Manuscript

Christoph Völtzke

2023-01-27

1. Introduction

Since early 2020, governments across the world have required people to alter their social behavior in an attempt to slow transmission of the SARS-CoV-2 virus (National Institute for Public Health and the Environment, 2022). In the United Kingdom, the government has implemented a variety of measures over the course of the pandemic, including three separate national "lockdowns' ' as well as other local and national measures. In addition, guidance has been issued on risk mitigation measures during social interactions, including meeting outdoors, maintaining space between people, frequent handwashing or use of hand sanitizer, and the use of face coverings (masks). Yet, since the pandemic's very beginning news reports and social media posts alerting society about potential offences against these measures have been just as frequent as their implementation. For that reason, governments and research institutes have been increasingly interested in surveying peoples' compliance with government regulations. Accordingly, in the United Kingdom, the center of mathematical modelling of infectious diseases has brought forward the CoMix social contacts survey trying to gain insights into the number of social contacts people had during the COVID-19 pandemic (Gimma et al., 2021; Jarvis et al., 2020). However, as the survey is carried out, using nonprobability sampling, estimates may be severely biased, posing the questions of: (1) How can inference be reliably drawn from nonprobability survey samples and (2) what measures help mitigate the bias caused by nonprobability sampling?

In order to answer these questions posed, the following assignment assesses weighting as a means to obtain reliable estimates from nonprobability survey sampling. More specifically, it discusses poststratification weighting and carries out the analysis of the CoMix social contacts data, using poststratification weighting to the three-way cross tabulation of age, gender, and household type. With respect to the analysis, we find poststratification weighting to alter the analyses' estimates only slightly. Note that this may indicate that the estimates may still be severely biased and that inference can still not be reliably drawn, regardless of the proposed weighting procedure. Regarding the CoMix social contacts survey, we find seniors and people living alone to be the most socially isolated in times of the COVID-19 pandemic. Moreover, analysis of the number of pre-pandemic contacts reveals that government regulations exhibited lasting effect on people's social behavior, effectively reducing the number of contacts by approximately 76%. Thus, despite the suspected unreliability in the analysis' estimates, strict government social distancing regulations can be seen as a viable tool to reduce the amount of social interactions.

Our assignment is ordered as follows: Section 2 discusses the CoMix dataset as well as the other data files used in the assignment's analysis. Moreover, section 2 reviews the weighting approach taken in this study and assesses the advantages and limitations of poststratification weighting in a nonprobability based framework. Section 3 involves the analysis and, lastly, section 4 discusses the results found in the preceding section. Note: before running the code in any of the following sections, it is important to first go through Appendix A: Data preparation.

2. Method

2.1 Study Design and Reporting of contacts

The CoMix survey is an online survey that runs weekly, with alternating panels of participants (of about 1,500 participants) submitting the survey bi-weekly. We focus on part of the collected data: data from adults, within the period from March 2020 until August 2020. Panel participants were asked to fill in the survey for a maximum of 10 rounds, so over a maximum period of 20 weeks (Gimma et al., 2021).

The survey was used to gather information regarding the number of contacts people had with others, as well as auxiliary information about these contacts such as place of contact (e.g. home, work, or school), age & gender of contacted person, and duration of contact. On top of that, some additional (demographic) information on the panel participants was also collected. This includes gender, age, and household type.

The participants were asked to report the contacts they had on the day prior to filling out the survey. They were asked to report both the total number of contacts, as well as to provide additional information on each separate contact. These additional questions were added for use in a risk assessment. They focus, for example, on specifying the time spent with the contact and whether or not the contact was physical.

2.2 Additional datasets

To put the CoMix data into further context and to be able to draw inferences regarding social contacts behavior in the UK additional datasets are needed. These datasets are supposed to explain certain changes in behavior as a reaction to the COVID-19 pandemic and respective government measures to reduce social contacts.

First, recent population data is needed to assign poststratification weights to the participants of the CoMix survey. Therefore, population-level data was drawn from the UK's Office for National Statistics. Data on age and gender distribution were drawn from the "Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland' ' dataset and data on single-person households from the "Families and Households" dataset. As this was the only participant information given in the survey, we did not conduct other socio-demographic variables to subgroup the participants.

The CoMix survey was designed to assess changes in social contacts behavior of the UK population over several months as a reaction to the COVID-19 pandemic. Therefore, we acquired the daily new cases of COVID-19 in the UK of the same time period and plotted them alongside mean contact data to illustrate the relationship between mean contacts and cases. The information on COVID-19 cases was provided by the UK government online coronavirus dashboard, which we acquired using the covidregionaldata R package.

Data on the stringency of government measures are based on the COVID-19 Stringency Index. The index is a composite measure of nine government response metrics, including school closures; workplace closures; cancellation of public events; restrictions on public gatherings; closures of public transport; stay-at-home requirements; public information campaigns; restrictions on internal movements; and international travel controls. The index is carried out as part of the Oxford Coronavirus Government Response Tracker project.

Another additional dataset is the POLYMOD survey which contains data of social contact behavior in a pre-pandemic setting. The CoMix survey adapted the study design based on the POLYMOD survey. Therefore, the same questions on social contacts behavior are asked and the POLYMOD survey can be used as a baseline for social contact behavior prior to the COVID-19 pandemic. Following, we can compare the main outcome variables of the CoMix survey with the data provided by the POLYMOD survey to understand the degree of change in social contact behavior between pandemic and non-pandemic times.

2.3 Weighting Approach

Among survey researchers, it is well established that probability based surveys outperform their nonprobability counterparts when comparing their accuracy to population level statistics (Cornesse et al., 2020). Hence, many researchers examining the accuracy of probability and nonprobability sample surveys have attempted to eliminate biases in nonprobability sample surveys by weighting their sampling distribution (e.g. Mercer et al., 2018; Wang et al., 2015). Although Cornesse and colleagues (2020) note that a majority of these studies found that weighting did not sufficiently reduce bias in nonprobability based surveys, we follow the literature's general advice, weighting our sample by means of poststratification.

2.3.1 Poststratification in nonprobability sample surveys

While originally designed for the adjustment of sampling errors and potential non-response bias in probability samples surveys, poststratification is equally employed in a nonprobability based framework. In both cases, it forms units into homogeneous subgroups after observation of the sample and employs auxiliary information to align the sampling distribution of these subgroups to the population distribution. Ideally, this information would be the true values of the proportions falling into the subgroups but usually the only information available are estimates based on administrative records, previous census results, or previous much larger surveys. Still, as a method for removing biases due to non-random causes, poststratification can have great value although the gains are, according to Smith (1991), greater in institutional surveys and often quite small in social surveys.

With respect to its applicability in nonprobability sample surveys, the literature, however, remains rather skeptic. While Berrens et al. (2003) conclude that weighting reduced the bias in a nonprobability sample survey sufficiently, Gittelman and colleagues (2015) show that poststratification weighting successfully reduced biases in a probability sample survey but was less successful across a number of nonprobability survey samples and even increased bias in one instance. This is in line with Cornesse et al. (2020) who find that the inaccuracy of nonprobability samples cannot be reliably solved by weighting procedures. Instead, the literature indicates that to get accurate estimates from nonprobability sample surveys by weighting, the availability of variables that predict the outcome of interest are more important than which statistical method is used (Mercer et al., 2018).

2.3.2 Weighting the CoMix data

Regardless of these disenchanting findings, we performed poststratifiation weighting to the three-way cross-tabulation of age, gender, and household type. Regarding the assessment of these variables, four age brackets were formed, assigning respondents to young adults (18-24 years old), adults (25-44 years old), mature adults (45-64 years old) and seniors (65 years and over). For gender, participants were divided in male and female participants. Respondents identifying as another gender were excluded from the analysis since the UK's office for official statistics does not comprise population level statistics about non-binary citizens. Lastly, household type was assessed by separating participants living in single person households from respondents not living in single person households.

Note that the selection of variables was informed by the sparsity in additional information found in the CoMix dataset. Yet, in line with the recommendation by Mercer and colleagues (2018), our analysis found all three variables to be significant predictors of our dependent variable, i.e. the number of contacts per person.

```
##
## Table 5: Poststratification Variables
## ====================================================
##                               Dependent variable:
##                          ----------------------------
##                               Number of Contacts
## ----------------------------------------------------
## 25-44 years old                     -0.051
##                                     (0.064)
```

```
##
## 45-65 years old                  -0.039
##                                   (0.063)
##
## 65 years and over              -0.678***
##                                   (0.064)
##
## Male                            -0.105***
##                                   (0.027)
##
## Multi People Household          0.209***
##                                   (0.050)
##
## Constant                        18.364**
##                                   (6.759)
##
## -----------------------------------------------------
## Observations                     22,842
## Log Likelihood                -47,826.460
## Akaike Inf. Crit.              95,666.930
## =====================================================
## Note:                  *p<0.05; **p<0.01; ***p<0.001
```

Note here also that a time fixed effects predictor (date) is added to the regression model since the CoMix dataset comprises of repeated cross-sectional data. However, to circumvent a predictor inflated output, date is omitted from this and all following regression outputs. Instead, the full regression outputs are listed in Appendix B.

Overall, poststratification weighting caused only slight alternations in the survey's estimates. For example, the overall number of mean contacts aggregated over the entire 19 week period changed by just 0.01 when adjusted for poststratification weighting.

```
## # A tibble: 1 x 2
##   mean.wgt  mean.non.wgt
##   <svystat> <svystat>
## 1 2.580821  2.5943
```

3. Results

3.1 Social distancing in times of Covid-19: Seniors and people living alone are among the most socially isolated

3.1.1 Analyses not involving interaction

In our exploratory analysis, we find a number of trends in the amount of respondents' social interactions. Observed over a 19 week period (March 2020 to August 2020), contacts generally follow stark fluctuations best described by a polynomial expression. Following a sharp increase in weekly Covid-19 cases, contacts fall until they reach an all time low in week 18, only to be soaring again with falling new cases. The two following figures plot these fluctuations in weekly social contacts. Note that the first dashed, vertical line indicates the government's lifting of strict lockdown measures and the second a further repeal of regulations.

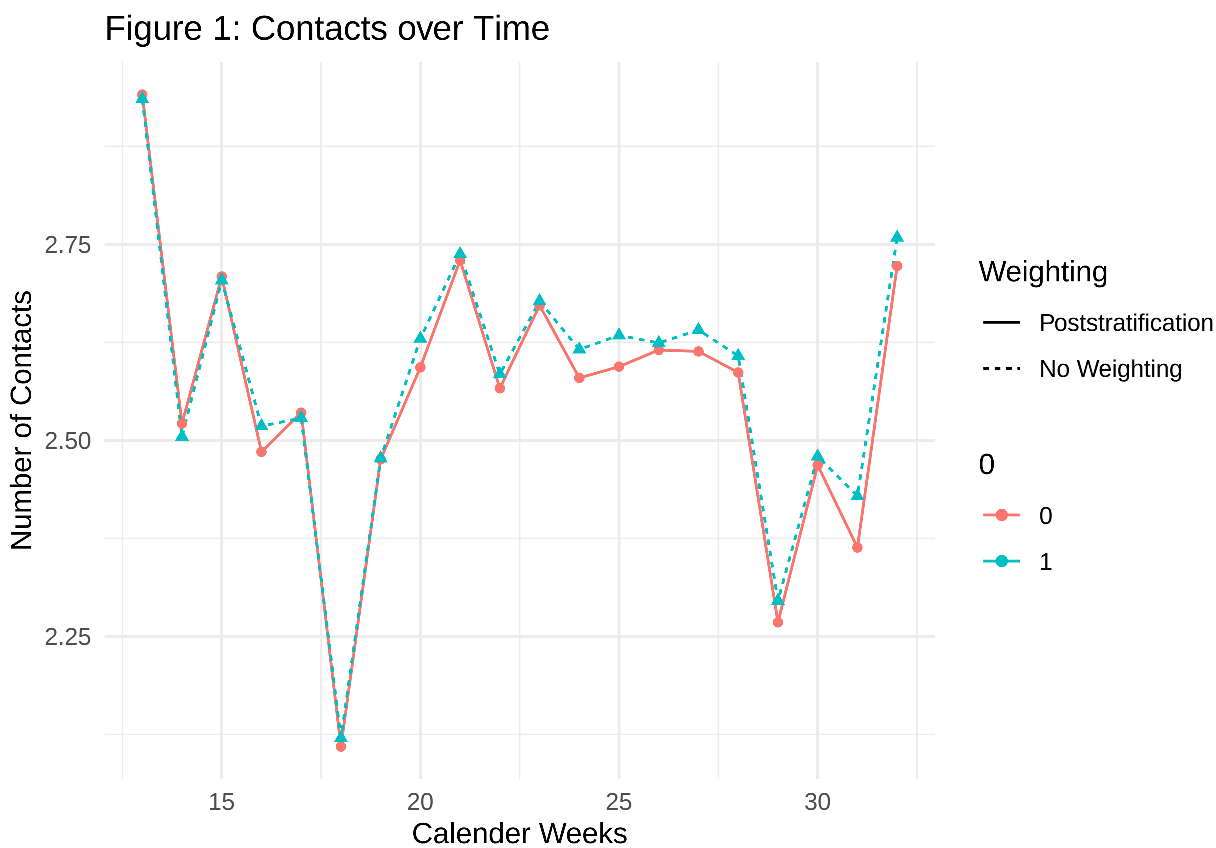
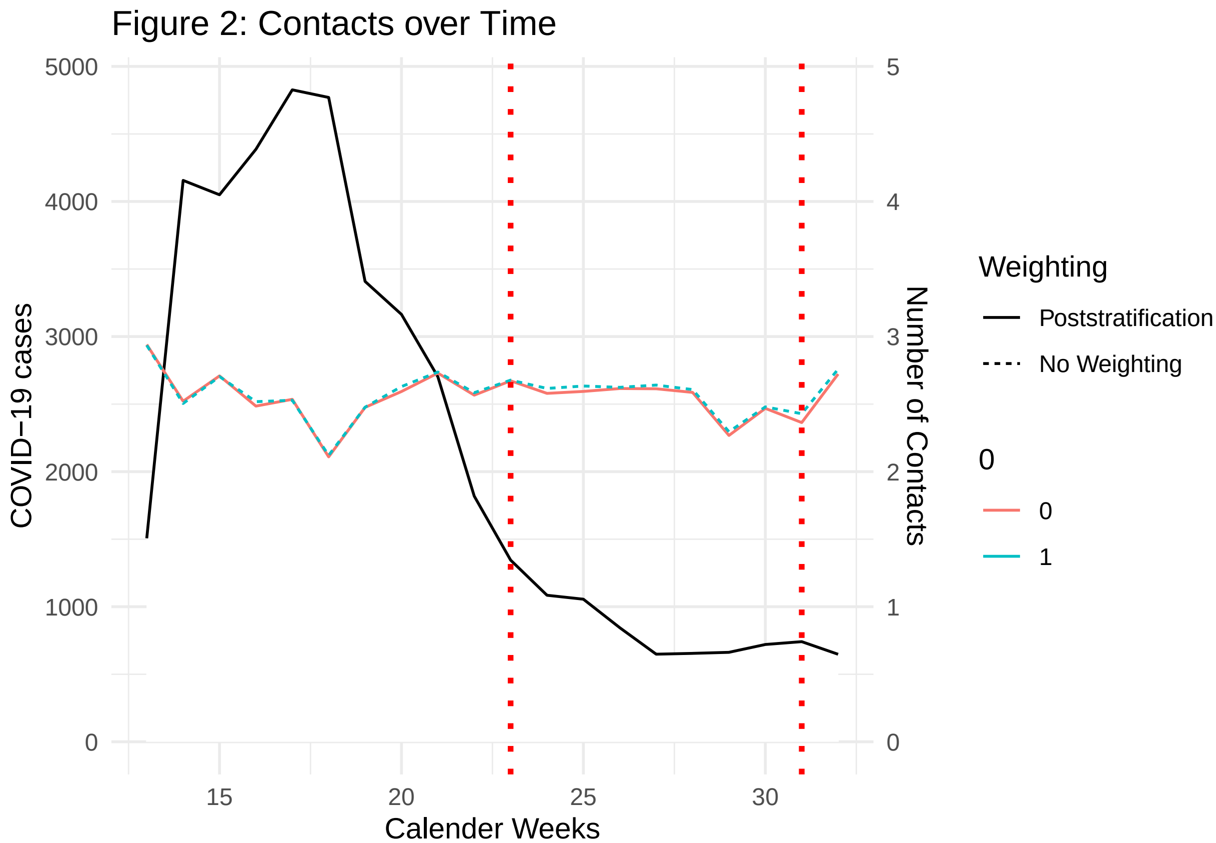


Figure 1: Contacts over Time
Number of Contacts
2.75
2.50
2.25
15
20
25
30
Calender Weeks
Weighting
Poststratification
No Weighting
0
0
1

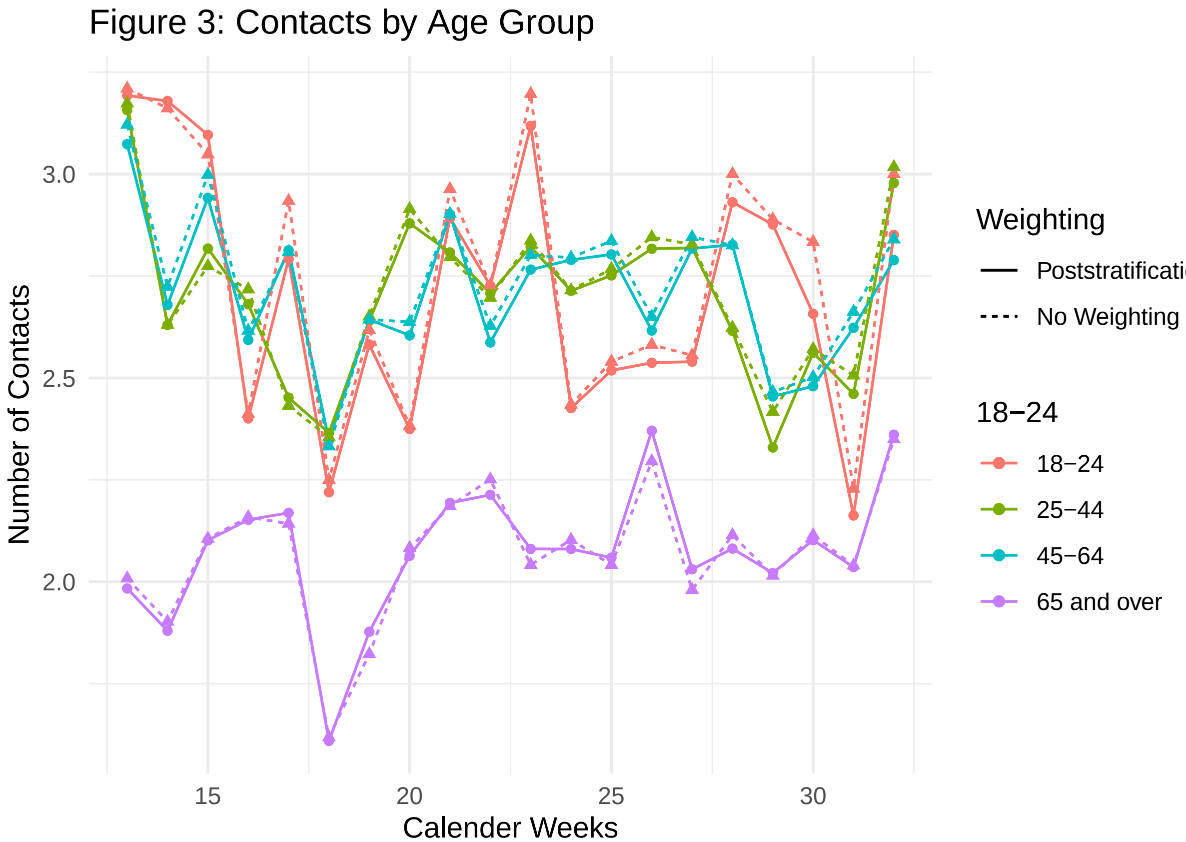


Interestingly, the data indicates a reduction in peoples' social interactions with tightening social distancing measures although the relationship is not found to be significant. The data also suggests a statistically significant positive association between COVID-19 infections and number of social contacts, indicating that people may be unwilling to reduce social interactions when COVID-19 infection rates are high if government regulations do not force them to.

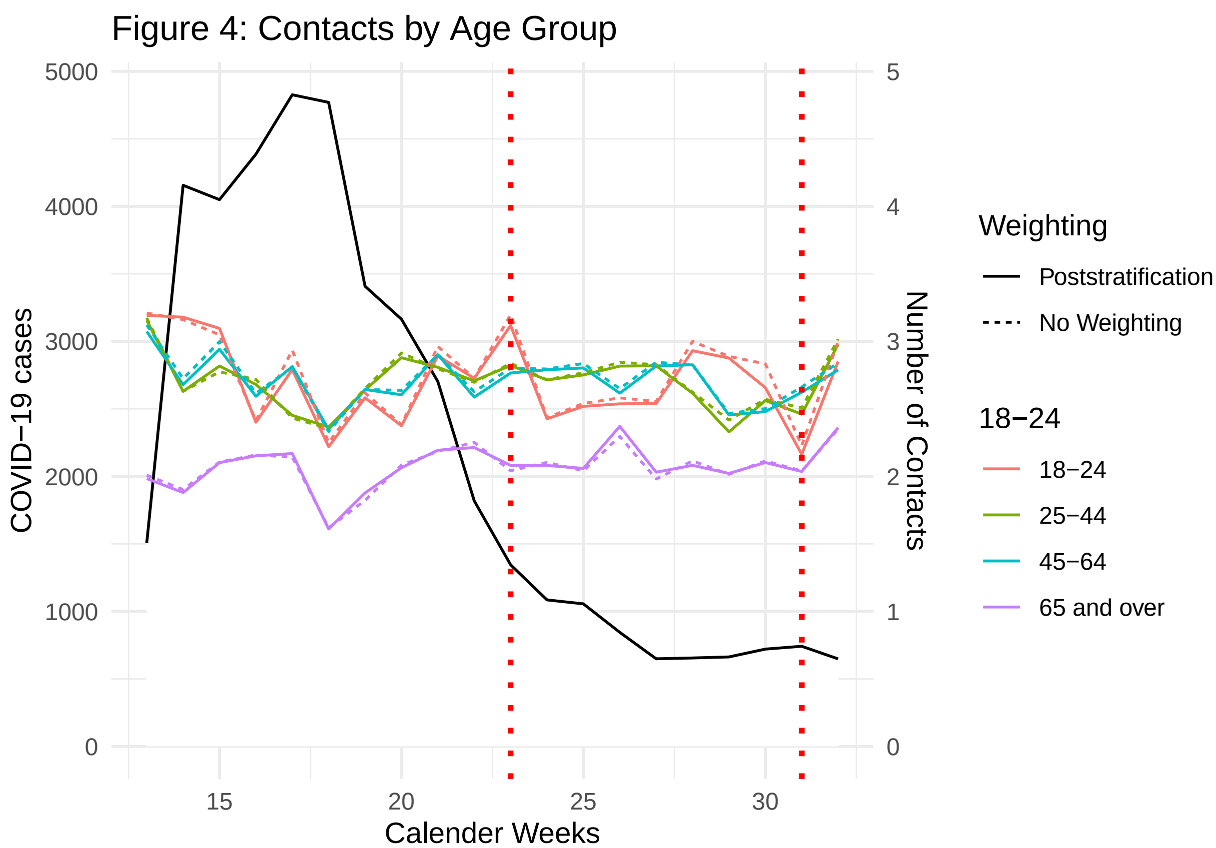
```
##
## Call:
## svyglm(formula = contacts ~ stringency_index + cases_new + date,
##     design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       7.236e+01  1.449e+01   4.995 5.94e-07 ***
## stringency_index  1.569e-03  5.061e-03   0.310    0.757
## cases_new        -9.533e-05  1.593e-05  -5.985 2.20e-09 ***
## date             -3.785e-03  7.738e-04  -4.892 1.00e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 3.733398)
##
```

```
## Number of Fisher Scoring iterations: 2
```

Looking at the differences across age groups, we find seniors (people aged 65 and over) to be the by far most socially isolated group. Compared to young adults (18-24 years old), they exhibit an average of 0.63 contacts less



per day (see table 6).




```
##
## Table 6: Contacts by Age Group
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                         Number of Contacts
## -----------------------------------------------
## 25-44 years old              -0.059
##                              (0.064)
##
## 45-65 years old              -0.062
##                              (0.064)
##
## 65 years and over           -0.728***
##                              (0.064)
##
## Constant                    18.648**
##                              (6.764)
##
## -----------------------------------------------
## Observations                 22,842
## Log Likelihood             -47,851.010
## Akaike Inf. Crit.          95,712.030
## ===============================================
## Note:               *p<0.05; **p<0.01; ***p<0.001
```

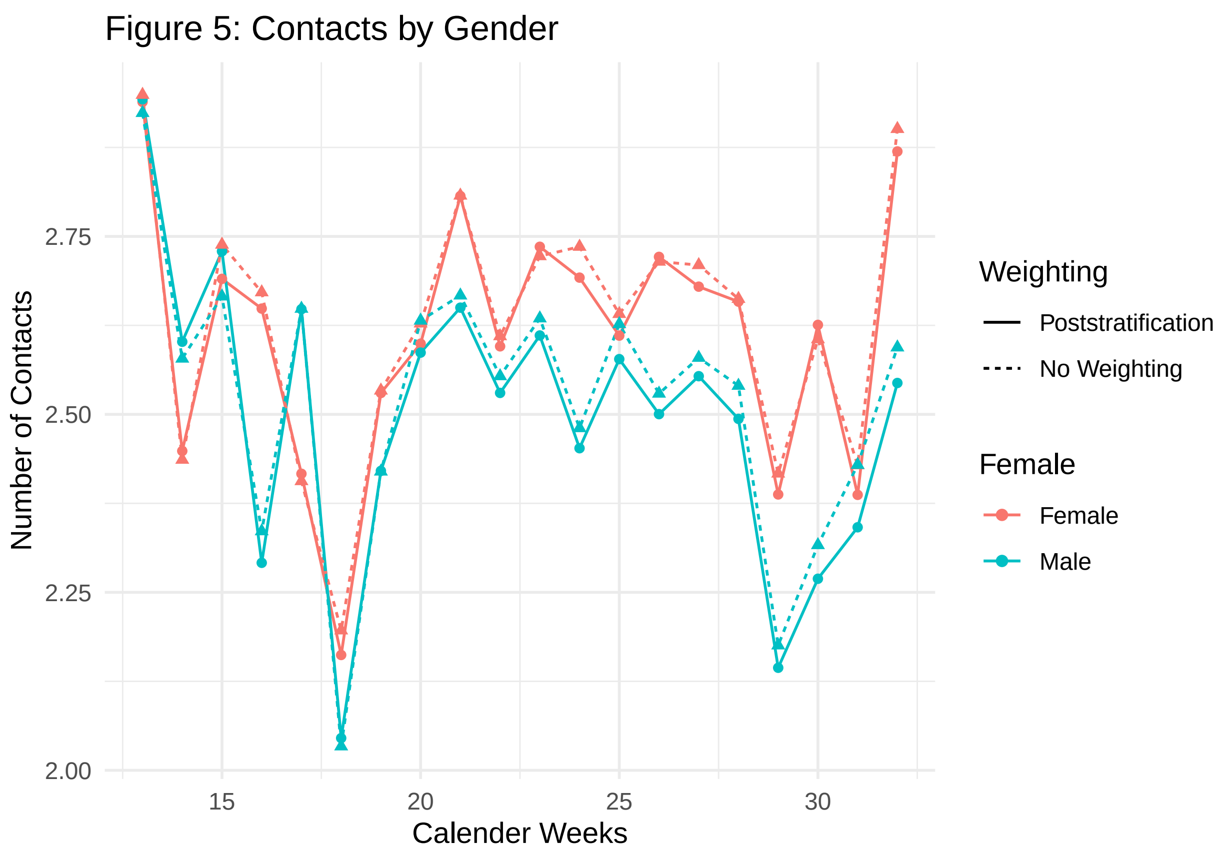


Figure 5: Contacts by Gender
Number of Contacts
2.00
2.25
2.50
2.75
15
20
25
30
Calender Weeks
Weighting
Poststratification
No Weighting
Female
Female
Male

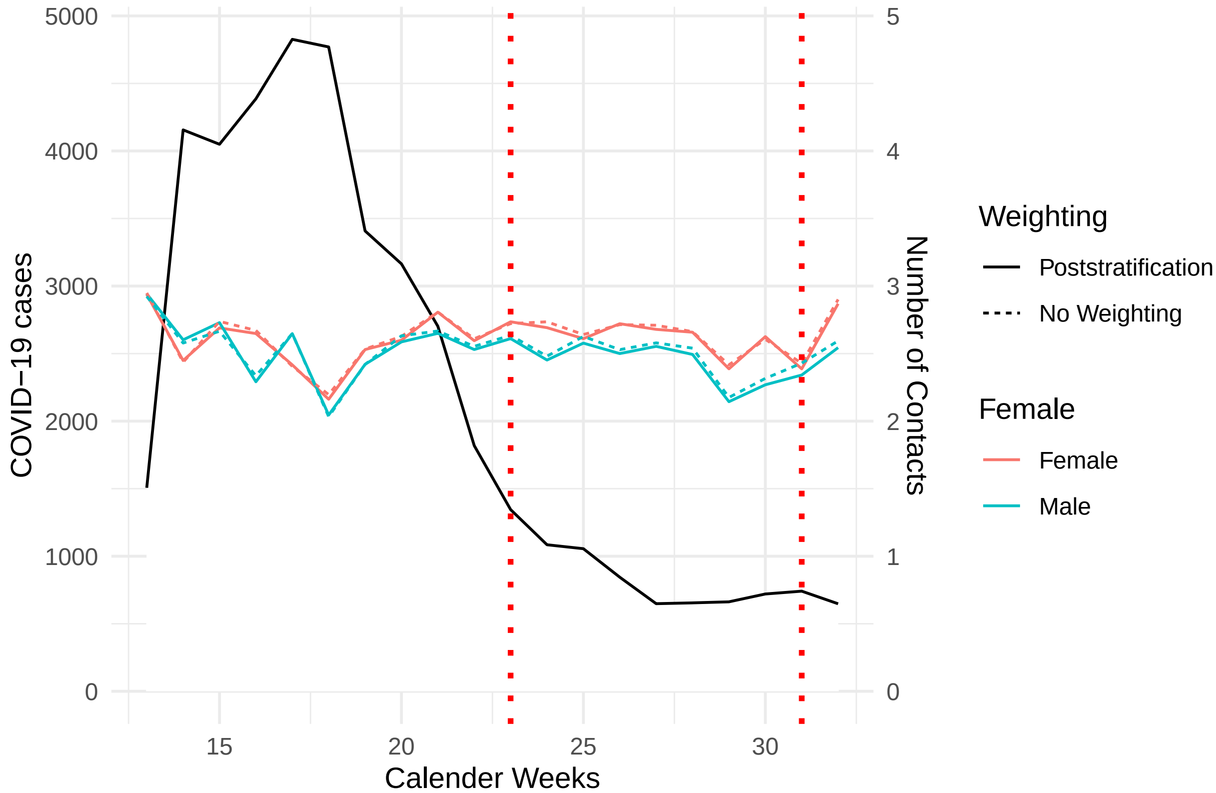
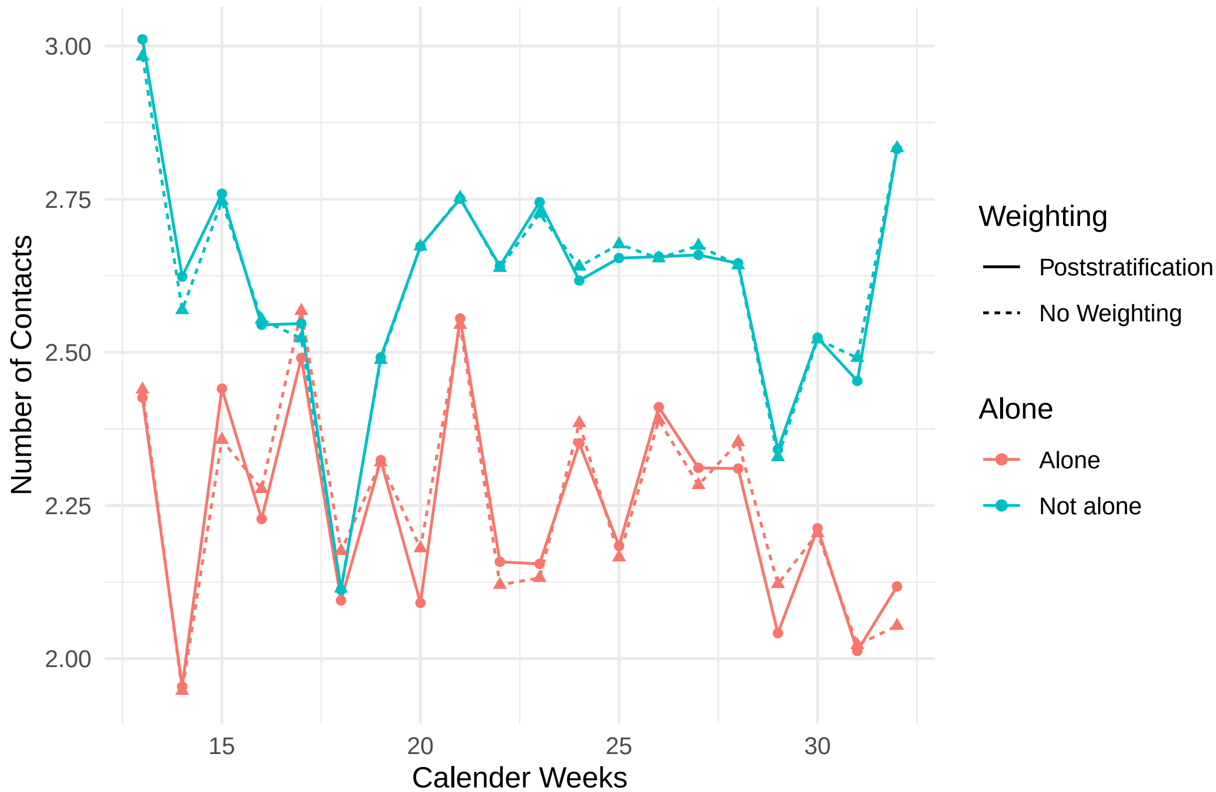
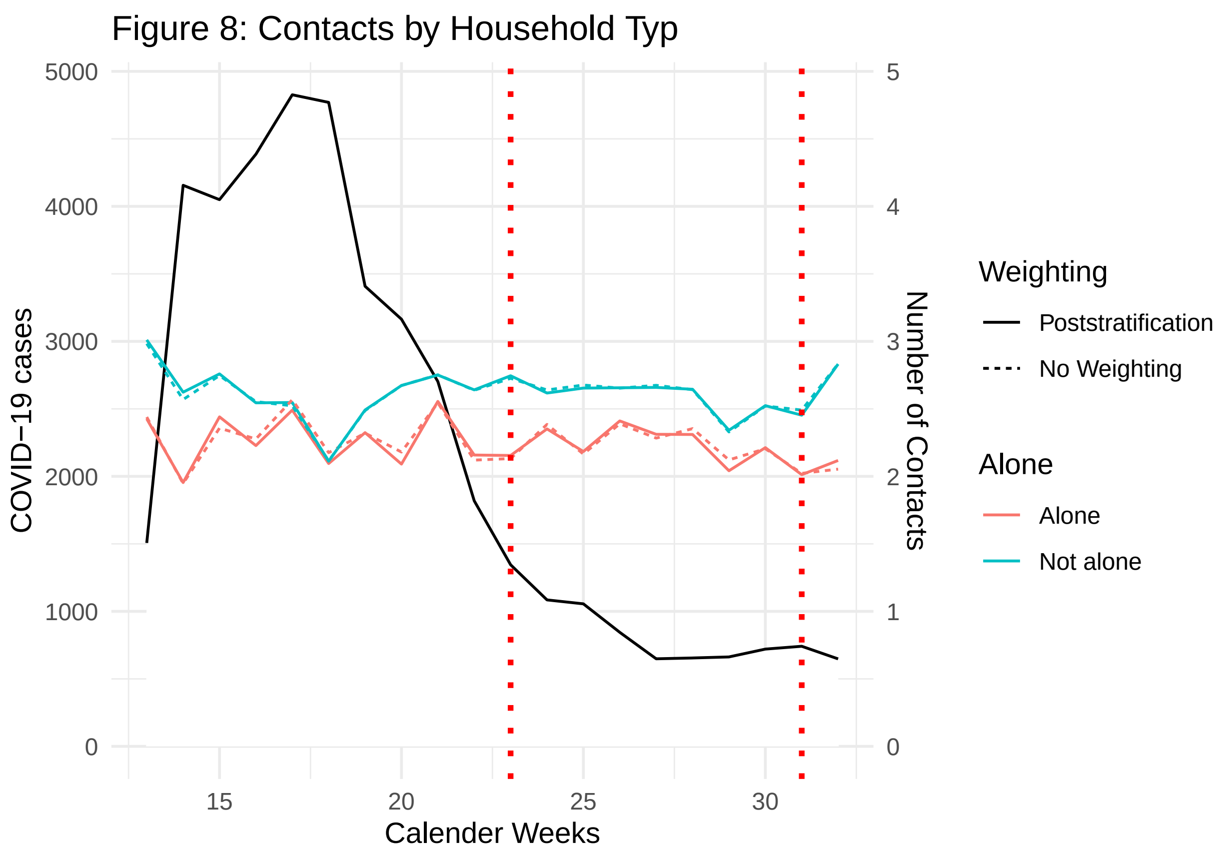


Figure 6: Contacts by Gender

```
##
## Table 7: Contacts by Gender
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                          Number of Contacts
## -----------------------------------------------
## 25-44 years old              -0.083**
##                               (0.027)
##
## 45-65 years old              22.913***
##                               (6.870)
##
## -----------------------------------------------
## Observations                  22,842
## Log Likelihood              -48,100.780
## Akaike Inf. Crit.           96,207.550
## ===============================================
## Note:               *p<0.05; **p<0.01; ***p<0.001
```

Figure 7: Contacts by Household Typ






```
##
## Table 8: Contacts by Household Type
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                          Number of Contacts
## -----------------------------------------------
## 25-44 years old               0.382***
##                               (0.049)
##
## 45-65 years old              20.731**
##                               (6.874)
##
## -----------------------------------------------
## Observations                  22,842
## Log Likelihood              -48,049.210
## Akaike Inf. Crit.           96,104.430
## ===============================================
## Note:               *p<0.05; **p<0.01; ***p<0.001
```

3.1.2 Analyses involving interaction

Since the data indicates elderly people to be more socially isolated in times of the covid pandemic, we created a dummy variable, indicating whether a respondent is a pensioner or not. The following analysis looks at potential interactions between age and gender as well as between age and household type.

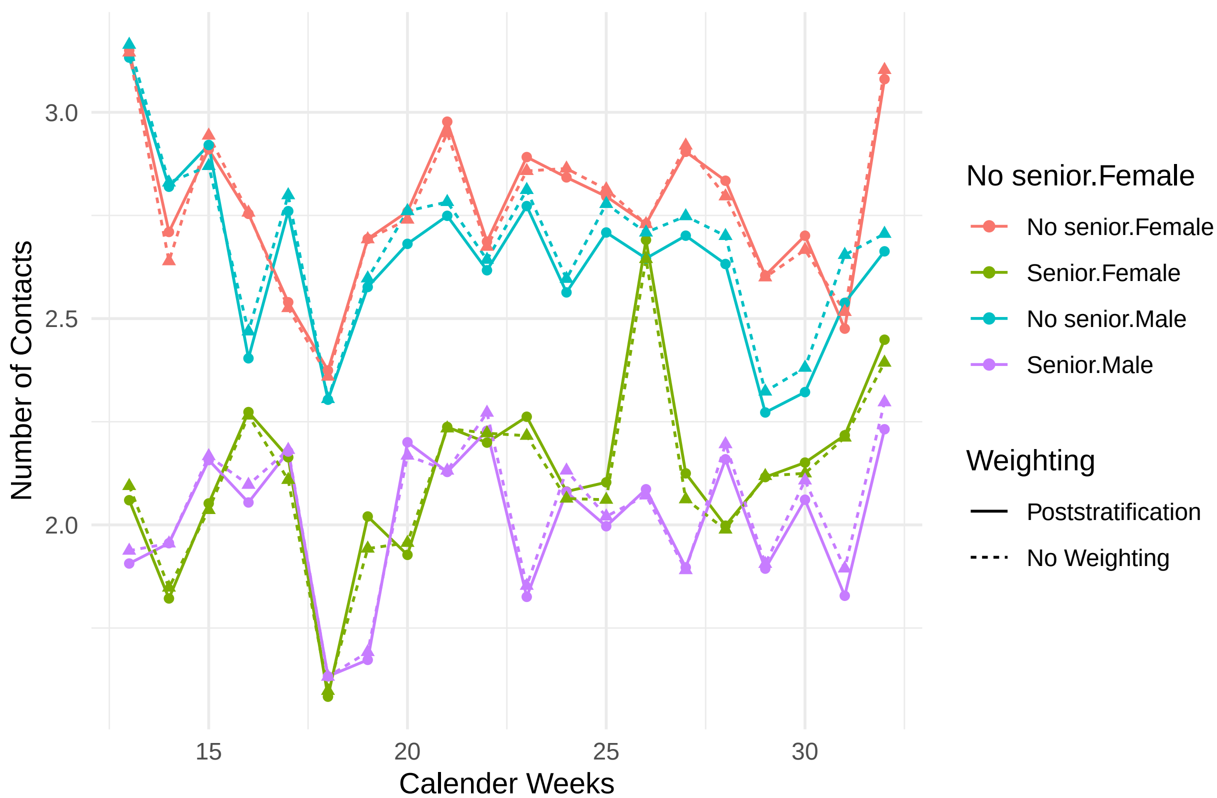
Inspecting the first interaction plot, we find that across age groups women consistently display significantly more contacts than men. However, table 11 indicates no significant interaction between age and gender.

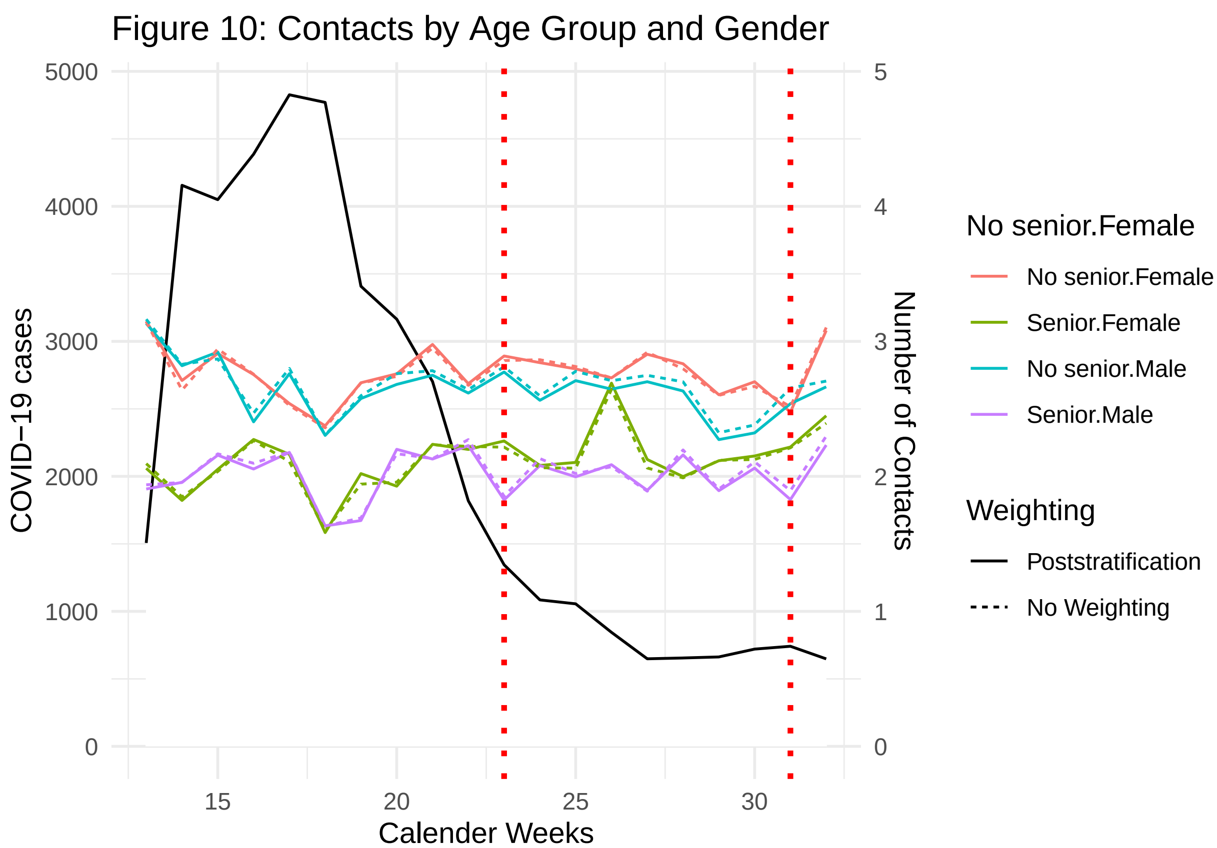
```
##
## Table 9: Contacts by Location
## ========================================================
##                                 Dependent variable:
##                            ----------------------------
##                                 Number of Contacts
## --------------------------------------------------------
## factor(location)Leisure               1.037*
##                                       (0.502)
##
## factor(location)Otherplace           -0.370***
##                                       (0.044)
##
## factor(location)School               -0.049
##                                       (0.299)
##
## factor(location)Transport             0.450
##                                       (0.264)
##
## factor(location)Work                 1.432***
##                                       (0.174)
##
## Constant                             20.865**
##                                       (6.741)
##
## --------------------------------------------------------
## Observations                          22,842
## Log Likelihood                      -47,855.180
## Akaike Inf. Crit.                   95,724.360
## ========================================================
## Note:                      *p<0.05; **p<0.01; ***p<0.001
```

```
##
## Table 10: Contacts by Periods
## =================================================================
##                                          Dependent variable:
##                                     -----------------------------
##                                          Number of Contacts
## -----------------------------------------------------------------
## factor(periods)Lockdown_1_ease                 0.151**
##                                                (0.051)
##
## factor(periods)Reduced_restrictions            0.265**
##                                                (0.086)
##
## Constant                                      58.357***
##                                               (13.850)
##
## -----------------------------------------------------------------
## Observations                                   22,842
## Log Likelihood                               -48,100.280
```

```
## Akaike Inf. Crit.                                   96,208.560
## =============================================================
## Note:                                 *p<0.05; **p<0.01; ***p<0.001
```

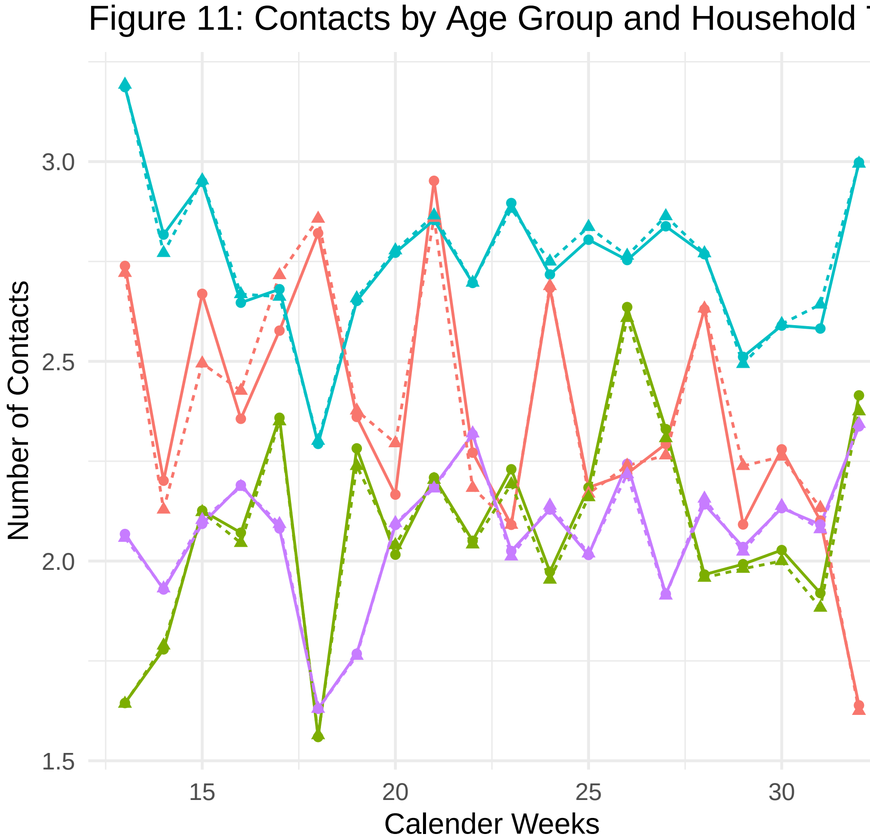
Figure 9: Contacts by Age Group and Gender





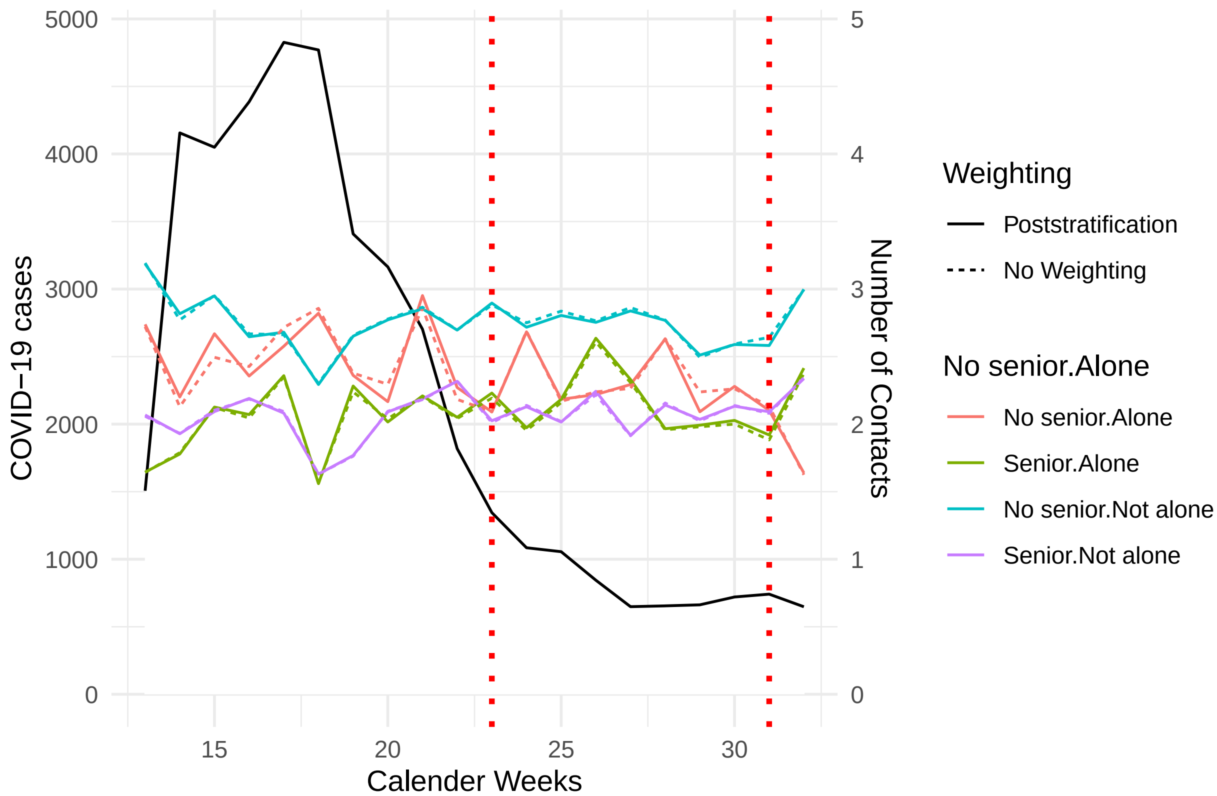

```
##
## Table 11: Contacts by Age Group and Gender
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                         Number of Contacts
## -----------------------------------------------
## 25-44 years old              -0.669***
##                               (0.044)
##
## 45-65 years old              -0.097**
##                               (0.033)
##
## 65 years and over             -0.023
##                               (0.058)
##
## Constant                     19.725**
##                               (6.781)
##
## -----------------------------------------------
## Observations                  22,842
## Log Likelihood              -47,843.770
## Akaike Inf. Crit.           95,697.530
## ===============================================
## Note:               *p<0.05; **p<0.01; ***p<0.001
```

Contrarily to the first interaction plot, the data do, however, indicate a significant interaction between age and household type. In fact, the amount of social contacts of people between the age of 18 and 64 seems to be de-



pendent on the household type they are living in.

Figure 12: Contacts by Age Group and Household Type




```
##
## Table 12: Contacts by Age Group and Household Type
## ===============================================
##                        Dependent variable:
##                   ----------------------------
##                        Number of Contacts
## -----------------------------------------------
## 25-44 years old             -0.297**
##                              (0.092)
##
## 45-65 years old             0.386***
##                              (0.072)
##
## 65 years and over           -0.432***
##                              (0.097)
##
## Constant                    18.456**
##                              (6.804)
##
## -----------------------------------------------
## Observations                 22,842
## Log Likelihood             -47,818.600
## Akaike Inf. Crit.          95,647.190
## ===============================================
## Note:              *p<0.05; **p<0.01; ***p<0.001
```

3.2 Descriptive Statistics CoMix survey and Polymod survey

Our next goal is to compare the mean number of social contacts per person in the CoMix survey to the POLYMOD survey which serves as the pre-pandemic baseline in this assignment. Through these analyses, we aim to show to what extent the social contact behavior changed due to a variety of implemented measures by the government.

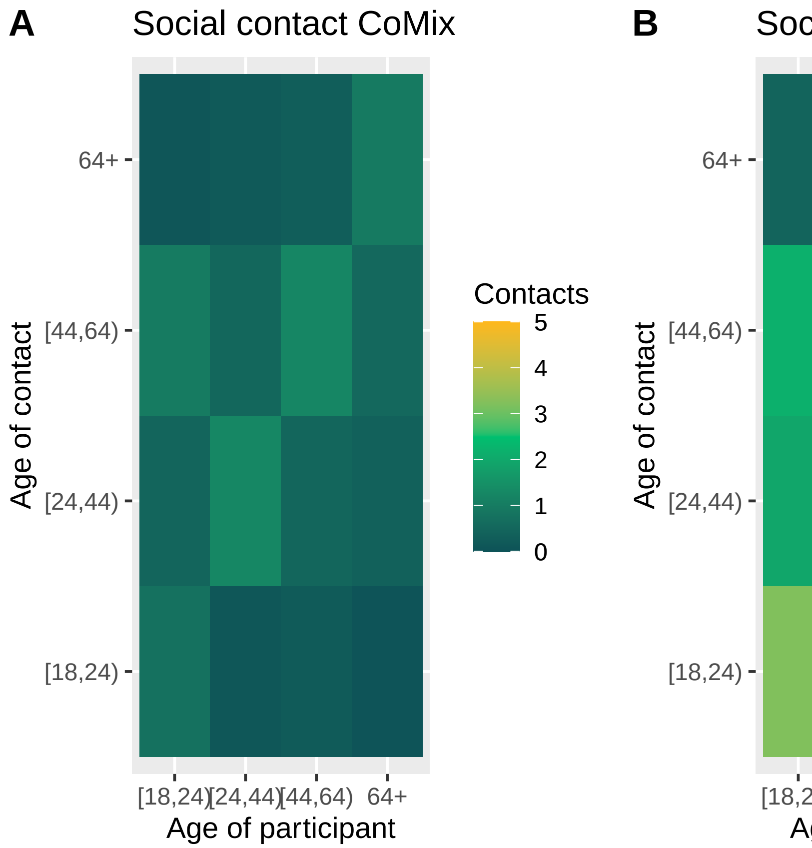
Moreover, we calculated the mean number of contacts stratified by age groups, gender, type of contact, and location of contact. This subdivision provides further insights into whether there are differences in behavior in distinct subgroups of individuals or through special situations. Another common feature in social contact research is the use of contact matrices (Jarvis et al., 2020). These contact matrices show the average total number of daily reported contacts made by participants in different age groups with individuals in other age groups. We constructed age-stratified contact matrices for four age groups (18 to 24, 25 to 44, 45 to 64, 65+). Further, we investigate to what extent the poststratification of the CoMix data influenced the results by comparing the mean number of social contacts per person of the unweighted CoMix data to the weighted CoMix data. As the CoMix data was conducted as a quota sample by market research institutes we only expect small differences in the outcome variables.

3.2.1 Overall mean comparison

The mean number of social contacts per person per day measured during the CoMix study was 2.581 compared to 10.796 contacts measured in the POLYMOD study (see Table 13). With respect to all participants, this has resulted in a very large (76%) reduction in the total number of contacts per day. Concerning our additional research interest to investigate how poststratification of the CoMix data influenced the results, it can be observed that the difference between weighted (2.581) and unweighted (2.594) data is almost negligible with a difference in mean contacts per day of 0.013.

3.2.2 Mean comparison by strata

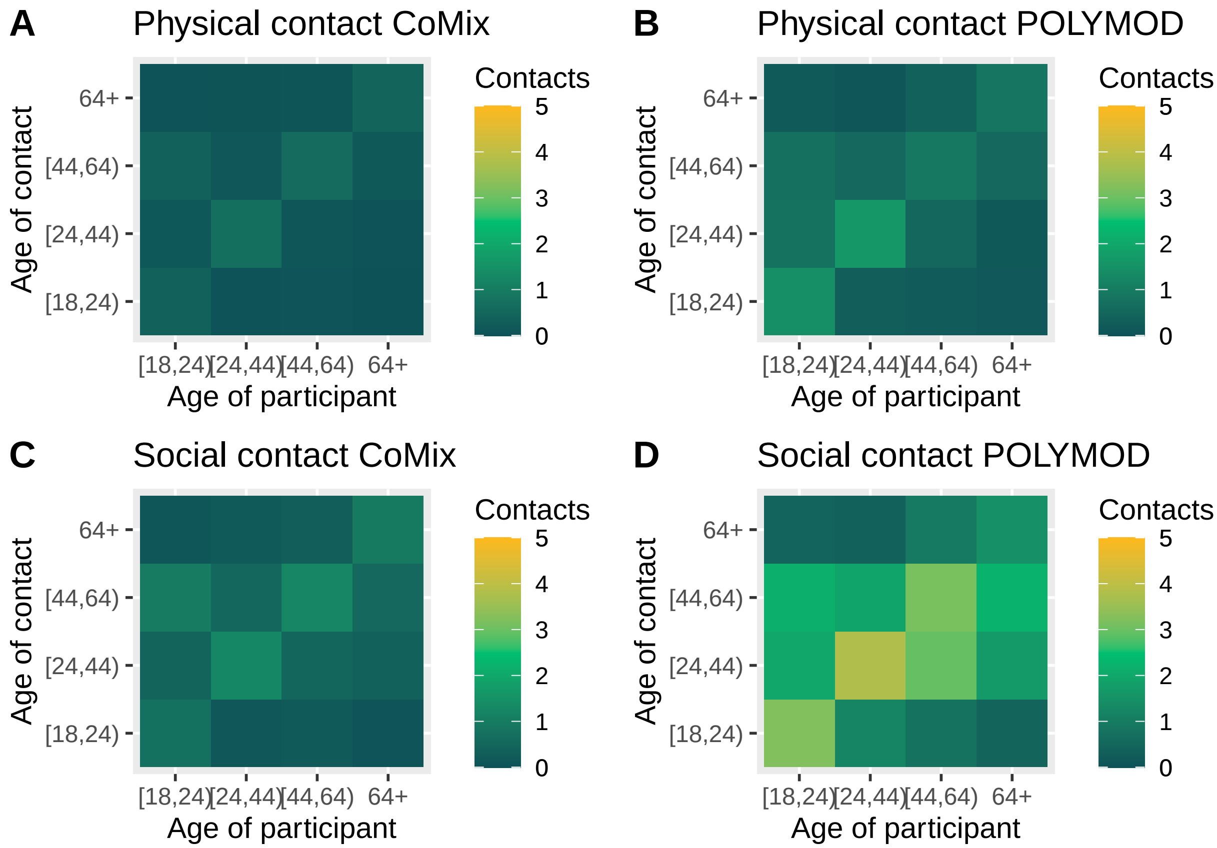
The reduction in mean contacts between POLYMOD and CoMix was also consistent across the stratified subgroups by age, gender and household type. However, the difference in mean contacts between the POLYMOD and CoMix survey was highest for the strata of women between 18 and 24 living in a at least two person household (9.976) and lowest for the subgroup of women over 65 living alone (3.502). In addition, when rounded for three digits there was no difference across the strata between weighted and unweighted data. Therefore, we only continue with the post stratified data in further analysis and neglect the unweighted data.



Matrices - CoMix and POLYMOD - (Overall)

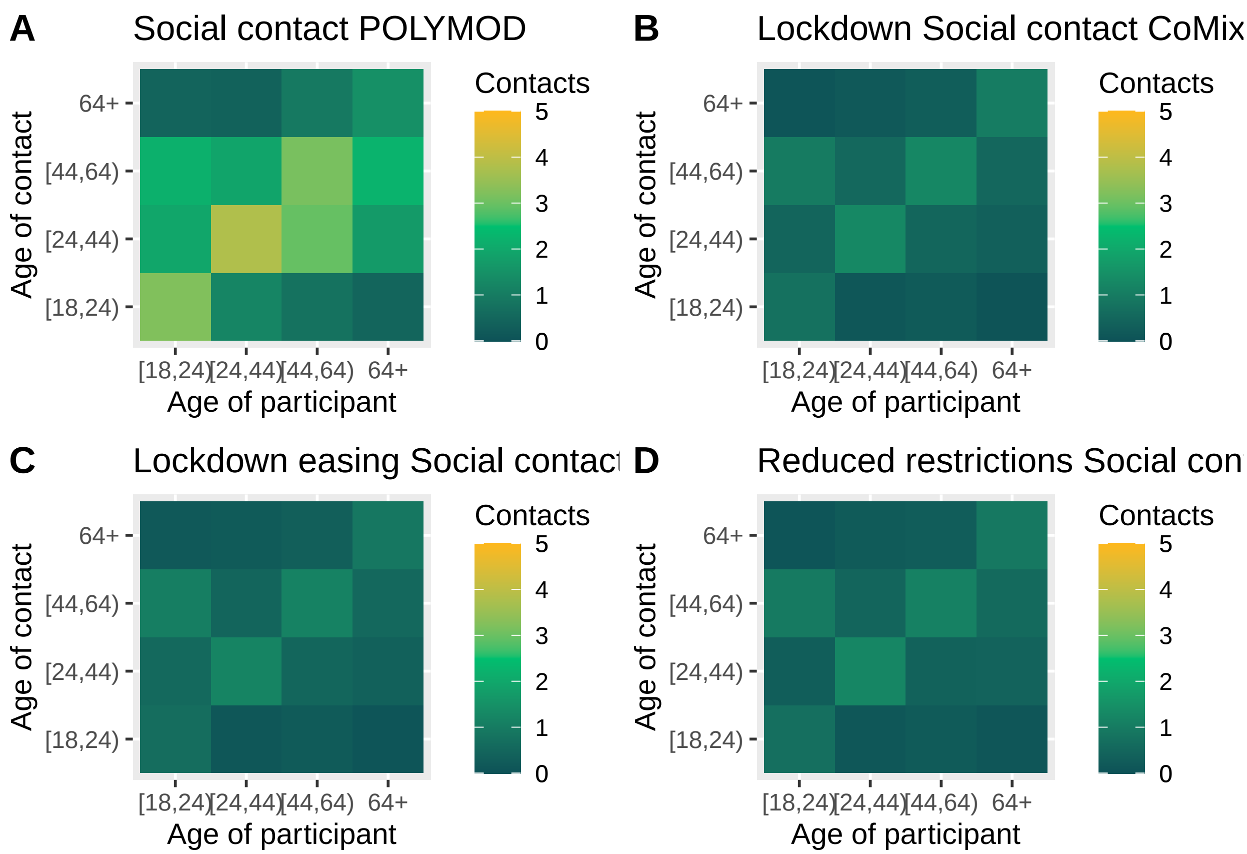
3.2.3 Mean comparison by type of contact

Next, we investigated the difference in the type of contact as previous studies assumed that physical contact is related to the basic reproduction number of COVID-19 (Jarvis et al., 2020). The CoMix data showed a large reduction compared to the POLYMOD data for physical contact as well as non-physical contact. Moreover, the ratio of physical and non-physical contact in the CoMix data showed that physical contact accounted for 45,7 % of all contacts. In line with this, the ratio of physical and non-physical contact in the POLYMOD data showed that physical contact accounted for 41,3 % of all contacts. Therefore, showing only small differences in the observed ratio.



3.2.5 Mean comparison by time period As the CoMix data was conducted over a time frame of several months we further investigated if the mean number of contacts changed due to different stringency levels of non-pharmaceutical interventions. We defined three time periods to compare descriptive statistics and calculate contact matrices. The three periods were categorized as lockdown, lockdown easing, and relaxed restrictions which were adapted from previous work on social contact behavior in the United Kingdom (Gimma et al., 2021). However, the time periods had different lengths and especially the period of reduced restrictions only consisted of a two-week period. Over the three different time periods, there is almost no difference in mean contacts with the lowest mean in the lockdown with easing period (2.562) and the highest mean in the full lockdown period (2.594).

The social contact matrices for each period supported these results and showed as well only small differences in the contact patterns between age groups. Additionally, the differences between participants stratified by gender and age groups showed only small differences between the time periods as well.



4. Discussion ## 4.1 Implications In this assignment, we analyzed a longitudinal social contact survey that assessed changes in contacts from a representative sample of the UK population from the early stages of the COVID-19 pandemic. The investigated period lasted from late March 2020 to mid-August 2020 and included periods of different stringency measures by the government including a full lockdown and reduced restrictions.

4.1.1 Summary of findings

In our analyzes, we mainly focused on the mean contacts as our main outcome variable and analyzed if there are changes in different subgroups or settings. The main results indicate that the mean contact rates remained low throughout the whole time period compared to pre-pandemic surveys regarding social contacts (Mossong et al., 2008). Further, we investigated if there are relationships between the mean number of social contacts measured in the CoMix survey and COVID-19 cases. The results indicated a small positive relationship between the mean number of contacts and new COVID-19 cases. Therefore, showing that the increase in contacts directly leads to an increase in COVID-19 cases, confirming at least to a small extent that there is a connection between the transmission of SARS-CoV2 and the requirement of close contact. Directly related to this finding we examined the relationship between the mean number of social contacts and the stringency of strict physical distancing measures implemented by the UK Government. As these measures instructed the people to stay at home and avoid leaving their house we expected a reduction in social contacts when more stringent government measures are in place (based on government stringency index). This relationship was however not found to be statistically significant. A possible explanation is that measures were usually implemented after the soaring of new COVID-19 cases and as shown before the corresponding increase in the mean number of social contacts. Therefore, the effect might be time-delayed to the actual measurements and changes in social contacts might occur more slowly than the relatively abrupt changes in measures. Another explanation is that the fluctuations in stringency are perhaps too minor over the observed time period in this assignment as the period mainly consisted of full or reduced lockdowns. However, this assumption should still be investigated in future research over a longer time period with more diverse levels of government stringency. Additionally, we looked at three different time periods separating different measures as well (lockdown, lockdown easing, and relaxed restrictions). Regarding this assumption, we did not find a significant difference. A possible explanation is that changes in measures do not always lead to direct changes in social contact behavior. Moreover, this suggests as indicated before that the difference in stringency is too small across the three chosen time periods. On top of this, we investigated if there are differences between distinct subgroups of the population. The results indicated that older people generally have fewer social contacts than younger people and women generally have more contacts than men. Additionally, the number of social contacts seems to depend on whether people live in a single household. However, the differences were still rather small between the subgroups. Nevertheless, these findings can be used to understand the social contact behavior of certain subgroups and the effect of a pandemic on their social lives. Also, it can be used as an additional indication for new government measurements as there are certain trends of social contact behavior in single subgroups.

4.1.2 Pre-pandemic versus COVID-19 pandemic

To further put the results into context and to evaluate if government measures actually led to a decrease in contacts we compared the data to a pre-pandemic assessment of social contact behavior. The results indicate a large (76%) reduction in both physical and non-physical social contacts during the COVID-19 pandemic in comparison to pre-pandemic settings. Therefore, it seems that the introduced measures by the UK government have a huge influence on the social contact behavior of the participants. As we expect the sample to be representative of the UK population, it is very likely that the physical distancing measures are successful in reducing social contacts and corresponding also reducing the basic reproduction number, which should lead to a decline in COVID-19 cases. Moreover, we compared in which locations most contacts are made to understand the change in behavior across a variety of social life settings. First of all, more than 60% of contacts during the pandemic were made at home. Contrasting in pre-pandemic times only about 30% of the contacts were made at home. First of all, indicating that the actual number of social contacts with people not living in the same household is very low. It also indicates that most participants avoided having social contacts out of their homes, further supporting the success of physical distancing measures introduced by the UK government. On top of that, the largest reduction in comparison to the pre-pandemic assessment is in the work setting. This finding further indicates the trend to working from home.

4.2 Limitations

There are several limitations to this assignment. First of all, regarding the survey mode, using a quota-based sample is likely to be biased. For example, only individuals who have consented to participate in marketing surveys are asked, the method is biased towards people with access to the internet and it is likely that individuals who are adhering to physical distancing measures are more likely to respond to this survey. Therefore, it is possible that this survey is flawed regarding selection bias and overestimates certain subgroups of the population. On top of this, the CoMix data only included limited auxiliary data that could be used for our poststratification weighting. Especially, the given socio-demographic information does not change the mentioned flaws regarding the selection bias. Therefore, this could be the reason why the poststratification weighting did not lead to significantly different results. Another limitation is regarding the limited time frame of observation as the pandemic almost lasts two years already and we only examined a period of about five months. Investigating a longer time frame would enhance the chance of drawing strong inferences regarding the social contact behavior while the COVID-19 pandemic. These further analyses could lead to potentially different results regarding the additional datasets like the new COVID-19 cases, the stringency index, and the POLYMOD survey. Therefore, further research is needed to investigate the influence of the COVID-19 pandemic over time, and also further investigations with a probability sample should be conducted.

5. Conclusion

In terms of our aim of figuring out how to do inference from non-probability data, we were successful in implementing our post-stratification design with multiple socio-demographic variables. However, the weighting did not significantly influence our results. On top of that, we were able to find several relationships and differences between social contact behavior and different related data sets like the new COVID-19 cases and the POLYMOD data. Especially, the decrease of mean social contacts during the pandemic in comparison to pre-pandemic periods provides valuable information regarding the change of behavior and the success of government measurements.