JOURNAL OF COMMUNICATION IN HEALTHCARE
https://doi.org/10.1080/17538068.2022.2026055

Psychological and situational profiles of social distance compliance during COVID-19

Michael Robert Haupt [a], Staci Meredith Weiss[b,c], Michelle Chiu[b], Raphael Cuomo[d], Jason M. Chein[b] and Tim Mackey [e]

[a]Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA; [b]Department of Psychology, Temple University, Philadelphia, PA, USA; [c]Department of Psychology, University of Cambridge, Cambridge, UK; [d]Department of Anesthesiology and Division of Infectious Diseases and Global Public Health, University of California San Diego School of Medicine, La Jolla, CA, USA; [e]Global Health Program, Department of Anthropology, University of California, San Diego, La Jolla, CA, USA

ABSTRACT

Background: Health advice in the wake of the COVID-19 pandemic has called upon the public to re-evaluate risk associated with recently routine behavior. However, differences in demographics, situational circumstances, and psychological dispositions create inequities in how people are able to respond to risks presented by the virus.
Method: Within a sample of 482 Americans, we examined the frequency of behavior reconceptualized as 'risky' by CDC public health guidelines released on 30 March 2020. We applied a cluster analysis using a data-driven persona framework from the field of user-design research, using only situational and dispositional (i.e. psychological) variables to identify profiles of individuals.
Results: This profile approach unpacked important variability in the evaluation of risk for COVID-19 contagion, as well as adherence to public health guidance. Profiles engaged in high-risk behaviors were more likely required to work on-site and report higher financial impact related to the pandemic.
Conclusion: Applying the profile approach facilitates personalized communications tailored to the psychological and situational circumstances of each profile that can promote compliance with public health guidelines and guide policy decisions. These results also suggest that risk-taking behaviors within the context of COVID-19 may also be driven by factors related to economic inequity since those who are more likely to be essential workers do not have the ability to remain as compliant to social distancing compared to those with higher economic status. Recommendations for policies promoting federally mandated paid leave policy in the US and employing qualitative methods for future research is discussed.

KEYWORDS
Health inequities; COVID-19; communication; human characteristics; health risk behaviors; public health; psychometrics; data-driven personas

Background on COVID-19

Since March 2020, the lives of citizens across the United States (US) and around the globe have been upended by the emergence of the COVID-19 pandemic. For some countries, this change has occurred even sooner, with China reporting cases of the virus as early as November 2019 [1]. On 23 March, the Centers for Disease Control and Prevention (CDC) distributed the first set of guidelines for how individuals can mitigate their 'risk' for Coronavirus infection and contagion. This advice introduced 'social distance' into public discourse, imploring individuals to minimize physical proximity to others outside of their household by maintaining 6 feet of distance with others when interacting outdoors or in public areas.

On 2 April, the CDC updated its recommendations, advising the use of face coverings or masks in situations 'at risk' for violations of social distance and urging avoidance of unnecessary exposure at visits to businesses or public spaces [2]. The tone of public health authorities shifted from suggesting augmented activity to imploring minimized activity, such that recently routine activities were newly associated with confronting 'risk' of harm to oneself or others. In the absence of a federal mandate, by 10th April, more than 95% of the American population was under advisement to minimize their activity and risk for infection, as state governments and local municipalities enacted 'Stay at Home' recommendations or 'Shelter-in-place' orders [3]. These recommendations advised citizens to stay indoors and only venture outside of one's residence for 'essential' errands. However, despite the cooperation by government, business and authorities to promote (and in some cases, enforce) social distancing, many members of the public show signs of restlessness towards a stay at home orders, even as confirmed case counts and the

CONTACT Michael Robert Haupt mhaupt@ucsd.edu Department of Cognitive Science, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA
Supplemental data for this article can be accessed at https://doi.org/10.1080/17538068.2022.2026055.

mortality rates did not experience sharp declines in some regions of the country [4].

Recent studies show evidence, for instance, that Black/African Americans have been disproportionately infected by COVID-19 [5], while data from New York City show higher COVID-19 related death rates for both Black/African and Hispanic/Latino persons [6]. People categorized as essential workers in Health, Food/Agriculture, and Infrastructure industrial sectors might exchange employment security at the risk of higher exposure to the public, thus raising their odds of contracting and spreading the virus [7]. Additionally, over a quarter of private-sector workers in the United States do not receive paid sick leave (including over 30% of workers in the South and Midwest), which could cause further spread of the disease that disproportionately impacts certain at-risk populations [8,9]. Hence, social determinants of health such as poverty, ethnicity, employment status, healthcare access, and other known factors are likely exacerbated and accentuated for already vulnerable populations in the face of COVID-19.

Assessing variability in risky behavior and public health compliance using data-driven personas

In order to promote the description of our sample rather than pigeonhole people by explanations for newly adopted compliance attitudes and risk behaviors [10], we employ a data-driven 'personas' approach to characterize individual differences in reactions to social distance guidelines due to the COVID-19 pandemic. Personas provide a description of hypothetical users that present information such as demographics, behaviors, and attitudes about user groups [11]. Personas are used across a wide range of industries and have guided user design for developing educational software [12] and healthcare technologies [13,14], promoting digital accessibility among people with disabilities [15], and have also been used by technology-focused companies such as Sony [16] and Microsoft [11,17]. A study using a panel of design experts lists audience focus (i.e. developing a product for the development of users and their goals), the challenging of long-standing and often incorrect assumptions about users, and the prevention of self-referential design (i.e. realizing that the user is different from the designer) as some of the most important benefits of personas [18]. While personas were originally created using qualitative approaches (e.g. focus groups, ethnographies), more recent studies have shown that personas produced using quantitative analysis are just as effective, less subjective and more reliable [19–21], and have been successful in identifying types of users based on large-scale datasets [22,23].

This paper extends the use of data-driven personas to produce profiles that outline individual differences in how US citizens perceive risk during the COVID-19 pandemic and comply with social distancing guidelines. The intention behind using a persona-based approach is to characterize layperson understanding of COVID-19 and assist public health and government officials in composing communications and policies related to the virus. The profiles can also be used in conceptualizing potential users for virtual tools and apps designed for public use.

Methods

Our survey was developed and piloted in a small sample of undergraduates for distribution via Qualtrics. It used a range of existing psychological measures and batteries, as well as our own and other newly developed items, which were created to interrogate aspects specific to COVID-19. Participants were consented and compensated at a rate of $7/hour for completing the 30-minute survey, recruited from Amazon Mechanical Turk (MTurk).

Participants

A total of 514 participants were recruited from MTurk with the aim to have a sample size approaching $n =$ 500 (aiming for a power of .80 for even smaller effect sizes, using the association between traditional Benthin risk inventory and BFI found in prior studies as a basis for our exploratory measure), after filtering participants who did not pass data quality checks [24]. Recruitment occurred between April 30th and May 2nd following the first full month of quarantine within the United States. The final sample size was $n = 482$ after removing poor quality and/or incomplete responses. Data quality checks included: speed outliers and/or incomplete responses (determined using Mahalanobis distance, $M = 32.5$ minutes to complete, $n = 9$); response to the question 'Estimate the Date you first modified your behavior due to the Coronavirus' (to identify malingerers, $n = 20$); and duplicate MID or completion codes ($n = 3$).

Our sample demographics roughly reflected the ethnic breakdown of the American public according to 2014–2018 Census data, with a significant underrepresentation of individuals identifying as ethnically Hispanic [25]. Age within the sample ranged from 18 to 73 with an average age of 37.12 (SD = 11.33) and 59% of participants identified as men. Additionally, 71.2% of participants were White, 19.7% Black, 6.0% Asian, and 7.1% Hispanic. See Appendix A for more demographic information about the full sample. Ethics approval for this study was granted by Temple University. MTurk participants were compensated based on standard survey-taking rates on the platform.

Self-report measures

COVID-19 Risk Taking Inventory (*CRI*). A 10-item questionnaire was developed, adapted from the structure and format based on the Benthin Risk Perception Scale [26], to assess a set of activities (e.g. attending a gathering of more than five people) that under the conditions of the COVID-19 pandemic are considered 'risky'. Items were developed to target the discrepancies in activities identified as 'risky' per CDC guidelines in April, as opposed to those released originally on 23 March. Each activity item was followed by four questions that asked about the frequency of engagement since the estimated date when respondents first began modifying their behavior due to COVID-19, a risk-benefit comparison, and risk propensity toward the self or others. The Cronbach's alpha for the full scale was above 0.90 for all the 10 items, across each of the four sub-questions: risk behavior frequency, cost–benefit evaluation, risk to self, and risk to other. This measure served as our primary outcome variable. Factor analysis was used to identify which behaviors were high-risk, low-risk and essential (travel-related items were not included in the further analysis). For a full detailed index of the items and questions used, see Supplementary Information.

Situational Factors. In assessing situational factors that may influence self-reported behavior, participants were asked to respond to questions relevant to Living Space Access ('How many rooms within your current residence do you feel comfortable relaxing or spending time in that are not your bedroom? This can also include outdoor spaces that are on your property') and Perceived Scarcity of resources ('It has been difficult for me to get needed resources (food, toilet paper) due to the Coronavirus'), the latter of which was rated on a 7-point Likert scale ranging from 1 ('strongly disagree') to 7 ('strongly agree').

Big Five Personality Inventory. A 23-item questionnaire based on the Big Five Inventory (BFI) [27–29] was used to evaluate participants across the Big Five personality dimensions (extraversion, agreeableness, conscientiousness, neuroticism, openness).

Empathy. Empathy was assessed by having participants respond to the 7-item Perspective Taking subscale taken directly from the Interpersonal Reactivity Index [30], with each item rated on a 5-point Likert scale ranging from 1 ('does not describe me well') to 5 ('describes me well').

Behavioral Inhibition/Behavioral Activation Scales. The Behavioral Inhibition/Behavioral Activation Scales [31] index the two motivational systems of Behavioral Inhibition and Behavioral Activation [32,33]. The BIS scale includes a subscale measuring sensitivity to aversive motivation (e.g. 'criticism or scolding hurts me quite a bit'). The BAS scale measures sensitivity to the mechanism underlying appetitive motivation by using three subscales, namely: drive (e.g. 'When good things happen to me, it affects me strongly'), fun-seeking (e.g. 'I often act on the spur of the moment'), and reward responsiveness (e.g. 'When I get something I want, I feel excited and energized').

k-Means clustering

Groups for the analysis were created using *k*-means clustering using variables related to psychological and situational circumstances. In order to investigate the variables that contribute most to risk-taking behavior, we implemented a Shapley Value regression, which assesses the relative importance of all the independent variables within a model by first computing all possible combinations of the independent variables, and then determining how much each variable contributes to the total R^2 of the model (see [34] for a more detailed description). First, all relevant ordinal variables (12-criterion model) were entered, accounting for 43% of the variance in risky behavior (see Table B1 in Appendix for full regression results). However, variables in this model were highly correlated – following the correlation output, we pruned variables for non-significance and multicollinearity violations (Conscientiousness was detected as a variance inflation factor greater than 10). The variables remaining in the leaner, subsequent regression (7-criterion model) guided the selection of input variables to be used in the final clustering model. Since variables with larger values contribute more to the distance measure in k-means clustering than variables with smaller values [35] we converted the psychometric scales into binary variables using the sample median score pertaining to each trait, such that participants with scores below the median are classified as low level and participants with scores above the median are classified as high level. By converting the variables into binaries, we prevented scales with larger ranges from overcontributing to the model. Additionally, Living Space Access (number of common spaces) was re-coded into a binary variable from the original 5-point scale, wherein responses were split into two groups, those who responded between 0 and 3 and those who responded 3 or above. We also re-coded Age into a 4-point scale based on quartile scores of the original continuous numeric variable.

The final cluster model is the result of an iterative process, which tested different combinations of input variables and group number. In total, 20 models were created and tested during analysis. A model was considered viable if it met the following criteria:

(1) Each group within the cluster model must show differentiation from one another
(2) Every input variable must have unique relationships with each group in the model (i.e. no two

input variables should have identical correlations with each group)

(3) The contribution of each input variable within the model must be statistically significant, and

(4) The distribution of the total sample must not be overly concentrated in one group.

Appendix D depicts multiple comparison results between cluster means of each input variable to test for differentiation between clusters. In order to observe distinctions among groups with overlapping characteristics (e.g. comparing two groups in the same age cohort but differ in psychological or situational circumstances), we allowed for some input variables do not have statistically significant differences between a limited number of groups.

Results

Factor analysis

A factor analysis was conducted to group activities together with each factor labeled using the item-level mean for activity risk assessment. See Appendix E for further description. Activities labeled as *essential* included going to the grocery store, going outside without a mask, and exercising outside in public. *Low-risk* activities included returning home without washing hands, meeting a friend while maintaining social distance and visiting a public space. Finally, *high-risk* activities included attending a gathering with more than five guests, not in your household, as well as interacting with a stranger for essential purposes.

Cluster analysis

Groups for the analysis were created using *k*-means clustering using variables related to psychological and situational circumstances. The final input variables used in the model in this paper are introversion scores measured by the BFI, sensation seeking scores measured by BIS/BAS, perspective-taking empathy scores from the IRI scale, age, living space access (whether or not the participant lives in a residence with more than two common spaces), and perceived scarcity (how much participants agree with the statement 'it has been difficult for me to get needed resources (food, toilet paper) due to the Coronavirus' (1 Strong disagree to 7 Strongly agree)). To ensure that there was adequate differentiation between clusters, only models that showed a majority of statistically significant comparisons were considered. Additionally, an ANOVA was conducted on input variables to assess the significant contribution of each variable. Complete results from the cluster analyses are presented in Appendix D. All SPSS syntax used to produce the k-means cluster model are also included as supplementary material. Table 1 shows high-level summaries and a label for each cluster, based on both input and selected outcome variables. See Appendix C for a more detailed look at demographic and behavioral variables related to each cluster. Table 2 shows the percentage of each cluster that at least somewhat agrees with statements that reflect attitudes towards quarantine restrictions, and Table 3 details messaging implications for each profile cluster.

Risky behavior and activity by cluster

We identified the risk-taking propensity of clusters by first providing each subject with a regression score for their risk-taking, given their subject-specific cost-to-benefit evaluation. The clusters were ranked 1–7 by the proportion of individuals who engaged in activities they rated as 'risks outweigh the benefits' – this is the order in which they are subsequently presented. A triadic split was applied to discriminate which clusters had the lowest (clusters 7 and 2) and highest risk-taking propensity (clusters 4 and 3), and labeled them as 'risk averse' and 'risk inclined', respectively. Moderate risk-takers were labeled as 'compliant' (clusters 5, 6 and 1).

Using ANOVA, we examined cluster differences in frequency of activity engagement across all CRI items (overall activity) and three factors of risk activities, determined by the factor analysis. Overall activity differed by cluster, F (6) = 3.062, SS = 40.387, P = 0.006, $\omega^2 = 0.025$, such that individuals belonging to high-risk clusters 3 and 4 reported engaging more frequently in all activities, particularly evident relative to cluster 7. High-risk activity differed by cluster, F (6) = 14.974, SS = 48.657, P < 0.001, $\omega^2 = 0.148$, such that individuals belonging to cluster 3 reported engaging more frequently in high-risk activities relative to all other individuals (the significance was marginal when compared to individuals from high-risk clusters, 3 and 4). Individuals belonging to risk-averse cluster 7 engaged in significantly less high-risk activity relative to the individuals in clusters identified as compliant and high-risk. Low-risk activity differed by cluster, F (6) = 7.365, SS = 23.364, P < 0.001, $\omega^2 = 0.073$, such that individuals belonging to clusters 3 and 4 reported engaging more frequently in low-risk activities relative to individuals in all other clusters. Individuals belonging to risk-averse cluster 2 engaged in significantly less activity relative to the individuals in clusters identified as compliant and high-risk. Essential activities did not differ by cluster, F (6) = 1.286, SS = 18.761, P = 0.262, $\omega^2 = .002$. See Appendix E for further analysis of cluster group comparisons of employment-related situational variables.

Table 1. Summary of cluster groups based on input variables and demographic differences.

Cluster groups	Psychological and situational summary based on input variables	Demographic and quarantine compliance summary (based on select output variables, see Appendix for corresponding data) SA = Sample average
Cluster 1 (*n* = 53) 'Unsure, yet Careful Boomer'	Middle in terms of introversion and highest in age. Also moderate in sensation seeking and perceived scarcity. Lower on empathy and not likely to have large living space	Skews low income (42% vs. 35% SA) and majority live in suburbs (55%). Also leans towards Conservative (57%). Shows lower levels of trust toward organizations like CDC compared to other clusters (70% vs. 81% SA), but still remains cautious against COVID. More likely to get COVID information from cable news relative to other clusters. Least likely to say that going to a public space or using public transit is high risk. Most likely to say that employment status did not change since COVID (67% vs. 57% SA). Skews mainly White (87%)
Cluster 2 (*n* = 58) 'Introverted Millennial'	Highly introverted and likely to live in a large living space. Still fairly young and very low in sensation seeking. Lower-middle end in empathy. Has low levels of perceived scarcity during the pandemic	Younger, largely skews Millennial and Gen Z. Least likely to be employed full-time (55% vs. 75% SA). Mostly Liberal (67%) and majority live in urban or suburban areas (90%). Takes quarantine seriously and is supportive of CDC. Complies with quarantine guidelines but does occasionally go outside for leisure despite being the least likely to agree that going outside for non-essential needs is ok (29% vs. 48% SA)
Cluster 3 (*n* = 93) 'Rural-leaning Millennial'	Not too introverted and very young. Scores low in both sensation seeking and empathy. Not likely to live in a large living space and has scores moderately high in perceived scarcity	Most rural relative to other Clusters (33% vs. 22% SA). Majority middle income but also skews lower (94% earn less than $99,999 a year). Leans Conservative (64%) and more likely to have children relative to other clusters (60% vs. 49% SA). Recognizes risks of COVID but generally does not practice strict quarantine compliance. Has been impacted by COVID financially and has trouble finding resources. Still takes Uber and public transit but might be related to higher likelihood of being an essential worker (25% vs. 16% SA). Majority White (58%) but also skews Black American (29% vs 17% SA)
Cluster 4 (*n* = 109) 'Financially-Impacted Gen Xer'	Low introversion and very high in sensation seeking. Middle age range with moderate levels of perspective-taking empathy. Some live in a larger living space, but many others do not. Reports the highest in perceived scarcity	Mostly middle income (62%) but skews lower income as well (31%). More likely to be Conservative (63%). Majority live in urban areas (41%), but many also live in rural and suburban regions. Middle age and most likely to have kids (62% vs. 49% SA). Respects authority figures like CDC but is also very lax on complying with quarantine. Believes that a lot of restrictions are excessive. Most likely to be an essential worker (28% vs. 16% SA) and also be impacted financially by COVID (80% vs. 53% SA). Majority White (61%) but also skews Black American (31% vs. 17% SA)
Cluster 5 (*n* = 41) 'Work-Driven Young Adult'	Lowest in introversion and the youngest cluster. High levels of sensation seeking and fairly empathetic. Not too likely to live in a larger living space but also reports fairly low in perceived scarcity	Most likely to be working full-time (93% vs. 75% SA). Majority are Liberal (63%) and Millennial/Gen Z. Over 90% live in either urban or suburban areas. Respects quarantine compliance and mainly goes outside for essential needs. Reports the lowest amount of hours outside for leisure reasons (2.45 vs. 5.28 SA). Sees risk in public spaces and generally avoids them
Cluster 6 (*n* = 75) 'Cautious Suburbanite'	High in both introversion and age. Very likely to live in a larger living pace but reports moderately high in perceived scarcity. Fairly low in sensation seeking with moderate levels of empathy	Majority live in Suburbs (55%) but also skews rural (27%). More likely to have low income relative to other clusters (47% vs. 35% SA). Political orientation is fairly spread out. Very wary of COVID threat and perceives many activities as high risk. Generally avoids going outdoors. Has been impacted financially by COVID but still complies to quarantine
Cluster 7 (*n* = 53) 'Authority-respecting Boomer'	Moderate in introversion but very high in age, sensation seeking, and perspective-taking empathy. Most likely to live in a large living space and scores very low in perceived scarcity	Most likely to have 100k+ income compared to other clusters (28% vs. 15% SA). Liberal leaning (23% Very Liberal vs. 15% SA; however only 53% are Liberal overall) and more suburb/rural centered (77%). Compliant to quarantine and perceives many non-essential activities as risky. Main reason to go outside is for employment. Respects the CDC and WHO the most out of the other clusters. Least likely to say that COVID pandemic has impacted them financially (25% vs. 53% SA). Rarely uses public transit, but it might be because they have access to other modes of transportation. Least racially diverse: 89% White

Discussion

These profiles accounted for significant variance in non-essential behavior deemed non-compliant and at high risk for viral transmission, as well as identifying compliant behavior that confers low risk for contagion, capturing a spectrum of adherence to hygiene vigilance, social distance and protective equipment guidelines. Across all items, participants consistently viewed the perceived risk to other as greater to themselves, except for the use of public transit. We found a quadratic effect, such that individuals identified as normative compliant (the majority of participants) rather than risk-averse compliant (clusters 7 and 2, the lowest in risk behavior) or risk-inclined compliant (clusters 3 and 4, the highest in risk behavior) had the smallest gap in their assessment of risk to self-versus risk to others. It is conjectured that while most individuals are influenced by public health messaging aimed to invoke consistency in community response, perhaps at the expense of individual choice, the most risk-

Table 2. Proportion of each cluster that at least 'somewhat agrees' to each statement.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
While it is important to take precautions against the Coronavirus, it is also important that we do not give up our freedoms	60.4%$_{a}$	46.6%$_{a}$	65.6%$_{a,b}$	82.6%$_{b}$	56.1%$_{a}$	53.3%$_{a}$	54.7%$_{a}$
It is ok to go outside for non-essential trips as long as I'm careful	39.6%$_{a}$	29.3%$_{a}$	69.9%$_{b}$	65.1%$_{b}$	53.7%$_{a,b}$	41.3%$_{a}$	34.0%$_{a}$
The Corona virus has impacted me negatively from a financial point of view	45.3%$_{a,d,e}$	37.9%$_{a,b}$	75.3%$_{c}$	79.8%$_{c}$	43.9%$_{a,d,e}$	66.7%$_{c,d}$	24.5%$_{b,e}$
I feel that the social distancing measures has been excessive	39.6%$_{a,d,f,g}$	22.4%$_{a,b}$	66.7%$_{c}$	63.3%$_{c,d,e}$	36.6%$_{a,e,f,g}$	40.0%$_{b,f}$	20.8%$_{b,g}$
I take announcements and guidelines from the Center for Disease Control (CDC) and the World Health Organization (WHO) very seriously	69.8%$_{a}$	72.4%$_{a}$	83.9%$_{a,b}$	91.7%$_{b}$	78.0%$_{a,b}$	78.7%$_{a,b}$	92.5%$_{a,b}$
I catch up on national news or press conferences from federal government officials most days	56.6%$_{a}$	56.9%$_{a}$	76.3%$_{a,b}$	80.7%$_{b}$	61.0%$_{a,b}$	69.3%$_{a,b}$	73.6%$_{a,b}$
I follow local news or press conferences from state government officials most days	69.8%$_{a,c,d}$	58.6%$_{a,b}$	77.4%$_{a,c,d}$	87.2%$_{c}$	73.2%$_{a,c,d}$	56.0%$_{b,d}$	75.5%$_{a,c,d}$

Note: Values in the same row and subtable not sharing the same subscript are significantly different at $P < 0.05$ in the two-sided test of equality for column proportions. Cells with no subscript are not included in the test. Tests assume equal variances and are adjusted for all pairwise comparisons within a row of each innermost subtable using the Bonferroni correction.

averse individuals may be motivated by protecting their own health. Similarly, individuals who might otherwise be risk-inclined compliant or even actively non-compliant with CDC guidelines might be more responsive to public health messaging if they feel an activity might have a personal cost to them, beyond the utility of the community benefit of adherence.

Profiles of compliance from dispositional and situational factors

From conducting k-means clustering, we produced a model with seven cluster groups based on psychological and situational variables. In a social distancing context, cluster 1 ('Unsure, yet Careful Boomer', see Table 1) remains compliant yet cautious towards

Table 3. Messaging implications by cluster.

Cluster number	Messaging implication
1	For Cluster 1 ('Unsure, yet Careful Boomer') which shows moderate risk-taking propensity, higher age, and lower levels of empathy, messages that emphasize how the virus negatively impacts one's personal health are more likely to encourage social distancing compliance. Based on results shown in Table A4 in the appendix, which shows where each cluster receives COVID-related information, cable news networks might also be an effective channel to reach this cluster
2	Cluster 2 ('Introverted Millennial') is comprised of individuals high in introversion and living space, and lower in age, perceived scarcity, and risk-taking propensity. Since this cluster is already compliant with public health measures, messaging for this group should focus on ways of mitigating quarantine fatigue. More specifically, messages that showcase activities that are safe for social distancing and what type of outdoor environments have the lowest risk of spreading the virus can help this cluster maintain its higher level of commitment to personal and public health. Indoor activities should also be suggested to take advantage of this group's higher levels of introversion and likelihood of living in a larger space
3	Cluster 3 ('Rural-leaning Millennial') consists of individuals low in age, living space size, introversion, and empathy but high in risk-taking propensity and perceived risk. Since this cluster is more likely to go outside, whether due to work-related reasons or smaller living space which makes long periods of staying indoors difficult, messaging towards this cluster should focus on what precautions one can take when going outside. For those who have to take public transportation in order to commute to work, recommendations for how to stay safe in those environments should be emphasized. In counties where there are lower cases of infection and there are non-essential businesses open such as restaurants or bars, guidelines for how to identify the risk level of spaces (e.g. amount of open space, adherence to mask-wearing ordinances, etc …) can help mitigate spreading of the virus for those who feel the need to go outdoors. Due to the lower levels of empathy, risks to personal health should be focused on when communicating caution concerning COVID-19
4	Similar to Cluster 3, Cluster 4 ('Financially-Impacted Gen Xer') is also high in risk-taking propensity and perceived scarcity while low in introversion. Recommendations for this cluster should continue to emphasize precautions one can take while commuting on public transit and identifying risk levels of outdoor spaces. Since this cluster is more likely to have children, recommending safe activities that can be done as a family can encourage higher levels of compliance among multiple members within the same household. Cluster 4 also has moderate levels of empathy, which indicates that information related to the threat from the virus should emphasize both personal risk and risk to the larger community. In addition, to promote wellness in the relatively extraverted individuals in this cluster and their families, remote options for socializing should be emphasized, particularly for contacting those not essential to their families' social network and older or vulnerable adults at high risk for COVID-19 complications
5	Cluster 5 ('Work-Driven Young Adult') is lowest in both age and introversion, while also low in living space size and perceived scarcity. Since this cluster is most likely to go outside for only essential needs and very likely to live in urban or suburban settings, recommendations should emphasize how to mitigate the risk of spreading the virus on public transit (especially in urban environments). Due to the low levels of introversion and smaller living space size, guidelines for identifying risk in public spaces such as bars or restaurants can also help this cluster avoid spreading the virus
6	Cluster 6 ('Cautious Suburbanite') is high in introversion, age, and living space with moderate levels of empathy and risk-taking propensity. Messaging for this cluster should emphasize how the virus impacts both individual health and the community. Additionally, since this cluster has high levels of introversion and is more likely to live in a larger space, messages that give recommendations of activities that can be conducted indoors or on one's outdoor property could be effective in decreasing quarantine fatigue and increase the likelihood that they will minimize the amount of time spent outside
7	Cluster 7 ('Authority-Respecting Boomer') is high in age, empathy, and living space size while low in perceived scarcity and risk-taking propensity. Similar to Cluster 2, recommendations for this cluster should focus on mitigating quarantine fatigue since people in this cluster already exhibit high levels of adherence to social distancing measures. Messaging for this cluster should recommend activities that can be done in large living spaces or remote socializing and also emphasize the benefit that maintaining social distancing has on the health and wellbeing of their community

authority, reporting less trust towards organizations like the CDC. Cluster 2 ('Introverted Millennial') is a younger group that reports high adherence to social distancing guidelines, both in how frequently they engage in outside activities and attitudes towards the need to quarantine (however, they also report going outside occasionally for leisure). Cluster 3 ('Rural-leaning Millennial') is a younger group who appear to be the least empathetic and most likely to live in a rural area relative to the other clusters. Members of cluster 3 show engagement in high-risk activities and are more likely to agree with statements such as 'I feel that the social distancing measures have been excessive' and 'It's ok to go outside for non-essential trips as long as I'm careful'. Cluster 4 ('Financially-Impacted Gen Xer') shows similar non-compliance attitudes and risk-inclined behaviors, despite being older and more sensation-seeking than cluster 3. However, it is noteworthy that both clusters 3 and 4 are more likely to be classified as essential workers and report spending higher percentages of their working hours outside or surrounded by other people. They also report the highest agreement with the statement 'The Coronavirus has impacted me negatively from a financial point of view' compared to all other clusters. This difference is statistically significant as well. Cluster 5 ('Work-Driven Young Adult') is the youngest cluster and also the most likely to be working full-time (up to 93%). This cluster demonstrates high compliance with quarantine measures and reports going outside for leisure the least among the other groups. Cluster 6 ('Cautious Suburbanite') is an older group that perceives many outdoor activities as risky and mostly avoids the outdoors. Cluster 7 ('Authority-Respecting Boomer') is another older group but has the highest average income relative to the other clusters, with higher levels of social distancing compliance in both attitude and behavior. Cluster 7 is also the least likely to report that the Coronavirus has impacted them negatively from a financial point of view (24.5% vs. 75.3% and 79.8% reported by clusters 3 and 4, respectively).

Limitations

While samples obtained from MTurk are widely used in the social sciences [24], it is important to note that although relatively demographically representative, our sample is not nationally representative of the American response to COVID-19. While our sample size was sufficient to detect even subtle effects, all measures reported here were self-report, and thus, are susceptible to self-report biases. Further, given the remote nature of their employment, MTurk respondents are more likely to have more computer experience than the typical American population. Thus, our sample is likely underestimating the impact of dispositional and situational factors on psychological state and risk-taking during COVID-19.

Conclusions and implications for future research, policy, and practice

Our cluster model shows that groups exhibited a wide array of behaviors and attitudes towards the COVID-19 pandemic, indicating that underlying psychological and situational factors could drive variability in behavioral compliance with public health guidance. Our analysis of risk-taking behavior also identified two 'risk-inclined' groups (clusters 3 'Rural-leaning Millennial' and 4 'Financially-Impacted Gen Xer') that exhibited a higher propensity for engaging in pandemic-related high-risk activities. Members of both these groups were more likely to be conservative-leaning and reported high perceived scarcity of goods, suggesting a potential political agenda for their risk engagement and attitudes; however, these were distinct populations, which differed significantly in age, sensation-seeking, and region density. Upon further investigation, we found that these risk-inclined clusters were more likely to report circumstances related to economic inequities (employment, perceived scarcity, and limited living space), along with their higher engagement in high-risk activities. It is possible that engagement in high-risk activities was partially driven by the circumstances that these individuals have encountered and may not be solely due to demographic factors and psychological dispositions. This interpretation is also supported when examining risk-averse groups such as clusters 2 'Introverted Millennial' and 7 'Authority-Respecting Boomer', who avoided high-risk activities but were also less likely to be affected by these situational factors.

In general, since situational circumstances may make individuals more vulnerable to contracting COVID-19, and in many cases, are outside the direct control of these individuals, these factors may be intractable issues that significantly impair the ability of certain population groups to adhere to stay at home orders and other outbreak measures. For example, if clusters 3 and 4 are the most likely to be essential workers and have to work on a job-site, then it is not surprising that they also engage in more high-risk activities compared to clusters 2 and 7, who are less likely to work around people. Clusters 2 and 7 are also the least likely to report that the COVID-19 pandemic has impacted them financially, while clusters 3 and 4 are the most likely, which suggests that the groups who engage more frequently in non-compliant behavior are also the ones who are most vulnerable to the pandemic. While the CDC created guidelines for businesses owners and employers promoting in-person health checks, social distancing and PPE use in the work place, and building barriers to prevent virus spread to mitigate harm to employees [36], these guidelines only act as recommendations and does not guarantee full adherence across employers and work places. With the exception of companies subject to the Family and Medical Leave Act (FMLA), in the US, there are no federal legal requirements for paid leave in the workplace [37]. The lack of legal requirements for paid leave can add pressure to employees to attend the workplace despite exhibiting covid-related symptoms in fear of losing employment, which is supported by recent studies showing that states that received access to paid leave from the Families First Coronavirus Response Act (FFCRA) saw around 400 fewer confirmed cases per state per day [38]. People of color and women are also more likely to be excluded from paid sick leave [39], further driving increased susceptibility to coronavirus exposure. Based on these findings, the authors argue that stricter federal policies requiring paid sick leave are needed in order to provide further protection for vulnerable groups during the remainder of the COVID-19 pandemic and for future health crises.

In evaluating demographic comparisons in infection and social distance compliance by racial compositions, 'exposure' and situational factors should also be considered. For example, Black Americans who are more susceptible to contracting COVID-19 also have higher representation in clusters 3 and 4. It is also worth noting that Black Americans, Hispanics, and those lower in income and education are more likely to report vaccine hesitancy in the US [40] despite being more vulnerable to COVID-19. This suggests that communications about the vaccine are not addressing the specific concerns and apprehensions of these groups. While the profile analysis presented in this study was initially designed for quarantine compliance, these profiles could also be used to develop messaging related to vaccine uptake or alternatively, this approach could be applied to a more recent dataset specifically designed to investigate vaccine hesitancy. Previous work using data-driven personas has also employed qualitative methods to reveal additional insights for personas that overcome limitations from relying solely on survey data [41]. Follow-up work can use the COVID profiles from the current study as recruitment guidelines for qualitative interviews or focus groups to gain more descriptive accounts from each cluster related to their experience of the pandemic and ascertain views on other public health topics such as vaccine hesitancy and mask use.

As defined by Farmer et al. [42] in their discussion of how public health practitioners can address social determinants of disease, 'structural violence' occurs when political, economic, and cultural structures are organized in ways that put individuals and populations in harm's way. In terms of structural violence, it is likely that economic and political systems that existed before the COVID-19 pandemic still impact disparities in health and public health guidance adherence within this new context. It is also worth noting that the COVID-19 profiles clustered together people who may not be normally grouped together in traditional media discourse (e.g. Black Americans and Conservatives) since the clustering criteria were based on quarantine-related factors. Future research assessing responses to health-related issues should consider adopting data-driven user personas as presented in this paper to establish more context-relevant and nuanced characterizations of groups within a population. As demonstrated in the results section, this approach can assist practitioners when developing messaging and policies related to public health concerns while also addressing health inequities.

Ethical approval

Informed written consent from participants have been obtained prior to the commencement of the study. Ethics approval for this study was granted by Temple University (IRB 25899, approved 9 July 2019).

Acknowledgements

We thank the participants for their willingness to share their time, disclose their behavior and reflect on their responses to COVID-19 during these uncertain times. We also thank Bryana Walker whose conversations with the first author helped shape the conceptualization of the study during its early stages.

Disclosure statement

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

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Notes on contributors

Michael Robert Haupt is a PhD student in Cognitive Science at UC San Diego and a graduate research fellow at the Global Health Policy and Data Institute. His current research uses social network analysis and machine learning techniques to analyze communication dynamics on social media.

Staci Meredith Weiss is a postdoctoral research associate with the Perinatal Imaging Partnership at Cambridge University and received her PhD from Temple University. Her current work involves a global study of the context of pregnancy and infant parenting amidst COVID-19.

Michelle Chiu is a graduate student in the Cognition and Neuroscience PhD Program.

Raphael Cuomo is a public health research scientist with board certification in public health from the National Board of Public Health Examiners.

Jason M. Chein is a professor in the Department of Psychology at Temple University, and the Director of the Temple University Brain Research & Imaging Center (TUBRIC).

Tim Mackey is the Director of the Global Health Policy and Data Institute and an Associate Professor at UC San Diego. He has extensive professional experience, including consulting for the World Health Organization, the US Department of State and others.

Data availability statement

The authors have published all supplemental materials, data and syntax/code anonymously at the following link: https://osf.io/7w59e/. Please also see appendix for detailed descriptive supporting the results presented in this manuscript.

ORCID

Michael Robert Haupt http://orcid.org/0000-0003-4985-7796
Tim Mackey http://orcid.org/0000-0002-2191-7833

References

[1] Allen-Ebrahimian B. Timeline: the early days of China's coronavirus outbreak and cover-up. Axios. 2020 Mar 18. Available from: https://www.axios.com/timeline-the-early-days-of-chinas-coronavirus-outbreak-and-cover-up-ee65211a-afb6-4641-97b8-353718a5faab.html

[2] Dwyer C, Aubrey A. CDC now recommends Americans consider wearing cloth face coverings in public. NPR.Org. 2020 Apr 3. Available from: https://www.npr.org/sections/coronavirus-live-updates/2020/04/03/826219824/president-trump-says-cdc-now-recommends-americans-wear-cloth-masks-in-public

[3] Mervosh S, Lu D, Swales V. See which states and cities have told residents to stay at home. The New York Times. 2020 Mar 31. Available from: https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html

[4] Marcus J. Quarantine fatigue is real. The Atlantic. 2020 May 11. Available from: https://www.theatlantic.com/ideas/archive/2020/05/quarantine-fatigue-real-and-shaming-people-wont-help/611482/

[5] CDC. Coronavirus disease 2019 (COVID-19). Centers for Disease Control and Prevention; 2020 Feb 11. Available from: https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html

[6] Mays JC, Newman A. Virus is twice as deadly for black and Latino people than whites in N.Y.C. The New York Times. 2020 May 7. Available from: https://www.nytimes.com/2020/04/08/nyregion/coronavirus-race-deaths.html

[7] NYTimes Editorial Board. Opinion | 'You're on your own,' essential workers are being told. The New York Times. 2020 Apr 20. Available from: https://www.nytimes.com/2020/04/20/opinion/osha-coronavirus.html

[8] Gould E. Lack of paid sick days and large numbers of uninsured increase risks of spreading the coronavirus. Economic Policy Institute. 2020 Feb 28. Available from: https://www.epi.org/blog/lack-of-paid-sick-days-and-large-numbers-of-uninsured-increase-risks-of-spreading-the-coronavirus/

[9] Weintraub K. Lack of paid sick leave a coronavirus threat. WebMD. 2020 Mar 10. Available from: https://www.webmd.com/lung/news/20200310/lack-of-paid-sick-leave-a-coronavirus-threat

[10] Rozin P. What kind of empirical research should we publish, fund, and reward?: A different perspective. Perspect Psychol Sci. 2020;4(4):435–9. doi:10.1111/j.1745-6924.2009.01151.x

[11] Pruitt J, Grudin J. Personas: practice and theory. Proceedings of the 2003 Conference on Designing for User Experiences; 2003. p. 1–15. doi:10.1145/997078.997089

[12] Dantin U. Application of personas in user interface design for educational software. Proceedings of the 7th Australasian Conference on Computing Education, Vol. 42; 2005 Jan. p. 239–47.

[13] LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inf. 2013;82(11):e251–e268. doi:10.1016/j.ijmedinf.2011.03.006

[14] Liu T, Du J, Cai H, Farmer R, Jiang L, Bu F. Personas in O2O mobile healthcare: a method for identifying and creating user groups. 2016 IEEE 13th International Conference on E-Business Engineering (ICEBE); 2016. p. 100–7. doi:10.1109/ICEBE.2016.026

[15] Loitsch C, Weber G, Voegler J. Teaching accessibility with personas. In: Miesenberger K, Bühler C, Penaz P, editors. Computers helping people with special needs. Springer; 2016. p. 453–60. doi:10.1007/978-3-319-41264-1_62

[16] Cooper A. The inmates are running the asylum. Morgan Kaufman; 1999.

[17] Manning H, Temkin B, Belanger N. The power of design personas. Cambridge (MA): Forrester Research; 2003. p. 43.

[18] Miaskiewicz T, Kozar KA. Personas and user-centered design: how can personas benefit product design processes? Des Stud. 2011;32(5):417–30. doi:10.1016/j.destud.2011.03.003

[19] Miaskiewicz T, Sumner T, Kozar KA. A latent semantic analysis methodology for the identification and creation of personas. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2008. p. 1501–10. doi:10.1145/1357054.1357290

[20] McGinn J, Kotamraju N. Data-driven persona development. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2008. p. 1521–4. doi:10.1145/1357054.1357292

[21] Sinha R. Persona development for information-rich domains. CHI '03 Extended Abstracts on Human Factors in Computing Systems; 2003. p. 830–1. doi:10.1145/765891.766017

[22] An J, Kwak H, Jung S, Salminen J, Admad M, Jansen B. Imaginary people representing real numbers: generating personas from online social media data. ACM Trans Web. 2018;12(4):27, 1–26. doi:10.1145/3265986

[23] Zhang X, Brown H-F, Shankar A. Data-driven personas: constructing archetypal users with clickstreams and user telemetry. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; 2016. p. 5350–9. doi:10.1145/2858036.2858523

[24] Anderson CA, Allen JJ, Plante C, Quigley-McBride A, Lovett A, Rokkum JN. The MTurkification of social and personality psychology. Pers Soc Psychol Bull. 2019;45(6):842–50. doi:10.1177/0146167218798821

[25] U.S. Census Bureau. U.S. Census Bureau QuickFacts: United States; 2020 [cited 2020 May 17]. Available from: https://www.census.gov/quickfacts/fact/table/US/PST045218
[26] Benthin A, Severson H, Slovic P. A psychometric study of adolescent risk perception; 1993.
[27] John OP, Donahue EM, Kentle RL. The big five inventory: versions 4a and 54 [technical report]. Berkeley: University of California, Institute of Personality and Social Research; 1991.
[28] John OP, Naumann LP, Soto CJ. Paradigm shift to the integrative big five trait taxonomy. Handb Pers Theory Res. 2008;3(2):114–58.
[29] Rammstedt B, John OP. Measuring personality in one minute or less: a 10-item short version of the big five inventory in English and German. J Res Pers. 2007;41(1):203–12.
[30] Davis MH. Measuring individual differences in empathy: evidence for a multidimensional approach. J Pers Soc Psychol. 1983;44(1):113.
[31] Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. J Pers Soc Psychol. 1994;67(2):319.
[32] Gray JA. A critique of Eysenck's theory of personality. In: A model for personality. Berlin: Springer; 1981. p. 246–76.
[33] Gray JA. The psychology of fear and stress. New York: Cambridge University Press; 1987.
[34] Budescu DV. Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. Psychol Bull. 1993;114(3):542.
[35] Norušis MJ. IBM SPSS statistics 19 statistical procedures companion. Prentice Hall; 2011.
[36] CDC. COVID-19 archived webpage. Centers for Disease Control and Prevention; 2021 Mar 8. Available from: https://www.cdc.gov/coronavirus/2019-ncov/community/guidance-business-response.html
[37] U.S. Department of Labor. Sick leave | U.S. Department of Labor; n.d. [cited 2021 Nov 18]. Available from: https://www.dol.gov/general/topic/workhours/sickleave
[38] Pichler S, Wen K, Ziebarth NR. COVID-19 emergency sick leave has helped flatten the curve in the United States. Health Aff. 2020;39(12):2197–204. doi:10.1377/hlthaff.2020.00863
[39] Heymann J, Sprague A, Earle A, McCormack M, Waisath W, Raub A. US sick leave in global context: US eligibility rules widen inequalities despite readily available solutions: study examines US paid sick leave and compares it to sick leave in other countries. Health Aff. 2021;40(9):1501–9. doi:10.1377/hlthaff.2021.00731
[40] Khubchandani J, Sharma S, Price JH, Wiblishauser MJ, Sharma M, Webb FJ. COVID-19 vaccination hesitancy in the United States: A rapid national assessment. J Community Health. 2021;46(2):270–7. doi:10.1007/s10900-020-00958-x
[41] Miaskiewicz T, Luxmoore C. The use of data-driven personas to facilitate organizational adoption – a case study. Des J. 2017;20(3):357–74. doi:10.1080/14606925.2017.1301160
[42] Farmer PE, Nizeye B, Stulac S, Keshavjee S. Structural violence and clinical medicine. PLoS Med. 2006;3(10):e449.