmj.m1642.

[50] T. W. Wong *et al.*, “The psychological impact of severe acute respiratory syndrome outbreak on healthcare workers in emergency departments and how they cope.,” *Eur. J. Emerg. Med.*, vol. 12, no. 1, pp. 13–18, Feb. 2005, doi: 10.1097/00063110-200502000-00005.

[51] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[52] P. Fusar-Poli, P. Brambilla, and M. Solmi, "Learning from COVID-19 pandemic in northen italy: Impact on mental health and clinical care," *J. Affect. Disord.*, vol. 275, no. January, pp. 78–79, 2020, doi: 10.1016/j.jad.2020.06.028.

[53] D. Fessell and C. Cherniss, "Coronavirus Disease 2019 (COVID-19) and Beyond: Micropractices for Burnout Prevention and Emotional Wellness," *J. Am. Coll. Radiol.*, vol. 17, no. 6, pp. 746–748, 2020, doi: 10.1016/j.jacr.2020.03.013.

[54] J. Raudenská *et al.*, "Occupational burnout syndrome and post-traumatic stress among healthcare professionals during the novel coronavirus disease 2019 (COVID-19) pandemic.," *Best Pract. Res. Clin. Anaesthesiol.*, vol. 34, no. 3, pp. 553–560, Sep. 2020, doi: 10.1016/j.bpa.2020.07.008.

[55] N. Restauri and A. D. Sheridan, "Burnout and Posttraumatic Stress Disorder in the Coronavirus Disease 2019 (COVID-19) Pandemic: Intersection, Impact, and Interventions," *J. Am. Coll. Radiol.*, vol. 17, no. 7, pp. 921–926, 2020, doi: 10.1016/j.jacr.2020.05.021.

[56] B. Ferreira, W. Silva, E. Oliveira, and T. Conte, "Designing personas with empathy map," *Proc. Int. Conf. Softw. Eng. Knowl. Eng. SEKE*, vol. 2015-Janua, no. May, pp. 501–505, 2015, doi: 10.18293/SEKE2015-152.