Hierarchical Cluster Analysis on the distribution of the EQ-5D-5L in different groups of perinatal people

Authors: Pallavi Aytha Swathi,[1,4] Ning Yan Gu,[2] Annette K Regan[2,3]

Affiliations: [1] University of Colorado Anschutz Medical School; [2] School of Nursing and Health Professions, University of San Francisco; [3] UCLA Fielding School of Public Health; [4] College of Arts and Sciences, University of San Francisco

OBJECTIVE: Pregnant and postpartum persons are particularly susceptible to pandemic-induced anxiety/depression, which can adversely affect maternal and infant health. We sought to evaluate whether unbiased hierarchical cluster analysis could identify distinct factors impacting the Health-Related Quality-of-Life (HRQoL) measures during the perinatal period and to explore the distribution of EQ-5D-5L profiles.

METHODS: Individuals who were pregnant any time since January 2020 (i.e., the beginning of the pandemic) were invited to participate in a national online survey between May and June 2021 (n=3,359, EuroQol grant: 260-2020RA). Variables collected including respondents' personal experiences with the COVID-19 as well as the experiences or diagnoses of their family members, friends, and people they know in other social circles. HRQoL was measured by the EQ-5D-5L and other HRQoL instruments. We used unbiased hierarchical cluster analysis to define and characterize mutually exclusive groups.

The distributions of the EQ-5D-5L utilities and the EQ-VAS scores in each group were compared using the standard t-test.

RESULTS: Among the 3,359 pregnant and postpartum participants, 14.62% reported they had COVID-19 themselves, 14% reported that their partners had COVID-19, 28% reported that their close family members had COVID-19 and 39% were concerned about being pregnant during the pandemic. The hierarchical cluster analysis classified participants into 3 optimally distinct groups. Group 1 (n=971) consisted of people who were impacted by their partner and/or family's diagnosis with COVID-19. Group 2 (n=736) were distinguished by race and their social circle's hospitalizations with COVID-19. This group also showed they had other health conditions including depression, gestational diabetes, high blood pressure and others. Group 3 (n=1652) were impacted by their family's death due to COVID-19 and/or hospitalized with COVID-19. Although there was no strong evidence of clustering of EQ-5D-5L utility values, we did find that many participants in Group 2 had slightly lower utility values compared to other groups.

CONCLUSIONS: The results suggest that the EQ-5D-5L and EQ-VAS are superior in that they are less likely to generate artefactual clusters since they do not seem to be driven by the clusters. Statistical learning algorithms may allow for improved classification of factors impacting the pregnant and postpartum women during the pandemic. This classification can be replicated and validated in other prospective cohorts.

1. INTRODUCTION.

Health-related quality of life (HRQoL) is a holistic concept that aims to capture a range of health status indices. To date, the impact of multimorbidity on HRQoL has been investigated based on two general categories of multimorbidity: i) the number of chronic conditions (count definition) and ii) the cluster of chronic conditions (cluster definition) [6, 7]. Although HRQoL scores decrease with an increasing number of co-occurring chronic conditions [8], the full impact of multimorbidity on HRQoL is unlikely to be captured by the simple count method [9]. Meanwhile, some specific clusters of multimorbidity, such as the combination of mental and physical conditions [10], have been shown to have a notable effect on HRQoL. However, the impact of the different definitions of multimorbidity on HRQoL in a primary care setting is still unclear [8].

Comparing how the aforementioned categorizations effect the sociodemographic profile and health status (HRQoL) will help in improving the health care planning to match healthcare services with patients' needs. Therefore, using a large, nationally representative dataset, this study examined the performance of cluster definitions and the distributions of EQ-5D-5L and EQ-VAS in perinatal populations.

2. METHODS.

We conducted a national, online cross-sectional survey of pregnant or recent pregnant US people between May and June 2021. Individuals were invited to participate using targeted internet advertising that is based on adults' search history or social media activity that indicated pregnancy or presence of new born in the household. Participants were eligible if they were pregnant any time since January 2020 (i.e., beginning of SARS-CoV-2 activity in the US), if they were a resident in the US, and were 18 to 45 years of age. This age range was based on NCHS birth rates in the United States in 2020[11]. Representativeness of the survey was boosted using Quota-based sampling based on age, race, and US region of residence.

The survey included items on health related quality of life (HRQoL), prenatal care, social support, pregnancy history, risk behaviors including substance use, social and demographic information, zip code of residence and experiences with Covid-19 infection and vaccines. Individuals were asked to self-report information on COVID-19 diagnoses for themselves, their family, friends, and social circles. Participants who were positively diagnosed with COVID-19 were asked to rate the severity of the illness. Individuals also self-reported concerns they experienced for their health, the health of their baby and family members along with concerns about being pregnant during the pandemic. Rurality of residence based on 2013 NCHS urban-rural codes was determined using zip codes. Survey items asked respondents to report on a 5-point Likert scale with each concern, from ‘strongly agree’ to ‘strongly disagree’.

2.1. Hierarchical cluster analysis

Cluster analysis is a significant technique for classifying a ‘mountain’ of information into manageable, meaningful piles. It is a data reduction tool that creates subgroups that are more manageable than individual datum. It examines the full complement of inter-relationship between variables. In cluster analysis, it is not known which elements fit into which clusters. The data is reviewed to define the grouping or clusters.

Cluster analysis, like factor analysis, makes no distinction between dependent and independent variables. The entire set of interdependent relationships is examined. Cluster analysis is the obverse of the factor analysis. Factor analysis reduces the number of variables by grouping them into a smaller set of factors, but cluster analysis reduces the number of observations or cases by consolidating them into a smaller set of clusters. Hierarchical cluster analysis is the major statistical method for finding homogeneous groups of cases based on the measured characteristics [28]. It starts with each case as a separate cluster, i.e., there are as many clusters as cases, and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. The clustering method uses the dissimilarities or distances between objects when forming the clusters.

A summary of the clustering as a general process is as follows:

The distance is calculated between all initial clusters. In most analyses, individual cases will build up the initial clusters.

Then, the distances are calculated again following the fusion of the two most similar clusters.

Step 2 is done over repeatedly until all cases ultimately turn into one cluster.

Distance can be measured in a variety of ways [28].

The squared Euclidean distance has been applied most frequently. The Euclidean distance between two values is the arithmetic difference [28].

The squared Euclidean distance is applied more frequently than the simple Euclidean distance to impose gradually greater weight on objects that are further apart. To determine how distance is measured, it is necessary to select the clustering algorithm, namely, the rules governing which points distances are determined to specify cluster membership [28,29]. Euclidean distance coefficient specifies the distance between units; the greater distance implies making diverse managerial decisions.

In this study we used Agglomerative hierarchical clustering to get mutually exclusive groups. This is a "bottom-up" approach, each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. This method builds the hierarchy from the individual elements by progressively merging clusters [28]. The first step is to determine which elements to merge in a cluster. Usually, we want to take the two closest elements, according to the chosen distance. To do that, we need to take the distance between elements and therefore define the distance between two clusters [29]. In our analysis, there were a total of 534 variables in the initial dataset. For the sake of the analysis, we chose the variables that seem to drive the Health Related Quality of Life measures in theory. In general, variables like mask wearing preferences, experiences with COVID-19 like if they were diagnosed with COVID-19 or hospitalized with it, if their spouse or family member was diagnosed or hospitalized, COVID-19 and flu vaccination preferences, if the respondent had other co-morbidities like gestational diabetes, high blood pressure or depression and if they had an income change due to the pandemic. We also included demographic variables like age, race and location of the respondent.

2.3. Statistical analysis.

Post-stratification weights were applied to the sample data by age, race/ethnicity, and US census region of residence. We calculated the weights based on US natality data for 2016-2019. We used quantile regression to access the distribution of EQ-5D-5L and EQ-VAS among the various groups of clusters and compare the HRQoL characteristics.

3. RESULTS

In total, 6089 individuals responded to the advertisement of the study. Of these, 429 (7.0%) were ineligible. Of the 5660 eligible participants, 697 (12.1%) did not consent to participate in the survey. Of the 4973 participants who consented to participate, 3392 (68.2%) completed the survey. The characteristics of participants are provided in Table A1. The weighted sample characteristics in terms of maternal age, race/ethnicity, US census region of residence, and rurality of residences were similar to the US population of births. The EQ-5D-5L health states range from −0.148 for the worst (55555) to 0.949 for the best (11111) and the EQ-VAS range from a scale of 0 to 100. Among the participants, the median EQ-5D-5L utility score was 0.88 and the median EQ-VAS score was 80. Overall, 14.62% reported they had COVID-19 themselves, 14% reported that their partners had COVID-19, 28% reported that their close family members had COVID-19. About 39.5% were concerned about being pregnant during the pandemic and 55% reported concerns that COVID-19 is a severe disease. The number of respondents that reported strong concerns for each category of concerns they faced during the pandemic is shown in Figure A2.

We found that there were four distinct classes among the overall respondents as shown in a scatter plot in Figure1.

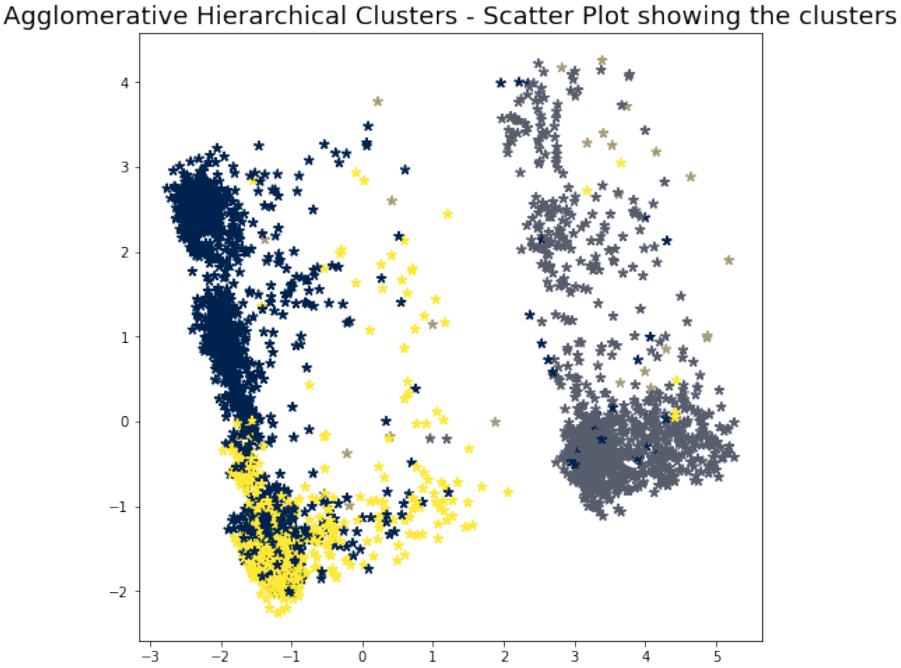

Figure 1: Scatter plot of the classes of the respondents.

Cluster 1 (n=1417), the biggest class consisted of people who did not report any major issues or concerns. Cluster 2 (n=940) were group were impacted by their family's death due to COVID-19 and/or hospitalized with COVID-19. Cluster 3 (n=26), the smallest group was distinguished by their own diagnosis of COVID-19 and if they were hospitalized during the pandemic. This group also were impacted by their social circle's hospitalizations with COVID-19 and showed they had other health conditions including depression, gestational diabetes, high blood pressure and others. The last group, Cluster 4(976) were categorized by their no one in their social circle being diagnosed with COVID-19 nor with anyone being hospitalized. Although we were able to get four distinct groups with each group showing an outcome related to the pandemic, in each of these clusters there was no evidence of EQ-5D-5L or EQ-VAS clustering or driving the clusters.

The optimal number of clusters was determined from the dendrogram as shown in Figure 2. The Y-axis has the Euclidean distances and X-axis has individual responses as an each work unit. We can note, as a first observation that the sizes of the clusters are really unbalanced.

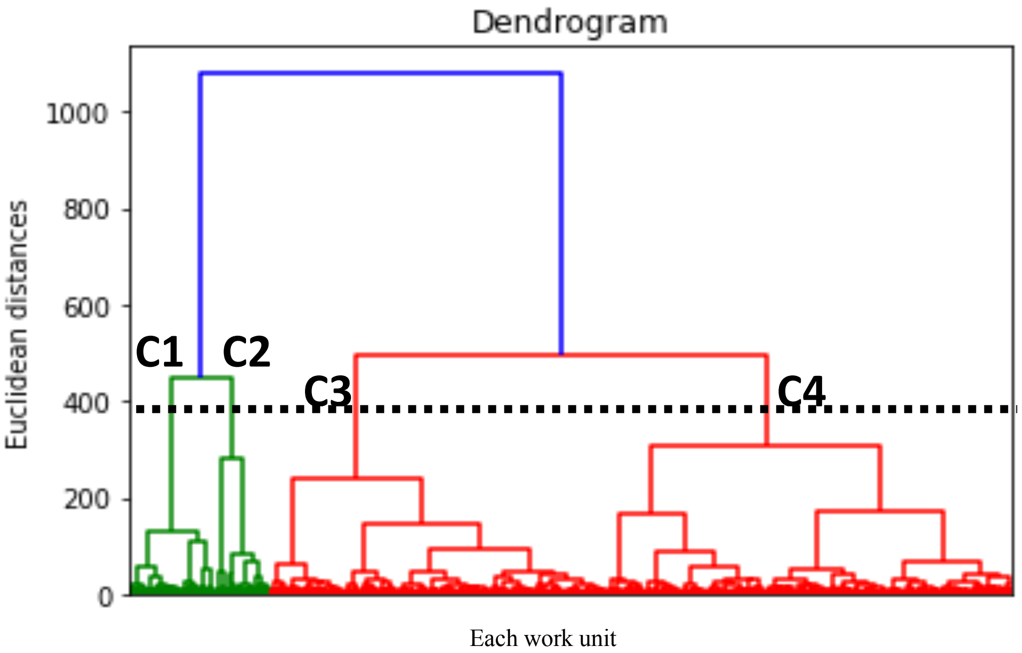

Figure 2: Dendrogram of hierarchical clustering showing 3 classes.

We can use a parallel coordinates plot to see how individual data points sit across all our variables. In the Figure 3, each color represents a different cluster. By looking at how the values for the variables compare across the clusters we can get a feel for what the clusters actually represent.

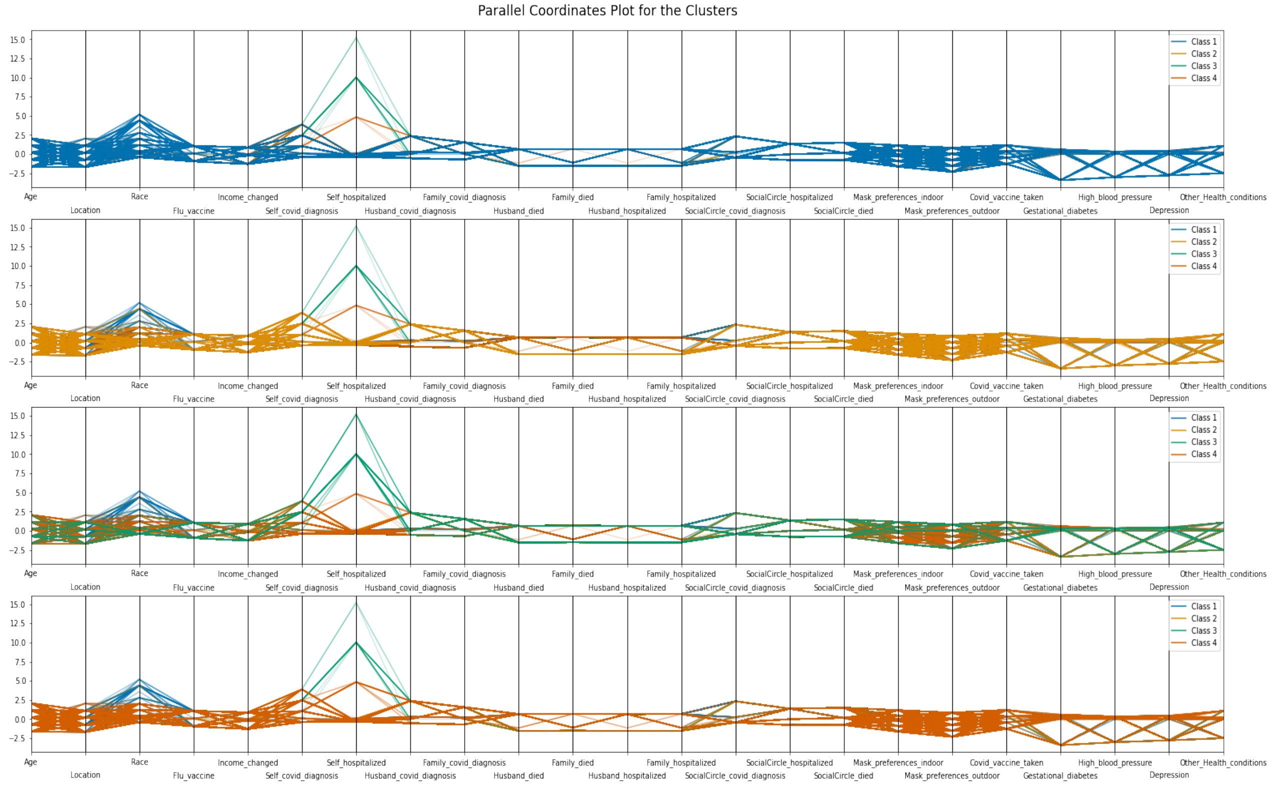

Figure 3: Parallel co-ordinates plot for each of the clusters. Each variable (as represented in x-axis) in the analysis has been scaled to have mean 0 and standard deviation of 1. The high and low values do not constitute any derivation in the cluster except those represent the actual values of the variables from the dataset.

The parallel co-ordinate plots are slightly confusing and seem to be overlapping. There is also a another useful piece of information coming out of the clustering: the centroids. We can now more clearly see the variation across the variables for each of the clusters. By looking at how the values for each variable compare across clusters, we can get a sense of what each cluster represents as shown in Figure 4.

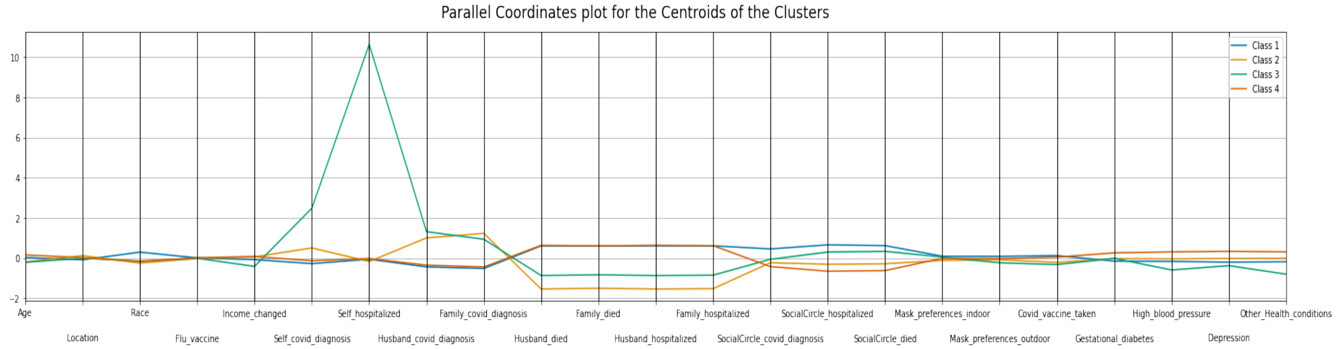

Figure 4: Parallel Co-ordinates for the Centroids. Each variable in the analysis has been scaled to have mean 0 and standard deviation of 1. The high and low values do not constitute any derivation in the cluster except those represent the actual values of the variables from the dataset.

Figure 4 represents for each cluster the average value (each variable has been scaled to have mean 0 and standard deviation of 1) of each of the initial variables.

From the plot, we can identify some groups here that are differentiating each clusters and are responsible for the respondents to be grouped in that cluster;

1. The first cluster (Class 1) seems to be consisting of all the respondents with no distinct variables or features differentiating from other clusters.

2. The second cluster (Class 2) seems to be driven very slightly, if the respondents' family members were diagnosed with COVID-19. It also consists of people whose family members were hospitalized during the pandemic and/or if they died.

3. The third cluster (Class 3) seems to be highly distinguished by respondents who have been self-diagnosed with COVID-19 and/or have been hospitalized during the pandemic. These class of people could have been highly impacted because on top of being pregnant and new parent, they were diagnosed with COVID-19.

4. Class 4 seems to be categorized by respondents who have had no one in their social circle diagnosed with COVID-19. These set of people have had positive experience during the pandemic.

Once we have these different classes, we can compare the EQ-5D-5L and EQ-VAS distributions in each of these clusters. The EQ-5D-5L and EQ-VAS values seem to be uniformly distributed in each of the clusters and do not drive the clusters. We should keep in mind that the sizes of classes are not equal.

The EQ-5D-5L distribution in each of the clusters is shown in Figure 5.

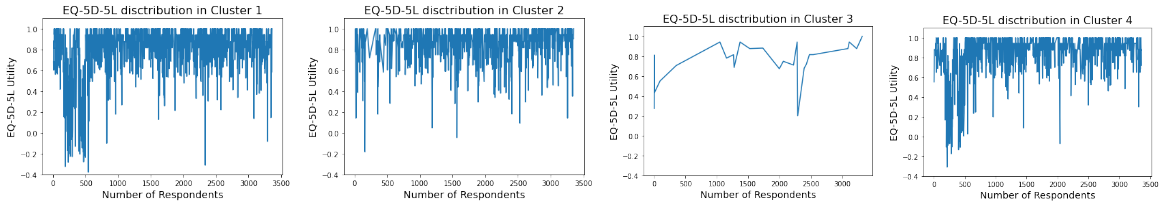

Figure 5: EQ-5D-5L distribution in each of the clusters. X-axis represents each unit and y-axis represents the EQ-5D-5L utility values. Cluster 1 consists of 1417 observations, Cluster 2 consists of 940 observations, Cluster 3 consists of 26 observations and Cluster 4 consists of 976 observations.

The EQ-VAS distribution in each of the cluster is shown in Figure 6.

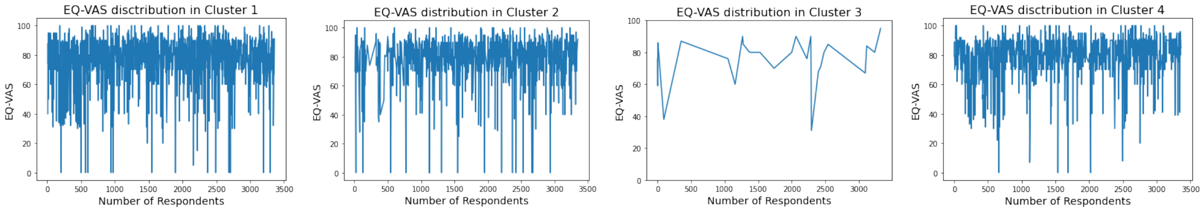

Figure 6: EQ-5D-5L distribution in each of the clusters. X-axis represents each unit and y-axis represents the EQ-VAS values. Cluster 1 consists of 1417 observations, Cluster 2 consists of 940 observations, Cluster 3 consists of 26 observations and Cluster 4 consists of 976 observations.

As seen from Figure 5 and Figure 6, the EQ-5D-5L and EQ-VAS distributions seem to be uniform in both the clusters although Cluster 1 and Cluster 4 has a slightly lower EQ-5D-5L utility values. There is no strong evidence of clustering of EQ-5D-5L utility values nor of EQ-VAS values.

4. DISCUSSION

This study aimed at identifying different subgroups of a population based on the health behaviors of pregnant people during the pandemic like COVID-19 diagnosis, mask wearing preferences, their family and social circles experiences with COVID-19, while also taking the influence of the parameters race, age and location into account along with the other health conditions. Four distinct classes were identified. All four classes show a unique pattern regarding the health behaviors. Class 1 and Class 4 represent a healthy cluster, showing a very healthy pattern and the highest item probabilities for healthy behavior categories. Class 2 and class 3 on the other hand show more unhealthy profiles in the sense that the respondents, their family and friends seem to be affected by COVID-19.

Cluster analysis is very exploratory and although comparisons with other studies are difficult because of different investigated health behaviors and methodological approaches, our results are in line with similar investigations. In line with previous studies, we identified an overall healthy cluster with class 1 and class 4 [1, 2, 4, 12, 13, 14, 15]. Similar to previous studies, we observed a clustering of behavior with people having very traumatic experiences with the pandemic[3, 4, 12,16]. This clustering becomes very evident for class 3 and partially for class 2 considering the fact that this class had respondents' family's experiences with COVID-19.

Scientific evidence on associations between clustering of health behaviors and HRQOL is sparse. Conry et al. [4] report a tendency for healthier clusters having a better quality of life. This result could not be replicated by our study. We found no clear association between a healthier behavior pattern and better physical or mental HRQOL. The differences in EQ-5D-5L and EQ-VAS between the clusters are too small to be considered as clinically relevant. This suggests that the EQ-5D-5L and EQ-VAS are superior in that they are less likely to generate artefactual clusters since they do not seem to be driven by the clusters.

4.1. Strengths and Weaknesses.

This study had several strengths and weaknesses. First, we collected data from a large, diverse national sample of pregnancies in the months following a SARS-CoV-2 infection. We were able to gather a range of information related to health, pregnancy and beliefs and concerns, while allowed us to comprehensively access the experiences of perinatal individuals and the impact on Health Related Quality of Life (HRQoL). The clusters is based on a hierarchical clustering model. This approach offers a clustering based on a statistical model instead of more arbitrary cluster criteria and thus might be more sophisticated than traditional clustering approaches [17]. Moreover, this methodological approach allows introducing covariates to factor in their influence on health behavior patterns. Another strength of this study is that the clustering is not only based on dichotomous variables like the absence or presence of an experience but also on polytomous variables. Also addressing the relevance of health behavior clusters by linking them to HRQoL outcomes like EQ-5D-5L and EQ-VAS, a clinically important outcome, adds value to this study.

Despite all this, our study has several limitations. One problem lies in the way the health behaviors are measured and operationalized. All information on health behaviors is self-reported and thus prone to information bias like recall-bias or social desirability-bias. While previous research has shown that self-report is a valid estimate while using the EQ-5D-5L scale and EQ-VAS[18], it can still be subjected to self-reporting bias. Further, our sample included only pregnancies from every state in the US, we have made efforts to ensure generalizability of our results, our sample was still restricted to those who interacted with the ads on social media and with internet access. This is why we cannot exclude selection bias in our sample. This could make our results specific. In large cohort studies with many variables like this one, we had to balance the tradeoff between accuracy and feasibility. Another limitation of this study concerns the reduced sample size this study that might result in a biased depiction of HRQoL. Taking this into account, the observed changes in physical and mental HRQoL might not necessarily reveal a true change on a population level. Although we adjusted our analyses for several variables, chances are high that HRQoL might have been influenced by a factor we did not adjust for, e.g., socio-economic status and prenatal care facilities provided to these pregnant people. Therefore, residual confounding cannot be ruled out.

5. CONCLUSIONS.

In conclusion, this study identified distinct patterns of health behaviors within a large population-based sample. The observed health behavior patterns and the socio-demographic characteristics of the identified clusters are in line with the few other existing international studies. Knowledge on specific clusters which are common in an perinatal population are an important step for comprehensive health promoting public health policies. The clustering of lifestyle factors like health behaviors can give valuable information on characteristics of target groups for primary preventions. The results also suggest that the EQ-5D-5L and EQ-VAS do not seem to be driven by the clusters and are less likely to generate clusters. Statistical learning algorithms may allow for improved classification of factors impacting the pregnant and postpartum people during the pandemic. This classification can be replicated and validated in other prospective cohorts. Further research should focus on linking identified clusters to important medical outcomes along with the EQ-5D-5L and EQ-VAS in order to identify vulnerable groups and to allow for individualized patient-centered primary prevention programs.

6. ACKNOWLEDGEMENTS

This work was supported by a EuroQoL Research Foundation grant #: 260-2020A (PI: Dr. Annette Regan). Thanks you to the EuroQoL Research Foundation and whole the research team at University of San Francisco.

7. REFERENCES.

1. Schuit AJ, Van Loon AJM, Tijhuis M, Ocké MC. Clustering of lifestyle risk factors in a general adult population. Prevent Med. (2002) 35:219–24. doi: 10.1006/pmed.2002.1064.

2. Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. Prevent Med. (2007) 44:124–8. doi: 10.1016/j.ypmed.2006.10.006.

3. Mawditt C, Sacker A, Britton A, Kelly Y, Cable N. The clustering of health-related behaviours in a British population sample: Testing for cohort differences. Prev Med. (2016) 88:95–107. doi: 10.1016/j.ypmed.2016.03.003

4. Conry MC, Morgan K, Curry P, Mcgee H, Harrington J, Ward M, et al. The clustering of health behaviours in Ireland and their relationship with mental health, self-rated health and quality of life. BMC Public Health (2011) 11:692. doi: 10.1186/1471-2458-11-692.

5. Zolfaghari F, Khosravi H, Shahriyari A, Jabbari M, Abolhasani A. Hierarchical cluster analysis to identify the homogeneous desertification management units. PLoS One. 2019;14(12):e0226355. Published 2019 Dec 18. doi:10.1371/journal.pone.0226355

6. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, Meinow B, Fratiglioni L. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev. 2011;10:430–439. doi: 10.1016/j.arr.2011.03.003.

7. Le Reste JY, Nabbe P, Manceau B, Lygidakis C, Doerr C, Lingner H, Czachowski S, Munoz M, Argyriadou S, Claveria A, et al. The European General Practice Research Network presents a comprehensive definition of multimorbidity in family medicine and long term care, following a systematic review of relevant literature. J Am Med Dir Assoc. 2013;14:319–325. doi: 10.1016/j.jamda.2013.01.001.

8. Fortin M, Lapointe L, Hudon C, Vanasse A, Ntetu AL, Maltais D. Multimorbidity and quality of life in primary care: a systematic review. Health Qual Life Outcomes. 2004;2:51. doi: 10.1186/1477-7525-2-51.

9. Wei MY, Kawachi I, Okereke OI, Mukamal KJ. Diverse Cumulative Impact of Chronic Diseases on Physical Health-Related Quality of Life: Implications for a Measure of Multimorbidity. Am J Epidemiol. 2016;184:357–365. doi: 10.1093/aje/kwv456. [PMC free article]

10. Kolappa K, Hendersona DC, Kishoreb SP. No physical health without mental health: lessons unlearned? Bull World Health Organ. 2013;91:3–3A. doi: 10.2471/BLT.12.115063. [PMC free article]

11. Martin JA, Hamilton BE, Osterman MJK. Births in the United States, 2020. NCHS Data Brief, no 418. Hyattsville, MD: National Center for Health Statistics. 2021. DOI: https://dx.doi.org/10.15620/cdc:109213.

12. Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. Prev Med. (2015) 81:16–41. doi: 10.1016/j.ypmed.2015.07.003.

13. De Vries H, Van 'T Riet J, Spigt M, Metsemakers J, Van Den Akker M, Vermunt JK, et al. Clusters of lifestyle behaviors: results from the Dutch SMILE study. Prev Med. (2008) 46:203–8. doi: 10.1016/j.ypmed.2007.08.005.

14. Schneider S, Huy C, Schuessler M, Diehl K, Schwarz S. Optimising lifestyle interventions: identification of health behaviour patterns by cluster analysis in a German 50+ survey. Eur J Public Health (2009) 19:271–7. doi: 10.1093/eurpub/ckn144.

15. Dodd LJ, Al-Nakeeb Y, Nevill A, Forshaw MJ. Lifestyle risk factors of students: a cluster analytical approach. Prevent Med. (2010) 51:73–7. doi: 10.1016/j.ypmed.2010.04.005.

16. Meader N, King K, Moe-Byrne T, Wright K, Graham H, Petticrew M, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. BMC Public Health (2016) 16:657. doi: 10.1186/s12889-016-3373-6.

17. Vermunt JK, Magidson J. Latent class cluster analysis. In: Mccutcheon AL, Hagenaars JA, editors. Applied Latent Class Analysis. Cambridge: Cambridge University Press (2002) 89–106. doi: 10.1017/CBO9780511499531.004.

18. Caparros-Gonzalez R.A., Luque-Fernández M.Á. Mental health in the perinatal period and maternal stress during the Covid-19 pandemic: Influence on fetal development. Rev. Esp. Salud Publica. 2020;94:e1–e2.

19. Milte CM, Thorpe MG, Crawford D, Ball K, Mcnaughton SA. Associations of diet quality with health-related quality of life in older Australian men and women. Exp Gerontol. (2015) 64:8–16. doi: 10.1016/j.exger.2015.01.047

20. Rabel M, Meisinger C, Peters A, Holle R, Laxy M. The longitudinal association between change in physical activity, weight, and health-related quality of life: results from the population-based KORA S4/F4/FF4 cohort study. PLoS ONE (2017) 12:e0185205. doi: 10.1371/journal.pone.0185205.

21. Dumuid D, Olds T, Lewis LK, Martin-Fernández JA, Katzmarzyk PT, Barreira T, et al. Health-related quality of life and lifestyle behavior clusters in school-aged children from 12 countries. J Pediatr. (2017) 183:178–83.e172. doi: 10.1016/j.jpeds.2016.12.048.

22. Li C, Ford ES, Mokdad AH, Jiles R, Giles WH. Clustering of multiple healthy lifestyle habits and health-related quality of life among US adults with diabetes. Diabetes Care (2007) 30:1770–6. doi: 10.2337/dc06-2571

23. Cockerham WC. Lifestyles, social class, demographic characteristics and health behavior. In: Gochman DS, editor. Handbook of Health Behavior Research I. Personal and Social Determinants. New York, NY: Plenum Press (1997). p. 253–65.

24. Kongsted A, Nielsen AM. Latent Class Analysis in health research. J Physiother (2017) 63:55–8. doi: 10.1016/j.jphys.2016.05.018.

25. Rabel M, Laxy M, Thorand B, Peters A, Schwettmann L and Mess F (2019) Clustering of Health-Related Behavior Patterns and Demographics. Results From the Population-Based KORA S4/F4 Cohort Study. Front. Public Health 6:387. doi: 10.3389/fpubh.2018.00387.

26. Ware JJr, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care (1996) 34:220–33. doi: 10.1097/00005650-199603000-00003

27. Schneider S, Huy C, Schuessler M, Diehl K, Schwarz S. Optimising lifestyle interventions: identification of health behaviour patterns by cluster analysis in a German 50+ survey. Eur J Public Health (2009) 19:271–7. doi: 10.1093/eurpub/ckn144.

28. Zhang Z, Murtagh F, Van Poucke S, Lin S, Lan P (2017). Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R. Ann Transl Med, 5(4):75 10.21037/atm.2017.02.05

29. Varouchakis E A, Theodoridou P G, Karatzas G P (2019). Spatiotemporal geostatistical modeling of groundwater levels under a Bayesian framework using means of physical background. Journal of Hydrology, 10.1016/j.jhydrol.2019.05.055

30. Sadeghravesh M H, Khosravi H, Ghasemian S (2016). Assessment of combating-desertification strategies using the linear assignment method. Solid Earth, 7: 673–683. 10.5194/se-7-673-2016

Dodd LJ, Al-Nakeeb Y, Nevill A, Forshaw MJ. Lifestyle risk factors of students: a cluster analytical approach. Prevent Med. (2010) 51:73–7. doi: 10.1016/j.ypmed.2010.04.005

31. Warkentin LM, Majumdar SR, Johnson JA, Agborsangaya CB, Rueda-Clausen CF, Sharma AM, et al. Weight loss required by the severely obese to achieve clinically important differences in health-related quality of life: two-year prospective cohort study. BMC Med. (2014) 12:1–9. doi: 10.1186/s12916-014-0175-5.

32. Fleary SA, Nigg CR. Trends in health behavior patterns among US adults, 2003–2015. Ann Behav Med. (2019) 53:1–15. doi: 10.1093/abm/kay010.

33. Burgard SA, Lin KYP, Segal BD, Elliott MR, Seelye S. Stability and change in health behavior profiles of U.S. adults. J Gerontol Series B (2018) doi: 10.1093/geronb/gby088.

34. R Core Development Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing (2010).

35. Ward Joe H (1963). Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association, 58 (301): 236–244. 10.2307/2282967.

8. APPEXDIX

Figure A1: Box Plot of each of the variable used in the analysis grouped by cluster

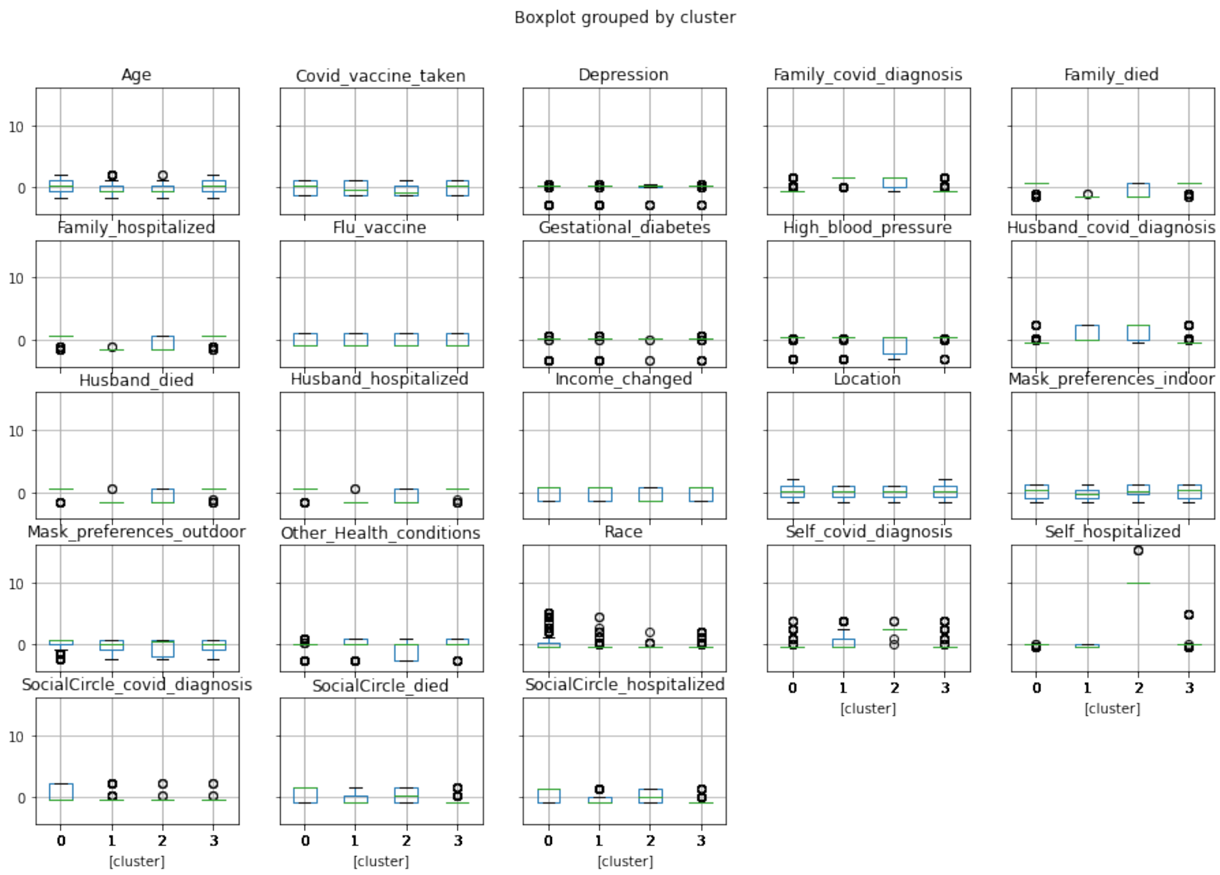

Figure A2: Strong concerns expressed by respondents. More number of respondents expressed concerns about COVID-19 being a serious disease followed by concerns about being pregnant in the pandemic. Least number of respondents expressed concern about their own health.

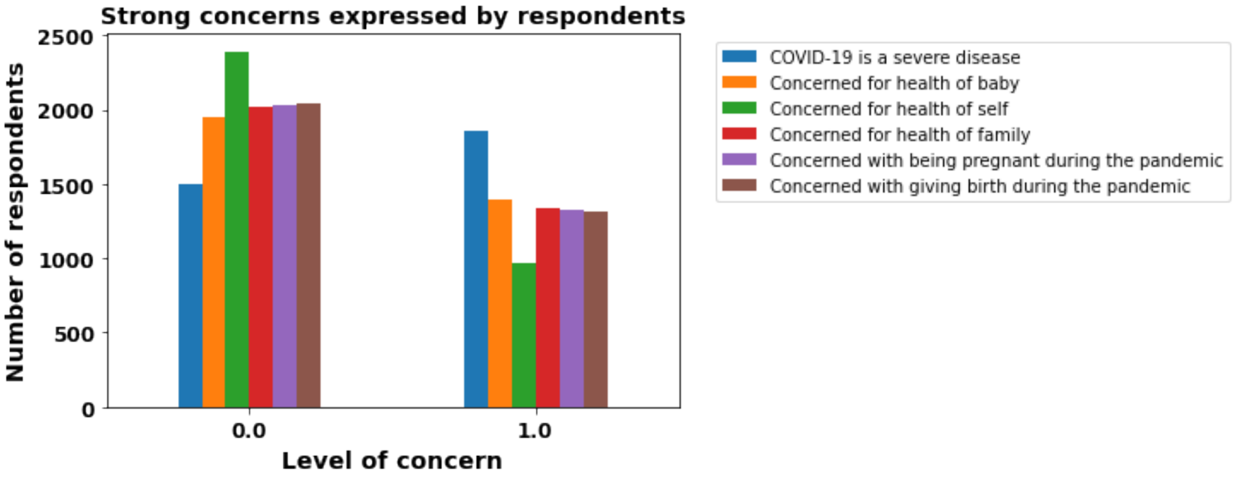

Figure A3: Location of pregnant and recently pregnant participants - May to July 2021

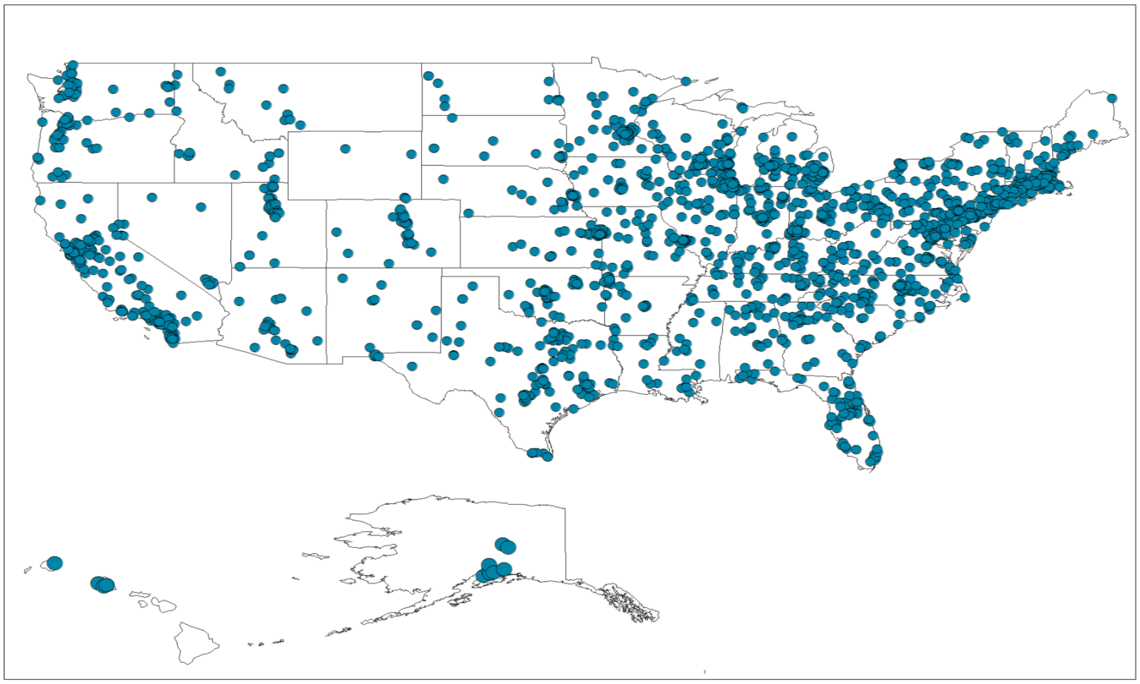

Table A1. Sociodemographic characteristics of participating pregnant and postpartum persons compared to 2016-2019 US birth statistics.

Characteristic	2016-2019 Births*	Sample
	N (%)	Weighted % (95% CI)
Total	3,695,063 (100)	2,213 (100)
Maternal age		
18-24 years	834,935 (22.6)	18.7 (16.2, 21.2)
25-29 years	1,078,097 (29.2)	31.2 (28.8, 33.5)
30-34 years	1,089,281 (29.5)	32.5 (30.1, 34.9)
35-39 years	572,598 (15.5)	15.0 (13.2, 16.7)
40-45 years	120,152 (3.2)	2.6 (1.9, 3.4)
Race/ethnicity		
White	1,921,589 (52.0)	51.1 (48.5, 53.8)
Hispanic or Latino/a/x	866,715 (23.4)	24.8 (22.1, 27.5)
Black	541,719 (14.7)	14.3 (12.0, 16.7)
Asian	244,034 (6.6)	8.3 (6.6, 10.0)
American Indian/Alaskan Native	28,109 (0.8)	0.6 (0.2, 1.0)
Native Hawaiian/Pacific Islander	10,026 (0.3)	0.1 (0.0, 0.2)
Multiple races	82,871 (2.2)	0.8 (0.1, 1.5)
Region of residence		
Northeast	591,148 (16.0)	16.1 (14.2, 17.9)
Midwest	774,826 (21.0)	20.9 (18.9, 22.9)
South	1,453,977 (39.3)	38.6 (36.0, 41.3)
West	875,112 (23.7)	24.1 (21.9, 26.3)
US Territory	--	0.3 (0.1, 0.5)
Rurality of residence		
Large central metropolitan	1,195,996 (32.4)	25.0 (22.5, 27.5)
Large fringe metropolitan	891,525 (24.1)	20.8 (18.6, 22.9)
Medium metropolitan	779,469 (21.1)	25.4 (23.2, 27.6)
Small metropolitan	330,240 (8.9)	11.6 (10.0, 13.2)
Micropolitan (Nonmetropolitan)	299,130 (8.1)	10.7 (9.1, 12.2)
Noncore (Nonmetropolitan)	198,703 (5.4)	6.5 (5.4 ,7.7)

*DATA SOURCE: United States Department of Health and Human Services (US DHHS), Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), Division of Vital Statistics, Natality public-use data 2016-2019, on CDC WONDER Online Database, October 2020. Accessed at http://wonder.cdc.gov/natality-expanded-current.html on Sep 21, 2021, 11:09:23 PM. We included births to pregnant persons aged 18-45 years (consistent with our eligibility criteria).