Shawn Stewart
Final Project

Exploring the relationship between number of comorbid conditions and COVID-19 protective health behaviors

Abstract
The COVID-19 pandemic has prompted a flurry of research into how different groups of people behave and what protective behaviors they take to reduce the spread of the virus. Burford et al. did a survey using convenience sampling and used logistic regression to calculate the odds ratios for a variety of demographics, using different protective behaviors as binary dependent variables. While the survey collected nuanced data, Burford et al. collapsed this data into aggregated categories for analysis. One of these categories looks at the number of comorbid conditions respondents have. Burford et al. ran their regression with comorbid conditions set as a binary variable, measuring whether people had zero comorbid conditions or one or more. I took their dataset and recategorized the counts of comorbid conditions to create two new variables, a factor with four levels (zero, one, two, or three or more comorbid conditions) and a factor with three levels (zero, one, and two or more). My logistic regression results show that there is evidence that the number of comorbidities, not just the presence or absence of any number of comorbidities, can affect the regression results and odds ratios.

Introduction
The COVID-19 pandemic has shed light on how different people react to health guidelines. The University of Texas Health Science Center at Houston began researching characteristics associated with protective health behaviors and reasons for leaving the home in order to better inform policy development and public health communication campaigns (Burford et al.). Although Burford et al. began this study when personal behaviors were one of the only ways to curb the spread of the virus, this investigation remains relevant, as social distancing and other personal measures remain useful tools to limit outbreaks of COVD-19 variants.

Burford et al. designed a survey to study how urbanicity and demographic characteristics affect individual's likelihood of engaging in certain behaviors during the COVID-19 pandemic, with the hope that these results can be used to target the least compliant groups with communications campaigns. They looked at two types of behaviors – reasons for leaving home, and protective behaviors. In the survey, they gave people a drop-down option for "reasons for leaving home" with the most common reasons they wanted to study and included an "other" category where the individual could fill in a text field with other reasons. The protective health behaviors the authors looked at were social distancing, hand washing, using hand sanitizer, using a mask, spending time inside the home, using disinfectant wipes, and wearing gloves.

For the predictor variables, they defined "urbanicity" by using the respondent's zip code and using the Department of Health and Human Services definition of urbanicity based on population density. Other demographics information included sex, race, age, education, income, children in the household, depressive symptom severity, and comorbidity. In this extension, I will look at the comorbidity category in more detail. While the survey collected detailed information about each of these categories, the authors aggregated the data into broad categories for their logistic regression, and by doing so made an assumption that the subgroups within the variable are homogenous.

Burford et al. did not explicitly state a hypothesis in their article. Rather, their study was exploratory in nature, with a focus on evaluating how urbanicity and various demographic characteristics correlate with reasons for the leaving home and protective behaviors used during the pandemic. The unspoken hypothesis is that urbanity, demographic characteristics, and people's behaviors are correlated, and that the correlations can be leveraged to identify less compliant groups of people.

Literature review and theory

Before basing further research on Burford et al.'s findings, it is important to verify the robustness of their results. While Burford et al. chose to make comorbidity a binary variable, other studies have broken this type of variable into categories instead, allowing for more granular analysis. In addition, people with multiple comorbid conditions may behave differently than people with only one comorbid condition.

Bish and Michie (2010) reviewed a number of studies and found that perceived susceptibility was positively associated with carrying out protective behaviors. During the COVID-19 pandemic the CDC has publicized that people with certain comorbid conditions may be more susceptible to developing severe cases of COVID-19 (CDC 2021). It is possible that people with multiple comorbid conditions may perceive themselves are more susceptible to severe COVID-19 and as a result may take more precautions than people with a single comorbid condition or with no comorbid conditions. While Burford et al.'s analysis lumps the groups together, the category of "people with one or more comorbid conditions" may not be homogenous.. If the group is actually heterogenous, split along lines of "number of comorbidities", this could affect the analysis.

O'Conor et al. also studied protective health behaviors and individual characteristics, and in their analysis they split "chronic health conditions" into categories of "one to two" or "three or more." In their analysis, these categories had statistically significant results, suggesting that health conditions may be a good predictor for some behaviors, and the two categories had somewhat different results, suggesting the number of health conditions may affect behavior.

Hypothesis

H_0 - Running the logistic models with the "comorbidity" category divided into additional subcategories will result in the same statistically significant results as when running the logistic model with a binary comorbidity category.

H_1 – Running the logistic models with the "comorbidity" category divided into additional subcategories will result in different statistically significant results than when running the logistic model with a binary comorbidity category.

Research Design

Burford et al. collected a survey via social media using convenience sampling. The researchers recruited participants through their own social media accounts, and since the authors are associated with The University of Texas Health Science Center at Houston, the respondents skew towards Texans (although other states are also represented). The unit of observation is individual people. The article attempts to discern how changes in urbanicity and demographics would change the odds of a single person participating in certain behaviors during the pandemic.

The authors point out the importance of understanding how individual characteristics are associated with behavioral decisions during the pandemic, as individual behaviors are one of the best methods for containing the spread of the virus. They argue that understanding how behaviors differ across different urbanicity and demographics can inform the development of policies and public health communications. However, given their convenience sampling method and skewed sample, the results that can be drawn directly from this study are limited. The results may direct future research, but the generalizability of the results is limited by the lack of a representative sample.

The online survey generated 2441 complete responses. Using this cross-sectional data, the authors first removed individuals with certain characteristics that might skew the results. They removed individuals who self-reported as essential workers, as essential workers might need to leave the house more and engage in protective behaviors differently than the rest of the population. If the respondent did not include a zip code, they were excluded as urbanicity could not be determined. If the respondent reported there was a stay-at-home order in place where they lived, they were excluded as this study is examining people's choices, and legal orders in place would limit people's choices. Similarly, anyone currently sick with COVID-19 was excluded, as they would have to quarantine, and individuals who reported they had a condition that prevents them from running errands were dropped from the dataset. After filtering out these conditions, the study was left with 1374 responses to analyze.

A summary of the survey respondents' characteristics shows that 72% of the respondents were female, 79% were white, and 85% of the respondents were college graduates, with 64% living in Texas. We can see in Table 1 that all categories have more than 130 respondents (except for the "missing" categories), so even though the sample is skewed towards white female college graduates, there are sufficient observations in each category to use for statistical analysis. A very small number of respondents in a single category would limit the statistical power of the analysis. The missing data is important, as the authors allow their R script to use the default listwise deletion. Although n=1374, the actual number of observations in the logistic models is smaller.

Table 1: Participant characteristics of the sample (n=1374)

Characteristic	Total % (n)
Urbanicity	
Urban	43.2 (593)
Suburban	33.3 (458)
Rural	22.0 (302)
Missing	1.5 (21)
Sex	
Male	28.0 (385)
Female	71.7 (985)
Missing	0.3 (4)
Race	
Non-Hispanic white	78.8 (1083)
Non-white	18.6 (256)
Missing	2.5 (35)
Age	
18 – 34	30.1 (414)
35 – 49	39.2 (539)
50 and older	29.9 (411)
Missing	0.7 (10)
Education	
Not a college graduate	15.3 (210)
College graduate or more	84.6 (1162)
Missing	0.1 (2)
Income	
<$50,000	10.1 (139)
$50,000 – $100,000	24.5 (336)
$100,000 - $150,000	24.7 (339)
Over $150,000	38.4 (537)
Missing	2.4 (33)
Children	
No	56.3 (774)
Yes	42.3 (581)
Missing	1.4 (19)
Depressive Symptom Severity	
None/mild	73.1 (1004)
Moderate to severe	20.1 (285)
Missing	6.2 (85)
Comorbidity	
None	58.7 (807)
1 or more	41.3 (567)
Missing	-

The authors created a logistic model for each dependent variable of interest. For example, there is a logistic model where the outcome variable is a binary variable measuring whether people use hand sanitizer, and there is another where the outcome variable is whether the respondents use a mask. Burford et al. then used the regression output to calculate the odds ratios and 95%

confidence interval for each predictor variable. Since each logistic model includes the same predictors, each model is "mutually adjusted for all other variables" (Burford et al. 2020). Using odds ratios allows for a substantive interpretation of the results. The authors also checked for multicollinearity by calculating the VIF for each model. All VIF results were between 1 -2, indicating multicollinearity is not an issue.

While the authors included a full set of demographics questions and gathered respondent's zip codes to determine urbanicity, in this extension I am primarily concerned with the "comorbidity" variable. The survey used conditional logic to guide respondents through the questions. The detailed questions about comorbidity were hidden behind a screening question which asks respondents to indicate whether a health professional has ever told them they have certain health conditions (and lists out all of the conditions the authors consider "comorbid conditions.") If the respondent answers "yes", they get an additional question asking which health conditions a health professional has told them they have.

I looked at the comorbid conditions data in more detail to determine where it makes sense to "split" the data. O'Conor et al. used a four-way split, dividing respondents into categories of none, one, two, or three or more. Looking at a table showing the number of respondents in each category, this seems reasonable for the Burford et al. dataset. There are more people in the first three categories, but the "three or more" categories still has over 60 respondents. Traditionally thirty or more observations is needed for most statistical tests, so this number of respondents seems sufficient. However, it would be better to have more respondents.

Table 2: Counts of Comorbid Conditions and Their Frequency (N= 1374)

# of Comorbid Conditions	Total % (n)
Zero	59 (808)
One	26 (358)
Two	10 (147)
Three or more	4 (61)

I also looked at the boxplot and barplot for the variable to visualize the spread of the data.

Figure 1: Boxplot of Counts of Comorbidity Variable with Four Levels

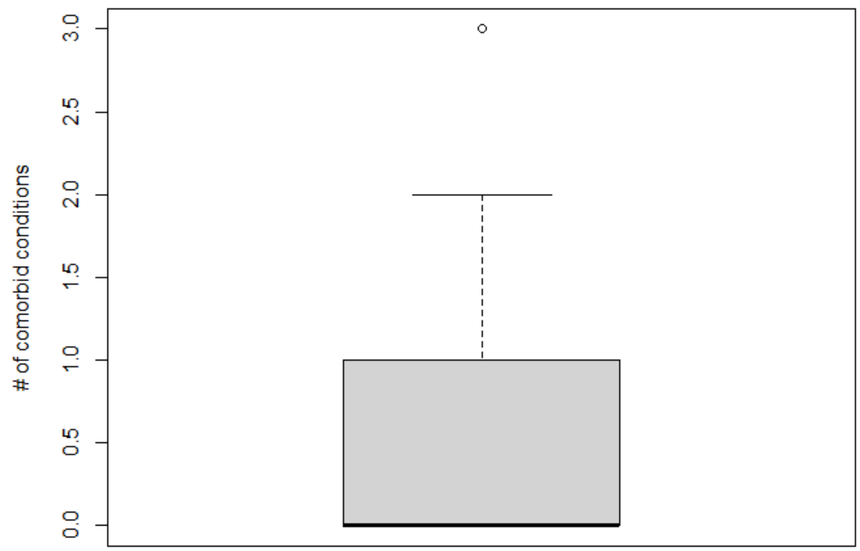
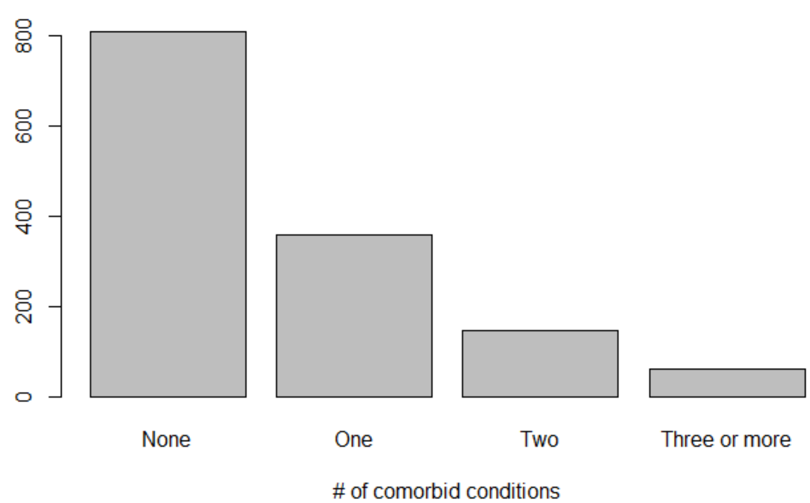


Figure 2: Barplot of Counts of Comorbidity Variable with Four Levels



The visualizations show the data is heavily skewed towards the right (zero comorbid conditions) and the data becomes very thin towards the "three or more" category. With the boxplot, we can see it identifies the "three or more" category as an outlier. I reorganized the data and visualized the variable with only three bins (zero, one, and two or more conditions).

Figure 3: Boxplot of Counts of Comorbidity Variable with Three Levels

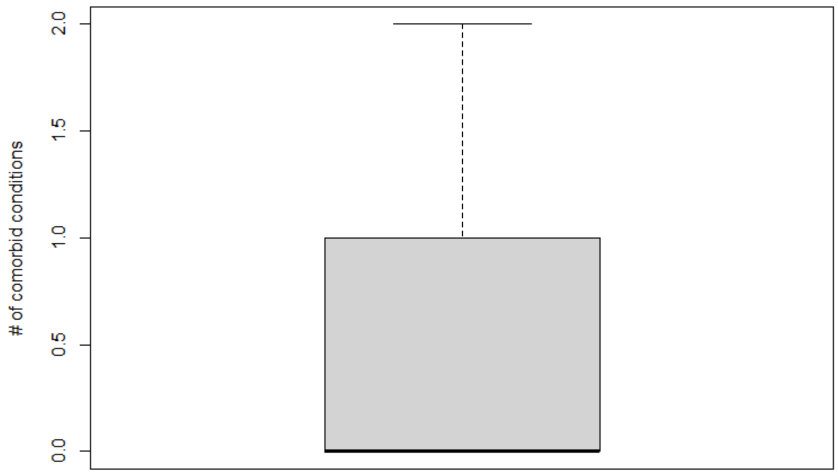
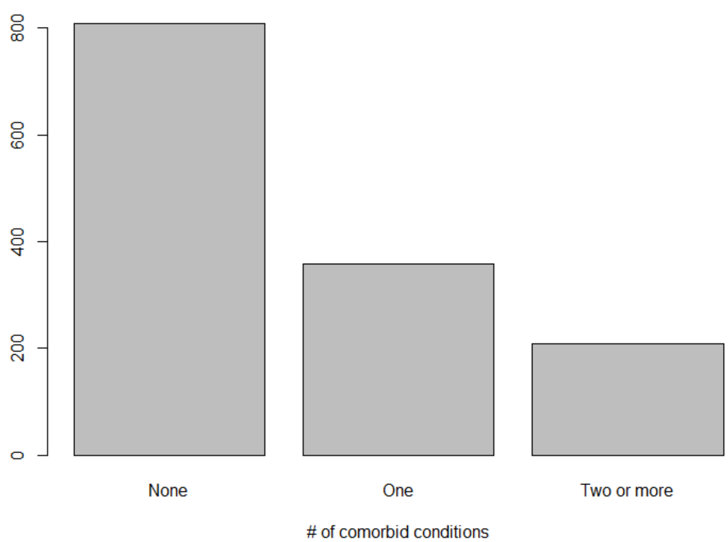


Figure 4: Barplot of Counts of Comorbidity Variable with Three Levels



The data is more evenly spread out when using three bins instead of four, although it is still heavily weighted towards "none" and few respondents reported more than one condition. The boxplot no longer detects an outlier.

In order to create get counts of number of comorbidities for each respondent, I changed the "NA" values for the second set of comorbidity questions to zero. These are the questions that respondents only see if they answered “yes” to the screening question, indicating that they have been told by a health professional that they have certain conditions. Respondents who answer “no” to the first question are coded as “NA” for the remaining, but conceptually, the NA is equivalent to zero because they already answered that they do not have any comorbid conditions.

I created a new categorical variable based on the counts of comorbid conditions and made sure to keep the referent the same as Burford et al so that our analysis could be compared. For both, the referent is “none”, allowing us to look at the odds someone with these conditions will participate in the outcome behaviors (reasons for leaving home and protective health behaviors) as compared to someone with no comorbid conditions.

I re-ran the logistic regression models, only changing the comorbidity variable. Everything else remained the same. I checked for VIF and calculated the odds ratios for each regression, to compare with Burford et al.’s results.

Empirical results

I replicated Burford et al.’s findings by running their logistic regression and calculating the odds ratios. I then ran a logistic regression with the multi-category comorbidity variables and calculated the odds ratios. The VIF for all models was between 1 and 2 for each variable, indicating that multicollinearity is not an issue. I ran the models twice, once with the comorbidity variable divided into four levels and once with it divided into three levels, and then compared with the original results.

All of the statistically significant variables in Burford et al.’s models had statistically significant counterparts in my models, but the statistical significance was not evenly distributed across the comorbidity levels. Some of the changes in statistical significance could be attributed to a smaller n, as decreasing the n decreases the statistical power and increases standard errors. In addition, I obtained statistically significant results for some levels of the comorbidity variables that did not generate a statistically significant result in Burford et al.’s models.

Table 2: Adjusted odds for reasons for leaving home during COVID-19 pandemic, original versus expanded models

Original Model versus Expanded Models						
	Original Model	*Expanded Model, Four Levels*			*Expanded Model, Three Levels*	
	AOR (95%CI)	*AOR (95%CI)*	*AOR (95%CI)*	*AOR (95%CI)*	*AOR (95%CI)*	*AOR (95%CI)*
	Comorbidity (*Referent=None*)	Comorbidity (*Referent=None*)			Comorbidity (*Referent=None*)	
	One or more	One comorbidity	Two comorbidities	Three or more comorbidities	One comorbidity	Two or more comorbidities
Grocery Shopping	**0.65*** (0.45, 0.96)**	**0.60** (0.40, 0.92)**	0.79 (0.44, 1.51)	0.68 (0.30, 1.88)	**0.60** (0.40, 0.92)**	0.76 (0.45, 1.34)
Exercise	**0.61*** (0.47, 0.79)**	**0.64*** (0.48, 0.86)**	**0.50*** (0.33, 0.73)**	0.74 (0.40, 1.37)	**0.64*** (0.48, 0.86)**	**0.54*** (0.38, 0.76)**
Walking Dog	1.13 (0.89, 1.45)	1.11 (0.84, 1.47)	**1.47 * (1.00, 2.16)**	0.63 (0.32, 1.19)	1.12 (0.84, 1.47)	1.19 (0.84, 1.68)
Other essential Shopping	1.18 (0.92, 1.515)	1.16 (0.88, 1.54)	1.09 (0.73, 1.61)	1.47 (0.81, 2.65)	1.16 (0.88, 1.54)	1.18 (0.83, 1.67)
Work	0.80 (0.53, 1.21)	0.89 (0.55, 1.40)	0.87 (0.44, 1.60)	**0.16 * (0.01, 0.78)**	0.89 (0.55, 1.41)	0.67 (0.34, 1.21)

N=1374

CI confidence interval, *AOR* adjusted odds ratio

* p<0.1 ** p<0.05 ***p<0.01

I included results that are statistically significant at p<0.1, indicated with a single asterisk, to show how the results changed when changing the comorbidity variable. Statistically significant results are bolded so that they are easier to read. Notably, none of the extension models show statistically significant results for all three subcategories of the comorbidity variable. In addition, when splitting the comorbidity variable into subcategories, we get statistically significant results for some additional outcomes. Both of these finding suggest that the number of comorbidities a person has can affect their behavior.

For grocery shopping, Burford et al. found statistically significant results for people with one or more comorbidity. In my extension, we see that the same association only for the first category (one comorbidity), and it has less statistical power, which makes sense due to the smaller n for the subcategory (358 in the subcategory, 566 in the original categories).

For the exercise outcome variable, Burford found that people with one or more comorbidities were about 61% as likely to exercise as people with no comorbidity. In the extension results, we can see that there are differences between people with one comorbidity versus two comorbidities. The odds for someone with only one comorbidity to exercise are a little higher than Burford et al.'s estimation, while the odds for someone with two comorbidities are lower. While Burford et al.'s model identifies that people with comorbidities are statistically less likely to leave the house to exercise during the pandemic, their model fails to note the differences between the subcategories of people. It makes sense that people with two comorbidities would be even less likely to leave the house to exercise, as they are dealing with additional health issues that likely affect what they are physically able to do.

When splitting comorbidity into four levels, the results show that people with one comorbidity are 64% as likely to leave the house to exercise (compared to someone with no comorbidities), while people with two comorbidities are only 50% as likely to leave to exercise. A fourteen-point difference is substantial enough to suggest there may be some underlying differences in how people with one comorbidity versus two comorbidities navigate the world. At the same time, the confidence intervals for all of the results (in the original analysis and in my analysis) are fairly wide, so no definitive conclusions can be drawn. The confidence intervals for one comorbidity and two comorbidities overlap, meaning it is possible that there is not truly a difference between the results for the two subcategories.

We see the same pattern when splitting the comorbidity variable into three levels, although combining the "two" and "three or more" levels dilutes the effect. In the four-level model, the two-comorbidity level has a 0.5 odds ratio and the three or more comorbidity level has a 0.74 odds ratio. In the three-level model, we see the odds ratio for the two or more category is 0.54, a little higher than the four-level model.

There are two cases where my four-level extension model obtained statistically significant results where the original model did not. The first is for the dog walking outcome variable. My analysis shows that people with two comorbidities are about 47% more likely to leave the house to walk a dog than the respondents with no comorbidities. This result is only significant at p<.1, but it is interesting that I obtained a statistically significant result where the original model did not. It is also interesting to look at the odds ratios between the two models. The original model estimated an odds ratio of 1.13 (although not statistically significant). In contrast, the odds ratios for my extended model vary across the subcategories. The result for the one comorbidity subcategory is the closest to the original model, with an odds ratio of 1.11. The two comorbidities and three or more comorbidities subcategories extend in opposite directions, with people with two comorbidities having higher odds of walking the dog (odds ratio of 1.47) and people with three or more comorbidities having much lower odds (odds ratio of 0.63). The original model's estimation as compared with the extension's estimation suggest that by aggregating the results, the original model smooths out the differences among subcategories. Of course, the results here are mostly not statistically significant (except for the two comorbidities subcategory) and the confidence intervals are wide.

So far, we have not seen any statistically significant results for the three or more comorbidities category. Due to having such a small number of observations for this category, this makes sense. It is much more difficult to obtain statistically significant results with a small n, so it is possible that if there were a relationship between having three or more comorbidities and the outcome variables, we may not see those results here due to the small n.

This makes the results for the "going to work" outcome variable even more noteworthy. While the original analysis did not show any statistically significant for the "going to work" variable, my analysis identifies a statistically significant result for people with three or more comorbidities. Significant at the $p<0.1$ level, my model estimates that people with three or more comorbidities are only 16% as likely as people with no comorbidities to leave the house for work. The confidence interval does not overlap with 1, which strengthens the findings. If the confidence interval overlapped with 1, then we would have less confidence about the direction of the odds. These results indicate that in 90% of samples, the odds of someone with three or more comorbidities leaving the home for work is 78% (or less) that of someone with no comorbidities. These results make sense in a substantive sense as well. People with health problems may be less able to work, and people with multiple health problems would be even less likely to be able to work.

Next, I looked at the results for the protective health behaviors outcomes.

Table 3: Adjusted odds for reasons for protective health behaviors during COVID-19 pandemic, original versus expanded models

Original Model versus Expanded Models						
	Original Model	*Expanded Model, Four Levels*			*Expanded Model, Three Levels*	
	AOR (95%CI)	*AOR (95%CI)*	*AOR (95%CI)*	*AOR (95%CI)*	*AOR (95%CI)*	*AOR (95%CI)*
	Comorbidity (*Referent=None*)	Comorbidity (*Referent=None*)			Comorbidity (*Referent=None*)	
	One or more	One comorbidity	Two comorbidities	Three or more comorbidities	One comorbidity	Two or more comorbidities
Social distancing	1.87 (0.87, 4.31)	1.47 (0.65, 3.65)	2.57 (0.71, 1.66)	4333492 (0, 1.43)	1.47 (0.65, 3.65)	**3.57*** **(0.99, 23.01)**
Washing hands	0.94 (0.61, 1.46)	0.87 (0.54, 1.42)	1.11 (0.41, 2.40)	1.10 (0.41, 3.85)	0.87 (0.54, 1.42)	1.11 (0.60, 2.163)
Use hand sanitizer	**1.39**** **(1.03, 1.88)**	**1.41*** **(1.00, 2.00)**	1.35 (0.85, 2.20)	1.33 (0.67, 2.83)	**1.41*** **(1.00, 2.00)**	1.34 (0.89, 2.06)
Use mask	1.19 (0.91, 1.55)	1.16 (0.86, 1.58)	1.08 (0.71, 1.66)	1.82 (0.89, 4.11)	1.16 (0.86, 1.56)	1.22 (0.83, 1.81)
Spend at least 2 – 3 hours at home	1.16 (0.89, 1.50)	1.15 (0.86, 1.55)	1.17 (0.78, 1.77)	1.12 (0.61, 2.13)	1.45 (0.86, 1.55)	1.15 (0.80, 1.67)
Use disinfectant wipes	1.04 (0.81, 1.32)	0.98 (0.74, 1.29)	1.14 (0.77, 1.68)	1.35 (0.74, 2.52)	0.98 (0.74, 1.29)	1.19 (0.84, 1.69)
Wearing gloves	1.02 (0.77, 1.35)	1.06 (0.76, 1.58)	0.87 (0.55, 1.36)	1.26 (0.63, 2.37)	1.05 (0.76, 1.45)	0.97 (0.65, 1.43)
N=1374						

CI confidence interval, *AOR* adjusted odds ratio

* $p<0.1$ ** $p<0.05$ *** $p<0.01$

The original analysis only found statistical significance for the "using hand sanitizer" behavior, and my four-level model found the same. However, when splitting out groups based on number of comorbidities, only people with one comorbidity were more likely to use hand sanitizer. The results in my extension also lose some statistical power, going from statistically significant at $p<0.05$ to significant at $p<0.1$. This could be due to the reduction in observations, as the original analysis had a larger group of observations since all respondents with comorbidities were aggregated.

The social distancing outcome bears looking at in a little more detail. In the four-level model, I obtained a very high coefficient and odds ratio for the three or more comorbid conditions level. This effect dissipates in the three-level model, and instead the model generates a statistically significant (at the $p<0.1$ level) result with the largest odds ratio out of all of the models generated. The unexpected result in the social distancing behavior could be due to a complete separation or quasi-complete separation issue, compounded by the small n in the three or more comorbidities category, or it may be a multicollinearity issue.

Overall, it looks as though there is some support for my hypothesis that splitting the comorbidity category into subcategories would yield different results than leaving the comorbidity category as a binary variable. The results in the original model and the extended model were not the same, so we can reject the null hypothesis. In particular, the extended model obtained statistically significant results for some outcome variables where the original model did not obtain statistically significant results.

Conclusion

This extension has shown that there is some evidence to support splitting out comorbidities by number of health conditions rather than treating them as a homogenous group in aggregate. Researchers should consider splitting out subcategories when there is a reason grounded in theory to do so. For example, the relationship seen between people with three or more comorbidities and the odds of leaving the house for work is something that makes sense substantively, and the differences between the subgroups could skew results if those subgroups are aggregated and treated as a homogenous category. However, while disaggregating the data showed some different relationships, Burford et al.'s findings still hold and were robust to the disaggregation.

Future research should consider not only how split the categories, but also how to operationalize variables such a comorbidity. Burford et al. did not discuss how they decided which conditions to include as comorbidities for this study, or how they decided on the survey question format. Other researchers suggest the relationship between comorbidities and protective behaviors is tied to self-perception of susceptibility (O'Conor et al. 2020). Burford et al. framed the question about comorbidity as whether a health professional had diagnosed the conditions. However, the diagnosis from a health professional may not be necessary for an individual to have a self-perception of susceptibility and a sense of their own health status. For example, individuals can calculate their own BMI without consulting a doctor and may develop a self-perception of susceptibility after learning from the CDC that people with high BMIs are at greater risk from COVID-19. The way Burford et al. operationalized the comorbidity variable may have underestimated the number of people with comorbidities by excluding those who did not get an official diagnosis from a health professional.

Sources

Bish, A., & Michie, S. (2010). Demographic and attitudinal determinants of protective behaviours during a pandemic: A review. British journal of health psychology, 15(4), 797-824.

Burford, K., Dooley, E., Lanza, K., & Knell, G. (2020). Associations of Urbanicity and Sociodemographic Characteristics with Protective Health Behaviors and Reasons for Leaving the Home During COVID-19. Health Psychology Bulletin, 4(1), 53–63. DOI: http://doi.org/10.5334/hpb.24

CDC. Coronavirus Disease 2019 (COVID-19). Centers for DiseaseControl and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fneed-extra-precautions%2Fgroups-at-higher-risk.html. Published April 29, 2021. Accessed May 1, 2021.

Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL https://www.jstatsoft.org/v40/i03/.

Hadley Wickham (2020). tidyr: Tidy Messy Data. R package version 1.1.2. https://CRAN.R-project.org/package=tidyr

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.3. https://CRAN.R-project.org/package=dplyr

John Fox and Sanford Weisberg (2019). An {R}Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

O'Conor, R., Opsasnick, L., Benavente, J.Y. et al. Knowledge and Behaviors of Adults with Underlying Health Conditions During the Onset of the COVID-19 U.S. Outbreak: The Chicago COVID-19 Comorbidities Survey. J Community Health 45, 1149–1157 (2020). https://doi.org/10.1007/s10900-020-00906-9

Sam Firke (2021). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 2.1.0. https://CRAN.R-project.org/package=janitor

Schratz, P. (2017): R package 'oddsratio': Odds ratio calculation for GAM(M)s & GLM(M)s, version: 1.0.2, doi: 10.5281/zenodo.1095472

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686