Analysis of contact patterns during COVID-19-epidemics

An population density area comparison with survey data

Tim-Yannik Sommer (5399298)

Ronja Alena Steinmetz (5282465)

Xin Yu (4704343)

Chair of Statistics

School of Business and Economics

Freie Universität Berlin

25.10.2021

Contents

List of Figures

List of Tables

Analysis of contact patterns during COVID-19-epidemics

An population density area comparison with survey data

Tim-Yannik Sommer
Ronja Alena Steinmetz
Xin Yu

Abstract

This paper supplements the debate around contact-restriction measures targeting the spread of the novel Sars-Cov-2 Virus by analysing how contact patterns differ geographically between densely, intermediate and thinly populated areas in Germany. The analysis makes use of survey data from the EViPan project collected for four German federate states between March and August 2021. Our research provides evidence for differences in contact patterns and limited evidence for differences in non-household contact patterns by population density area.

1 Introduction

The outbreak of the novel coronavirus (COVID-19) in China in December 2019 soon evolved into a global public health crisis which as of the course of 2020 affected more than 200 countries worldwide and is still pressing in most places. In an effort to handle the situation countries have been undertaking various preventive and emergency measures. These include public health interventions aiming at restricting physical contact for example by enforcing the closure of public locations or by restriction of personal contacts.

The reasoning behind these measures was that virus transmissions depend on the duration and manner of interpersonal interaction. Therefore, most governments decided to limit social contacts to a minimum. However, there are individual social, psychological and economic consequences associated with such strict measures. To justify these political measures, it is thus elementary to investigate broad contact patterns. Further, preventing future super-spreading events has to be a stated goal throughout the pandemic. Hence, knowledge of the transmission of SARS-CoV-2 as well as age and regional-specific infectiousness are of great relevance. For that purpose, social contact surveys can be used to draw conclusions about people's behavior during a pandemic.

Therefore, this paper aims to add another perspective to the debate by looking at differences in contact patterns by different population density areas in Germany. Specifically this paper provides a comparison of the total number, type and age structure of social contacts between different population density areas using survey data from the EViPan project between March and August 2021. The density areas under investigation are densely, intermediate and thinly populated areas.

The remainder of this paper is organized as follows: section 2 reviews the most important literature. Then section 3 presents the methodological approach of this study. The fourth section elaborates on the data. Section 5 discusses the main results and section 6 discusses potential limitations. The last section concludes.

2 Related Work

For policy makers, virologists, and other decision makers to take effective action a need to understand the transmission patterns of the COVID-19 virus and its potential heterogeneity arose. SARS-CoV-2 is primarily transmitted via droplets and aerosols, in a way that person-to-person contacts are a strong determinant of transmission dynamics (Jarvis et al. 2020). Non-pharmaceutical interventions (NPIs) focusing on the reduction of person-to-person contacts are one of the cornerstones of the pandemic response (Tomori, Damilola V. et al. 2021). Suppression of social contact in the workplace, schools, and other public spheres is the goal of these measures (Singh, Rajesh and Ronojoy Adhikari 2020).

Compared to the assessment on after-regulation effects, social contact data gathered through surveys can reflect a more real-time picture of the situation (Tomori, Damilola V. et al. 2021). Hence, contact surveys are an important way to assess social mixing and the impact of control measures such as quarantine, travel restrictions or social distance measures, or blockades in general (Mossong, Joël, et al. 2008). Heterogeneities in social contact networks plays are relevant in determining which interventions should be taken in order to successfully prevent the spread of a pathogen on endemic levels (Prem, Kiesha et al. 2017). The basic idea behind the combination of social contact data with epidemic models has been termed the "social contact hypothesis" (Wallinga, Jacco et al. 2006). Suggesting effective interventions based on epidemic modelling always need to consider the endemic social structure and age-mixing contact patterns since they are always used as a parametric predicator for building up transmission dynamic models (Prem, Kiesha et al. 2017).

Since social contacts have a strong assortative structure in age, the efficacy of these measures is dependent on both, the age structure of the population and the frequency of contacts between age groups across the population (Zhang, Juanjuan et al. 2020). Most papers focused on the contact structuring modelling based on mathematical models such as compartmental models for exploration of evolution in a specific country, as well as having a concentration on the calculation of basic reproduction number R (Ndaïrou, Faïçal et al. 2020, Batistela, Cristiane M. et al.2021).

However, because of geographically specialities, equal measures can have unequal outcomes when applied to regions with significantly differing age and social contact structures (Singh, Rajesh and Ronojoy Adhikari 2020). Furthermore, the asymptomatic transmission of COVID-19 is an important topic in understanding its transmissivity (Qiu, Jane 2020). Therefore, a close look into the regional specific contact patterns throughout the whole COVID-19 period, including population differences as well as regulation specialities, is essentially important for the next step of infectious disease research.

As one can see, the epidemiological literature has come up with several approaches to the analysis of social contact patterns. In this context, Poisson or negative binomial distribution models can be applied (Mossong, Joël, et al. 2008). Kiti et al. (2014) calculated the mean number of contacts in rural and coastal regions in Kenya. To determine different impacts of covariates (e.g. location) on mean contacts, they use an analysis of variance (ANOVA) design. In another study about social mixing patterns and disease transmission in South Africa, Johnstone-Robertson et al. (2011) rejected the assumption of normally distributed data and conducted a nonparametric analysis using Mann-Whitney U and Kruskal-Wallis tests. Indeed, to the best of our knowledge it is not clear which approach is the most appropriate. One can find arguments for and against models. Therefore, we opt for a nonparametric approach in our study and take advantage of the flexibility it provides.

3 Methodology

To examine statistical differences in the number of contacts among areas, we adapt a nonparametric approach. In general, nonparametric analysis requires less stringent assumptions about the distribution. However, it needs to be clarified whether a sample is dependent or independent. Since we test differences between areas individually for each wave, we argue that our sample fulfills the requirement to be independent. Note that in our survey not every participant was interviewed repeatedly, as this would otherwise indicate a blocked study design.

We continue by using the Kruskal-Wallis one-criterion variance analysis by ranks (Kruskal and Wallis 1952). Following Conover (1999, p.290) the null hypothesis tests that all of the k population distributions are identical and, thus can be formulated as:

$$H_0 : P(X_a > X_b) = 0.5$$

for all groups a and b from 1 to k. In contrast, the alternative claims that at least one of the populations yields larger observations than at least one of the other populations:

$$H_1 : P(X_a > X_b) \neq 0.5$$

with $a \neq b$ for at least one of all k groups. Hence, the latter indicates that the groups differ. Note, the test statistic uses a χ^2-distribution (see Conover 1999, p.289). Afterwards, Dunn's test is applied for a multiple comparison between the groups (Dunn 1964). In general, such post-hoc pairwise testing is a common method to explore which groups exactly differ. Hoffman (2019, p.402) emphasises that between the existing statistical tests, Dunn's validation is recommended as it allows for different sample sizes, as it occurs in our case. Similar to the Kruskal-Wallis test, the null hypothesis assumes the samples to originate from populations with identical distributions, otherwise there is statistical evidence for a difference between the groups, what is tested with Dunn's z-test-statistic (see Dunn 1964, p.247). In both tests k is the number of groups (3) and results in k-1 degrees of freedom. Further, we adjust the level of significance to guard ourselves against family-wise error rates using the Bonferroni correction (Dinno 2017). Finally, our statistical analysis will be conducted in R.

4 Data

The data used in this paper originates from the EViPan online survey. It was conducted as part of a larger program of work which aims to provide answers about the epidemiological characteristics of COVID-19. Using the population density categorization from GKPOL in the dataset we are able to explore differences in contact patterns for different population densities (densely, intermediate and thinly) over eighth waves between March and August of 2021. Survey respondents were asked to report the number of people they had contact with in the timespan between 5 AM of the previous day and 5 AM of the survey day. These included contacts of the questionee with individuals residing in the same household, contacts with individuals not residing in the same household and contacts with groups of people. For the first two categories of contacts 14 place options were given[1]. For the latter category contacts met at home, at work and elsewhere were recorded.

In the proceeding analysis the notation *all contacts* refers to the sum of three categories, i.e. contacts with individuals residing in the same household (*household contacts*), contacts with individuals not residing in the same household (*non-household contacts*) and contacts with groups of people *group contacts*.

The total number of observations in this study is 18.644 which contain information on 4.374 individuals from four German federate states, namely Baden-Württemberg, Bayern, Brandenburg and Berlin. The number of individuals observed in each wave can be seen in the top of table 1. It is highest in the beginning (n=2984) and decreases towards the last wave (n=2194) with the exception that wave five has the lowest number of observations (n=1753). Overall there is a loss of observations of around 25% between the first and the last wave which common for data acquired through survey. Details on observations per wave by federal state can be found in figure 4 in the appendix.

Table 1: Descriptives by wave

	3/3 - 10/3 (N=2984)	**31/3 - 7/4 (N=2695)**	**21/4 - 28/4 (N=2419)**	**12/5 - 19/5 (N=2190)**	**2/6 - 9/6 (N=1753)**	**23/6 - 30/6 (N=2209)**	**14/7 - 21/7 (N=2200)**	**4/8 -11/8 (N=2194)**
Age Group								
0-14	493 (17%)	412 (15%)	329 (14%)	291 (13%)	190 (11%)	288 (13%)	276 (13%)	257 (12%)
15-19	164 (5%)	142 (5%)	112 (5%)	102 (5%)	80 (5%)	113 (5%)	97 (4%)	98 (4%)
20-24	165 (6%)	131 (5%)	88 (4%)	68 (3%)	47 (3%)	71 (3%)	71 (3%)	60 (3%)
25-34	337 (11%)	286 (11%)	216 (9%)	197 (9%)	157 (9%)	189 (9%)	196 (9%)	196 (9%)
35-44	286 (10%)	250 (9%)	226 (9%)	181 (8%)	152 (9%)	183 (8%)	203 (9%)	197 (9%)
45-54	402 (13%)	353 (13%)	346 (14%)	325 (15%)	258 (15%)	312 (14%)	328 (15%)	318 (14%)
55-64	518 (17%)	530 (20%)	519 (21%)	487 (22%)	401 (23%)	500 (23%)	497 (23%)	521 (24%)
65-74	515 (17%)	490 (18%)	484 (20%)	448 (20%)	386 (22%)	459 (21%)	446 (20%)	458 (21%)
75+	104 (3%)	99 (4%)	97 (4%)	90 (4%)	82 (5%)	94 (4%)	85 (4%)	89 (4%)
Don't know	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)	0 (0%)
Prefer not to answer	0 (0%)	2 (0%)	2 (0%)	1 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Sex of Participants								
Female	1606 (54%)	1464 (54%)	1266 (52%)	1121 (51%)	874 (50%)	1120 (51%)	1135 (52%)	1121 (51%)
Male	1372 (46%)	1224 (46%)	1150 (48%)	1067 (49%)	875 (50%)	1086 (49%)	1061 (48%)	1068 (49%)
Missing	6 (0.2%)	7 (0.3%)	3 (0.1%)	2 (0.1%)	4 (0.2%)	3 (0.1%)	4 (0.2%)	5 (0.2%)
Household Size								
1	830 (28%)	770 (29%)	697 (29%)	664 (30%)	670 (38%)	672 (30%)	675 (31%)	705 (32%)
2	1212 (41%)	1110 (41%)	1039 (43%)	927 (42%)	675 (39%)	928 (42%)	932 (42%)	921 (42%)
3	494 (17%)	435 (16%)	379 (16%)	339 (15%)	241 (14%)	338 (15%)	343 (16%)	318 (14%)
4+	448 (15%)	380 (14%)	304 (13%)	260 (12%)	167 (10%)	271 (12%)	250 (11%)	250 (11%)
Region of Participants								
Densely populated area	1407 (47%)	1242 (46%)	1115 (46%)	1001 (46%)	820 (47%)	1040 (47%)	1035 (47%)	1023 (47%)
Intermediate density area	888 (30%)	817 (30%)	735 (30%)	686 (31%)	525 (30%)	671 (30%)	684 (31%)	677 (31%)
Thinly populated area	689 (23%)	636 (24%)	569 (24%)	503 (23%)	408 (23%)	498 (23%)	481 (22%)	494 (23%)

Moreover, details on the questionee's age and sex as well as household size by population density area were included in this analysis to understand systematic differences or similarities

[1]Specifically the survey allowed for the following options: Their own home, another persons' home, their workplace, places of worship, transportation, school, essential shops, non-essential shops, places of entertainment, sport facilities, parks, healthcare facilities, beauty places, and an option for other places

between populations (table 1). Compared to the overall age distribution in Germany one can see that the sample is not representative. Generally, elderly individuals are underrepresented when compared to their share in overall population in Germany. After receiving the raw data which contained information on age in five year intervals we decided for broader age groups based on their interpretational value. Each age group is expected to have certain similarities in lifestyle and contact behaviour. The first age group with individuals age 0 to 14 contains smaller children, most of whom are likely in (pre)school and are dependent on adults in their daily life, the second group, consisting of teenagers 15-19 differs in the sense that while still in school are more independent of adults with respect to their ability to have contacts outside of their own homes. The following age groups summarize individuals in higher education or training (age 20-24), young adults with/without small children who are likely to be part of the workforce (age 25-34), adults in the labour force (age 35-44, age 45-54 and age 55-64) and finally retired individuals (age 65-74) and elderly individuals (age 75+). Considering differences between the three density areas the sample population in thinly and intermediate density areas appear mostly similar whereas densely populated areas have fewer younger individuals.

Another difference is the importance of single households. These are generally higher in densely populated areas (39%) compared to intermediate (24%) and thinly populated (22%) areas. Another aspect is that while the share of female vs. male survey participants overall stays mostly the same throughout the eight waves it is rather unequal in thinly populated areas (see table 4 in the appendix) where females are overrepresented with 57%. Furthermore, it is worth noting that while the number of observations between waves varies, they stay the same between population density areas for each wave. Finally, missing values occured most variables. Handling of missing values will be explained in the following sections.

5 Results

In the following section, the number of contacts are described by category and region across all waves in section 5.1. Afterwards, in order to better understand the framework in which these contacts took place, the contact types by place are examined in section 5.2. Afterwards, the comparison of population density areas is used to reveal specific contact patterns between age groups, hence section 5.3 presents detailed contact matrices. Finally, section 5.4 shows the statistical testing.

5.1 Number of social contacts

To give a comprehensive overview, summary statistics will be presented for all contacts, non-household contacts and household contacts. In addition, two modified truncated variables for group contacts are introduced to analyse differences due to potential outliers. Here, we limit group contacts to 50 per age category (under 18, between 18-64, above 64). Thus, a person in a wave met a maximum of 150 contacts for each group contact variable (work, school, else), respectively. Consequently the maximal number of group contacts is limited to 450. Note, that the limit of 50 is set arbitrarily and could be chosen differently by other researchers.

Figure 1: Comparison of contact categories across waves and areas

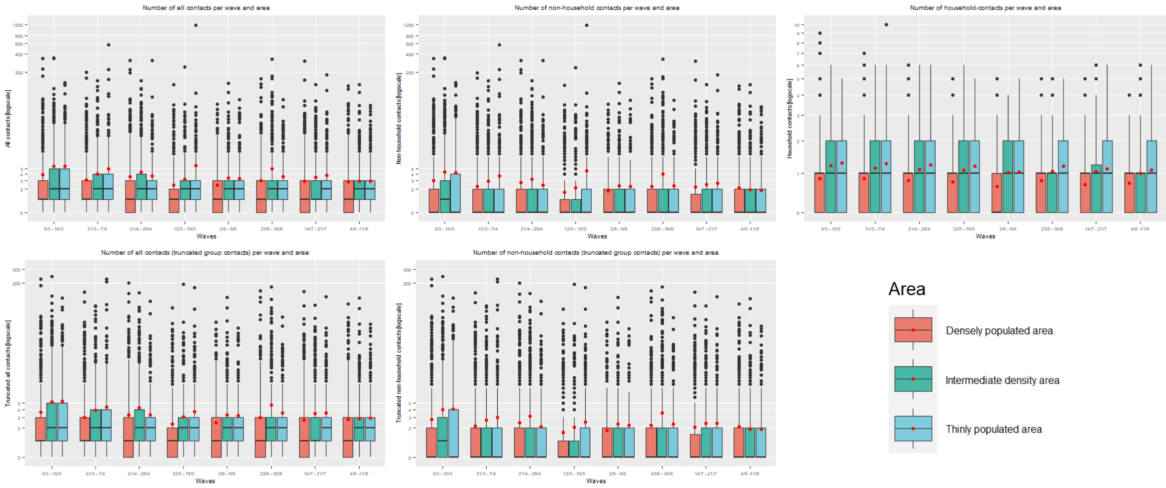


Figure 1 depicts the number of contacts for all regions across the waves. All boxes are characterized at the bottom by the first quartile and at the top by the third quartile for each distribution. Thus, the distance in the box displays the interquartile range. Further, the median represents the middle section of the dataset and is highlighted by a horizontal black line. In addition, red dots represent the respective mean values. The region below the box shows the area from the minimum value to the first quartile and above the box from the third quartile to the maximum value of the dataset. The black dots indicating outliers. Finally, to handle the large range between the observations, a logarithmic transformation is applied to the scale of the y-axis.

Considering all contacts, the medians in densely populated areas are lower compared to the other two areas across all waves. Concurrently, between intermediate and thinly populated areas the medians mostly tend to be equal. A similar difference in the case of densely populated areas seems to hold for the means. However, between intermediate and thinly populated areas no consistent trend can be observed. As one can see in figure 1, the continuous discrepancy between medians and means is striking. Apparently, many participants met very few or no contacts during the pandemic. Therefore, all boxes are closely located above the zero baseline. At the same time, outliers with a large number of contacts occur in every wave, and thus drive the mean upwards. Overall, there appears to be more variation in the data points for the first four waves and decrease over the last four waves.

Similarly to all contacts, an unvarying median can be determined for all regions for non-household contacts except for one deviation in the first wave. In addition, a comparable high interquartile range is observed for five waves, indicating a similar dispersion in the middle half of the data. In contrast, the picture for the mean for non-household contacts is no longer as clear as it was for all contacts. A clearly lower mean of densely populated areas in comparison to intermediate density and thinly populated areas only occurs in the first two waves. Afterwards, the mean values fluctuate in different directions and seem to almost align between the regions in the last four waves.

The difference between all contacts and non-household contacts can be assigned to household contacts. Looking at the medians, all show the same value across the regions with one exception in densely populated areas in wave 5. However, as already described, our data indicates more single households in densely populated areas and large households in thinly populated areas. Therefore, the mean values suggest a clear structure being lowest in densely populated areas, moderate in intermediate densely areas and highest in densely populated areas. In fact, despite potential fluctuation among participants, this pattern for all three regions remain constant.

5.2 Type of social contacts

In order to get a more detailed picture of the differences and similarities of contacts patterns between population density areas over time percentage stack bar plots are presented for all non - household and group contacts per wave for the three density areas (see figure 2). Each bar of the three plots stands for one of the eight waves respectively. The advantage of using percentage stacked bar plots is that they show the contacts by place as share of all contacts in that category whereas each colour stands for a different place. Hence, by using percentage stacked bar plots we are able to display the relative importance of different contact places for each population density area over time. It is important to mention that this way of data visualization does not give information on how contact numbers by place changed in terms of absolute values. It is for example possible that work contacts have a share of 50% in one wave but only 25% in another wave while the absolute number of work contacts is in fact constant. In this example scenario the change in percentage share would be due to an increase in the number of other types of contacts between the waves, hence making the number of work contacts less important relative to other contacts.

As one can see in figure 2 contacts patterns change over time for all three population density areas. Some of the changes can be explained by changes in measures preventing the spread of Covid-19, national holidays or school vacations. For example the importance of meetings at one's own home or someone else's home increase in the second wave for all three density population areas which is likely due to the Easter holidays (2nd to 5th April 2021). Also, contacts in entertainment facilities where not important until the fifth wave due to political restrictions. Furthermore, the importance of contacts happening in school decreases sharply for all three population density areas in the last wave. This observation is likely due to the start of summer vacations.

Figure 2: Non-household and group contacts by place

Most relevant to our research question is how these changes over time differ between population density areas. Interestingly, meetings taking place in essential shops (such as grocery stores) appear to be most important for densely populated areas among the three population density areas. This observation can likely be explained a better accessibility of shops in densely populated areas or the higher share of single households since lager households (especially families) tend to shop in bulk. One's own home appears to be more important for thinly populated areas during the third to fifth wave whereas other people's homes appears to be more important in densely and intermediate populated areas. Generally, this points towards an increase in the importance of meetings in private places during this time which are more difficult to regulate. Interestingly, there appear to be no substantial differences in contact numbers related to work facilities. Finally, differences in the importance of school contacts between intermediate and thinly populated areas on one side and densely populated areas on the other side can most likely be explained with different starting points of summer vacations in different federate states in Germany. A large share of individuals in our data set residing in densely populated areas live in Berlin where summer vacations started on 24th

June. Bayern and Baden Württemberg started school summer vacations around one month later on 29th July and 30th July respectively.

For the sake of consistency an analysis of overall contacts (i.e., non-household and group-contacts including household contacts) using percentage stacked bar plots can be found in figure 5 in the appendix. Compared to the contact pattern for non-household and group contacts we can see a similar pattern with the difference that contacts happening in one's own home are more important. This observation is most likely due to the fact that overall contacts include those contacts with individuals residing in the same household. Therefore, they add little practical value to the analysis. Since policy makers can only restrict non-household contacts we decided that the information given in figure 2 is of greater use concerning our research question.

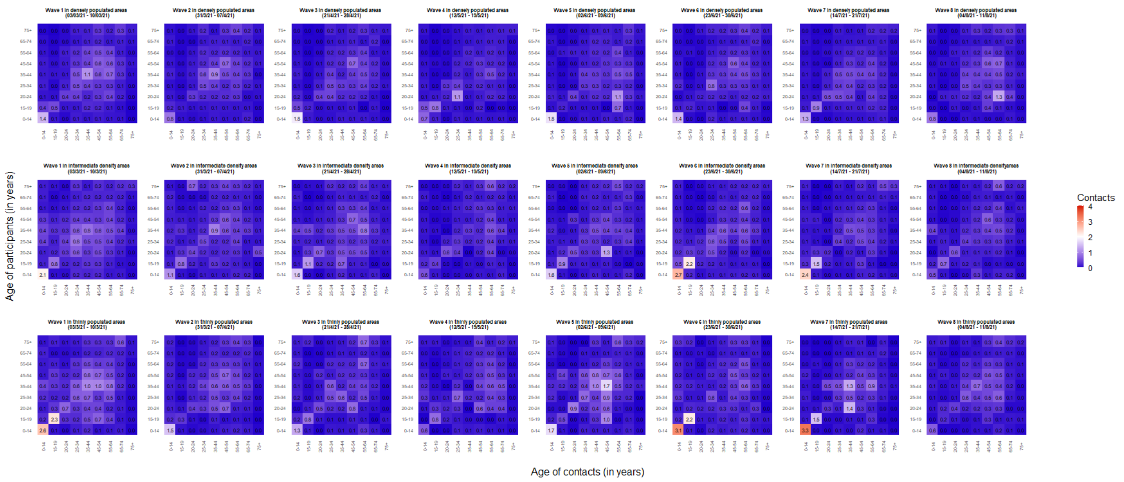
5.3 Contact Matrices

To explore different contact patterns across age groups we again restrict the analysis to non-household contacts. Thus, we compute contact matrices and calculate an average for the number of contacts between the respective age groups. Each contact matrix results from a list of participants and a list of contacts with individuals being linked by an ID number. For that matter, this study uses the 'socialmixr' package in R and the pre-implemented functions in it to construct symmetric contact matrices (Funk 2020).

Note that we made several research-specific decisions. First, certain age groups in the contact matrices could be strongly influenced by outliers. Therefore we continue working with the truncated variable for non-household contacts. Second, imputation methods were used for missing values on the age of participants and social contacts. Whenever no other information is known, a random number between 0 and 90 was sampled for each entry in the data sheet for the age. A similar procedure was applied to the age of group contacts. However, the survey participants estimated the age of group contacts in categories, i.e. a random number was drawn for the corresponding interval (under 18, between 18-64, above 64).

Figure 3 displays the contact matrices for non-household contacts broken down by area for all waves. Note, by the color scale the values are automatically put into a relation to make them more comparable across waves. Following this, blue squares indicate low values and red squares indicate high values. The transition is shown in white. When looking at the calculated average number of non-household contacts two characteristics stand out. First, across all waves the diagonals show the most contacts, i.e. individuals primarily meet contacts in the same age group. This feature is particularly pronounced in the youngest age group (0-14 years) and appears to be lowest in the oldest age group. Second, the areas apart the diagonal should be the focus of attention. In fact, contacts strongly occur in parent-child generations. Here, the generations of the age groups 0-14 and 35-44, 14-19 and 45-54, as well as 20-24 and 55-64 stand out.

Figure 3: Contact matrices for non-household contacts across areas and waves



A pandemic with a changing infection process must always be viewed as a dynamic system. Therefore, certain events affect the pattern in a contact matrix. In the first wave, non-household contacts are relatively high in all three regions. Note, after three months the Covid-19 measures were temporarily relaxed. Hence, a broad purple picture arises across most age groups in all three areas, however this is more widespread in intermediate densely and thinly populated areas, and rather focused in densely populated areas. In wave two, which characterizes both the Easter period and the beginning of the third infection wave in Germany, the number of contacts drops significantly in all three regions. Afterwards, the contact pattern in waves three and four seems rather concentrated on some age groups in densely populated areas, while in intermediate and thinly populated areas it remains to be more strongly in the width. From wave five on, the differences in the breadth of the pattern between the regions seem to decrease. Note, this may be related to the waning of the third wave the advance of the vaccination campaign. Finally, the last wave is characterized by the summer vacations and a predominantly uniform picture between all regions.

5.4 Statistical Testing

We apply the Kruskal-Wallis test and Dunn's test at the level of significance $\alpha = 0.05$ to examine all five contact variables described in section 5.1. To conduct the statistical tests, we use the 'dunn.test' package in R (Dinno, 2017). The results for each test in each wave are presented in table 2 in the appendix. For all contacts, we find in every single wave statistical evidence that the areas are significantly different (χ^2-test: P $<$ 0.001). By looking into the individual comparison, we identify strong stochastic dominance between densely populated and intermediate densely areas as well as densely populated and thinly populated areas. As mentioned earlier, non-household contacts are of particular interest in our study. However, in the case of the non-household contacts, the areas significantly differ only in the first wave, and between densely populated and intermediate densely areas (z-test: P = 0.002) as well as densely populated and thinly populated areas (z-test: P = 0.004). Note that although the corresponding test statistics in waves two and four are close to the rejection region, not enough statistical evidence can be found to reject the null hypothesis. Overall, this shows there is only to some limited extend evidence of differences for non-household contacts between the areas across all waves. Consequently, the differences seem likely to be more attributable to household contacts. Indeed, a similar picture emerges for household contacts as for all contacts. Again, across all waves significant difference in values among the groups can be determined (χ^2-test: P < 0.001). In addition, the test statistics and P values between the non-truncated and truncated variables do not exactly equal, as waves six and seven show slight differences. Note that this is not due to a rounding issue. Hence, in this testing method outliers seem not to affect the results. Finally, for further epidemiological modeling, it can be summarized that there are statistically significant differences between areas with different densities for all contacts. However, we find only limited differences between non-household contacts. These special features would have to be taken into account when making future decisions.

6 Limitations

The results presented in the previous section are subject to a number of limitations. In fact, for our analysis to become traceable we had to simplify and restrict it. First, the results depend to a large extent on the information provided in the contact survey. While some studies use exogenous tracked mobility data, our approach is conditional on the veracity and accuracy of the participants. However, a non-negligible amount of participants declined or simply did not know the age of their contacts. Furthermore, the group contacts represent a large part of the non-household contacts whose age could only be specified imprecisely at large intervals. Both aspects could raise doubts about the accuracy of the contact matrices. To address these issues, we randomly sampled age information to fill missing values. Note that our selected method is not the optimal choice. On one hand, researchers are likely to make use of census data and follow a weighting approach for their observations. On the other hand, they could adopt more sophisticated imputation methods, for example based on the age information given by the participants. Second, social contact structures are individual and vary between regions. Hence, the calculated numbers of contacts can only be transferred to a limited extent. According to that, the contact matrices could probably show different patterns in other countries.

One could potentially expand this study by including more waves, if they are available. In fact, our beginning of the dataset represents the end of Germany's second COVID-19 wave. A longer observation period, i.e. from the beginning of November, would have provided insights into a broader lockdown period in Germany. Even some researchers have looked at comparing contacts before and during the COVID-19 pandemic. Furthermore, in addition to the social contact calculations, a few studies estimated a reproduction number (R_t). Hence, another decision than the political municipality size (GKPOL) could be made for the division of the regions, e.g. a classification by counties. For the robustness of the results, different regional clusters should be subject of future investigations. Moreover, diverging contact patterns between sex or working days and weekends may be part of the research. Finally, our analysis does not account for differences of the vaccination progress or in lockdown rules, the letter federally determined in Germany.

7 Conclusion

The analysis of contact patterns during the COVID-19 pandemic is of high value for decision makers to apply the right measurements to prevent future epidemics. Various methods have been developed in interdisciplinary science to track people's contact behavior. However, only a few studies examine differences between different population density areas.

Using EViPan survey data of eight waves between March and August 2021 from four states in Germany, we investigate the number of contacts, contact types, and contact patterns among age groups by population density areas, namely densely populated, intermediate densely and thinly populated areas - following the GKPOL classification.

Regarding the research question we conclude, there are to some extent statistical significantly differences between different population density areas. Our research provides evidence for differences in contact patterns and limited evidence for differences in non-household contact patterns by population density area. Further, we find social contact patterns to be more concentrated on some age groups in densely populated areas at the beginning of the time span, while in intermediate and thinly populated areas they remain to be more widespread. However, these age-specific differences decrease over time.

Finally, our results depend on several limitations. On the one hand, the quality of the data and the handling of the missing values should be mentioned here. On the other hand, the research-specific decisions we have made can be scrutinised. Hence, these issues should be subject to further analysis.

References

Beraud, Guillaume, et al. (2015):" The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases", PloS one, 10(7), e0133203

Batistela, Cristiane M. et al. (2021): "SIRSi compartmental model for COVID-19 pandemic with immunity loss", *Chaos, Solitons Fractals*, 142, 110388.

Conover, W.J. (1999): ”Practical Nonparametric Statistics”, Third Edition, Wiley, 288-297.

Dinno, Alexis (2017): dunn.test, https://cran.r-project.org/web/packages/dunn.test/dunn.test.pdf, request date: 25.10.2021.

Dunn, O. J. (1964): “Multiple comparisons using rank sums.” *Technometrics*, 6(3):241–252s.

Hoffman, Julien I. E. (2019): Biostatistics for medical and biomedical practitioners, Amsterdam, https://katalog.ub.uni-heidelberg.de/cgi-bin/titel.cgi?katkey=68379808sprache=ENG

Jarvis, Cristopher I. et al. (2020): ”Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK”, *BMC medicine*, 18(1), 1-10.

Kruskal, W. H. and Wallis, A. (1952): Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association.* 47, 583–621.

Meyer, Sebastian and Leonhard Held (2017): ”Incorporating social contact data in spatio-temporal models for infectious disease spread”, *Biostatistics*, 18(2), 338-351.

Mossong Joël et al. (2008): ”Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases”, *PLOS Medicine*,5(3), e74.

Ndaïrou, Faïçal et al. (2020): ”Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan”, *Chaos, Solitons Fractals*, 135, 109846.

Panovska-Griffiths, Jasmina et al. (2020): ”Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study”, *The Lancet Child Adolescent Health*, 4(11), 817-827.

Prem, Kiesha et al. (2017): ”Projecting social contact matrices in 152 countries using contact surveys and demographic data”, *PLoS computational biology*, 13(9), e1005697.

Qiu, Jane. (2020): ”Covert coronavirus infections could be seeding new outbreaks”, *Nature*.

Sharma, Mrinank et al. (2021): ”Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe”, *Nature communications*, 12(1), 1-13.

Singh, Rajesh and Ronojoy Adhikari(2020):*Age-structured impact of social distancing on the COVID-19 epidemic in India.*, arXiv.

Tomori, Damilola V. et al. (2021): *Individual social contact data reflected SARS-CoV-2 transmission dynamics during the first wave in Germany better than population mobility data-an analysis based on the COVIMOD study.*, medRxiv.

Wallinga, Jacco et al. (2006): ”Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents”, *American journal of epidemiology*,164(10), 936-944.

Zhang, Juanjuan et al. (2020): ”Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China”, *Science*, 368(6498), p.1481-1486.

Declaration of Authorship

Ich versichere, dass ich die Seminararbeit selbstständig verfasst habe. Andere als die angegebenen Hilfsmittel und Quellen wurden nicht benutzt. Die Arbeit hat keiner anderen Prüfungsbehörde vorgelegen. Es ist mir bekannt, dass ich bei Verwendung von Inhalten aus dem Internet diese zu kennzeichnen habe und einen Ausdruck davon mit Datum und Internet-Adresse (URL) als Anhang der Seminararbeit beizufügen habe.

T. Sommer

Tim-Yannik Sommer

Ronja Alena Steinmetz

Xin Yu

Berlin, den 25.10.2021

Appendix A

Table 2: Kruskal-Wallis and Dunn's test results for all areas and waves between March and August 2021

	3/3-10/3		**31/3-7/4**		**21/4-28/4**		**12/5-19/5**		**2/6-9/6**		**23/6-30/6**		**14/7-21/7**		**4/8-11/8**	
	T	P value	*T*	P value	*T*	P value	*T*	P value	*T*	P value	*T*	P value	*T*	P value	*T*	P value
All contacts																
K-W test	67.13	<0.001	37.59	<0.001	30.61	<0.001	55.22	<0.001	21.49	<0.001	21.21	<0.001	37.93	<0.001	21.65	<0.001
Dunn's test																
Dens - Inter	-6.52	<0.001	-5.06	<0.001	-4.89	<0.001	-5.33	<0.001	-3.68	<0.001	-3.85	<0.001	-4.79	<0.001	-3.28	0.002
Dens - Thin	-6.96	<0.001	-5.08	<0.001	-4.20	<0.001	-6.78	<0.001	-3.95	<0.001	-3.74	<0.001	-5.32	<0.001	-4.26	<0.001
Inter - Thin	-0.87	0.58	-0.37	>0.999	0.29	>0.999	-1.81	0.10	-0.51	0.91	-0.22	>0.999	-0.96	0.51	-1.20	0.35
N-hh contacts																
K-W test	14.64	<0.001	4.53	0.10	3.32	0.19	4.48	0.11	0.03	0.99	2.95	0.23	2.22	0.33	0.02	0.99
Dunn's test																
Dens - Inter	-3.27	0.002	-2.11	0.05	-1.46	0.21	-1.05	0.44	-0.15	>0.999	-1.68	0.14	-1.05	0.44	-0.08	>0.999
Dens - Thin	-3.02	0.004	-1.07	0.43	0.47	0.95	-2.10	0.05	0.02	>0.999	0.25	>0.999	-1.36	0.26	0.13	>0.999
Inter - Thin	0.00	>0.999	0.81	0.63	1.68	0.14	-1.07	0.43	0.14	>0.999	1.18	0.36	-0.39	>0.999	-0.05	>0.999
Hh contacts																
K-W test	110.16	<0.001	81.74	<0.001	82.56	<0.001	79.87	<0.001	78.92	<0.001	60.54	<0.001	88.09	<0.001	60.22	<0.001
Dunn's test																
Dens - Inter	-8.21	<0.001	-6.69	<0.001	-6.89	<0.001	-7.00	<0.001	-7.41	<0.001	-5.62	<0.001	-7.67	<0.001	-5.83	<0.001
Dens - Thin	-9.03	<0.001	-8.11	<0.001	-8.03	<0.001	-7.74	<0.001	-7.24	<0.001	-7.03	<0.001	-7.78	<0.001	-6.88	<0.001
Inter - Thin	-1.34	0.27	-1.78	0.11	-1.54	0.18	-1.29	0.30	-0.36	>0.999	-1.77	0.12	-0.86	0.58	-1.49	0.20
All c. truncated																
K-W test	67.13	<0.001	37.59	<0.001	30.61	<0.001	55.22	<0.001	21.49	<0.001	21.21	<0.001	37.92*	<0.001	21.65	<0.001
Dunn's test																
Dens - Inter	-6.52	<0.001	-5.06	<0.001	-4.89	<0.001	-5.33	<0.001	-3.68	<0.001	-3.85	<0.001	-4.79	<0.001	-3.28	0.002
Dens - Thin	-6.96	<0.001	-5.08	<0.001	-4.20	<0.001	-6.78	<0.001	-3.95	<0.001	-3.74	<0.001	-5.32	<0.001	-4.26	<0.001
Inter - Thin	-0.87	0.58	-0.37	>0.999	0.29	>0.999	-1.81	0.10	-0.51	0.91	-0.23*	>0.999	-0.96	0.51	-1.20	0.35
N-hh c. truncated																
K-W test	14.64	<0.001	4.53	0.10	3.32	0.19	4.48	0.11	0.03	0.99	2.95	0.23	2.22	0.33	0.02	0.99
Dunn's test																
Dens - Inter	-3.27	0.002	-2.11	0.05	-1.46	0.21	-1.05	0.44	-0.15	>0.999	-1.68	0.14	-1.05	0.44	-0.08	>0.999
Dens - Thin	-3.02	0.004	-1.07	0.43	0.47	0.95	-2.10	0.05	0.02	>0.999	0.25	>0.999	-1.36	0.26	0.13	>0.999
Inter - Thin	0.00	>0.999	0.81	0.63	1.68	0.14	-1.07	0.43	0.14	>0.999	1.18	0.36	-0.39	>0.999	-0.05	>0.999

Notes: For illustration purposes, several variables are abbreviated. From top to bottom (bold): all contacts, non-houshold contacts, household contacts, all contacts with truncated group contacts (50) and non-household contacts with truncated group contacts (50). Below Dunn's test, one finds the area comparisons for densely populated areas, intermediate densely areas and thinly populated areas. The variable T is a proxy for the χ^2 (Kruskal-Wallis) and z (Dunn's) test statistic. The degrees of freedom are calculated from k-1 (k = 3 groups) and yield 2 for all tests. Values with * highlight differences to non-truncated variables.

Figure 4: Descriptives per wave for all four federate states

Baden-Württemberg

	3/3 - 10/3 (N=711)	31/3 - 7/4 (N=631)	21/4 - 28/4 (N=571)	12/5 - 19/5 (N=507)	2/6 - 9/6 (N=408)	23/6 - 30/6 (N=502)	14/7 - 21/7 (N=510)	4/8 -11/8 (N=505)
Age Group								
0-19	193 (27%)	162 (26%)	139 (24%)	122 (24%)	85 (21%)	118 (24%)	118 (23%)	105 (21%)
20-64	427 (60%)	372 (59%)	338 (59%)	300 (59%)	247 (61%)	292 (58%)	310 (61%)	311 (62%)
65-74	75 (11%)	79 (13%)	77 (13%)	70 (14%)	64 (16%)	76 (15%)	70 (14%)	76 (15%)
75+	16 (2%)	16 (3%)	16 (3%)	14 (3%)	12 (3%)	16 (3%)	12 (2%)	13 (3%)
Don't know	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Prefer not to answer	0 (0%)	2 (0%)	1 (0%)	1 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Sex of Participants								
Female	419 (59%)	360 (57%)	318 (56%)	275 (54%)	213 (52%)	266 (53%)	276 (54%)	268 (53%)
Male	291 (41%)	270 (43%)	253 (44%)	232 (46%)	195 (48%)	235 (47%)	234 (46%)	236 (47%)
Missing	1 (0.1%)	1 (0.2%)	0 (0%)	0 (0%)	0 (0%)	1 (0.2%)	0 (0%)	1 (0.2%)
Household Size								
1	159 (22%)	148 (23%)	129 (23%)	117 (23%)	126 (31%)	115 (23%)	116 (23%)	139 (28%)
2	272 (38%)	233 (37%)	228 (40%)	199 (39%)	144 (35%)	205 (41%)	208 (41%)	195 (39%)
3	137 (19%)	124 (20%)	108 (19%)	103 (20%)	74 (18%)	92 (18%)	100 (20%)	81 (16%)
4+	143 (20%)	126 (20%)	106 (19%)	88 (17%)	64 (16%)	90 (18%)	86 (17%)	90 (18%)
Region of Participants								
Densely populated area	140 (20%)	112 (18%)	103 (18%)	86 (17%)	81 (20%)	87 (17%)	99 (19%)	94 (19%)
Intermediate density area	353 (50%)	321 (51%)	288 (50%)	270 (53%)	212 (52%)	263 (52%)	278 (55%)	258 (51%)
Thinly populated area	218 (31%)	198 (31%)	180 (32%)	151 (30%)	115 (28%)	152 (30%)	133 (26%)	153 (30%)

(a) Baden-Württemberg

Bayern

	3/3 - 10/3 (N=780)	31/3 - 7/4 (N=715)	21/4 - 28/4 (N=639)	12/5 - 19/5 (N=587)	2/6 - 9/6 (N=475)	23/6 - 30/6 (N=598)	14/7 - 21/7 (N=590)	4/8 -11/8 (N=594)
Age Group								
0-19	176 (23%)	148 (21%)	113 (18%)	96 (16%)	70 (15%)	117 (20%)	94 (16%)	89 (15%)
20-64	441 (57%)	403 (56%)	360 (56%)	336 (57%)	267 (56%)	339 (57%)	346 (59%)	348 (59%)
65-74	138 (18%)	134 (19%)	131 (21%)	125 (21%)	109 (23%)	113 (19%)	117 (20%)	122 (21%)
75+	25 (3%)	30 (4%)	34 (5%)	30 (5%)	29 (6%)	29 (5%)	33 (6%)	35 (6%)
Don't know	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Prefer not to answer	0 (0%)	0 (0%)	1 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Sex of Participants								
Female	407 (52%)	371 (52%)	315 (49%)	274 (47%)	204 (43%)	278 (47%)	291 (49%)	283 (48%)
Male	371 (48%)	342 (48%)	323 (51%)	312 (53%)	269 (57%)	319 (53%)	297 (51%)	309 (52%)
Missing	2 (0.3%)	2 (0.3%)	1 (0.2%)	1 (0.2%)	2 (0.4%)	1 (0.2%)	2 (0.3%)	2 (0.3%)
Household Size								
1	189 (24%)	176 (25%)	156 (24%)	150 (26%)	150 (32%)	154 (26%)	154 (26%)	164 (28%)
2	325 (42%)	312 (44%)	288 (45%)	271 (46%)	205 (43%)	256 (43%)	256 (43%)	265 (45%)
3	134 (17%)	123 (17%)	110 (17%)	96 (16%)	71 (15%)	96 (16%)	93 (16%)	92 (15%)
4+	132 (17%)	104 (15%)	85 (13%)	70 (12%)	49 (10%)	92 (15%)	87 (15%)	73 (12%)
Region of Participants								
Densely populated area	168 (22%)	144 (20%)	121 (19%)	128 (22%)	97 (20%)	127 (21%)	131 (22%)	126 (21%)
Intermediate density area	296 (38%)	279 (39%)	255 (40%)	223 (38%)	175 (37%)	237 (40%)	221 (37%)	237 (40%)
Thinly populated area	316 (41%)	292 (41%)	263 (41%)	236 (40%)	203 (43%)	234 (39%)	238 (40%)	231 (39%)

(b) Bayern

Berlin

	3/3 - 10/3 (N=985)	31/3 - 7/4 (N=888)	21/4 - 28/4 (N=803)	12/5 - 19/5 (N=705)	2/6 - 9/6 (N=583)	23/6 - 30/6 (N=743)	14/7 - 21/7 (N=728)	4/8 -11/8 (N=718)
Age Group								
0-19	178 (18%)	148 (17%)	117 (15%)	98 (14%)	69 (12%)	98 (13%)	92 (13%)	89 (12%)
20-64	561 (57%)	520 (59%)	469 (58%)	414 (59%)	345 (59%)	438 (59%)	436 (60%)	424 (59%)
65-74	205 (21%)	185 (21%)	186 (23%)	162 (23%)	138 (24%)	175 (24%)	175 (24%)	178 (25%)
75+	41 (4%)	35 (4%)	31 (4%)	31 (4%)	31 (5%)	32 (4%)	24 (3%)	27 (4%)
Don't know	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0%)	0 (0%)
Prefer not to answer	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Sex of Participants								
Female	495 (50%)	468 (53%)	410 (51%)	356 (51%)	301 (52%)	377 (51%)	369 (51%)	372 (52%)
Male	488 (50%)	417 (47%)	391 (49%)	348 (49%)	280 (48%)	365 (49%)	358 (49%)	345 (48%)
Missing	2 (0.2%)	3 (0.3%)	2 (0.2%)	1 (0.1%)	2 (0.3%)	1 (0.1%)	1 (0.1%)	1 (0.1%)
Household Size								
1	349 (35%)	327 (37%)	300 (37%)	284 (40%)	290 (50%)	294 (40%)	292 (40%)	294 (41%)
2	380 (39%)	346 (39%)	327 (41%)	272 (39%)	200 (34%)	302 (41%)	298 (41%)	294 (41%)
3	133 (14%)	109 (12%)	98 (12%)	82 (12%)	55 (9%)	85 (11%)	84 (12%)	74 (10%)
4+	123 (12%)	106 (12%)	78 (10%)	67 (10%)	38 (7%)	62 (8%)	54 (7%)	56 (8%)
Region of Participants								
Densely populated area	985 (100%)	888 (100%)	803 (100%)	705 (100%)	583 (100%)	743 (100%)	728 (100%)	718 (100%)

(c) Berlin

Brandenburg

	3/3 - 10/3 (N=508)	31/3 - 7/4 (N=461)	21/4 - 28/4 (N=406)	12/5 - 19/5 (N=391)	2/6 - 9/6 (N=287)	23/6 - 30/6 (N=366)	14/7 - 21/7 (N=372)	4/8 -11/8 (N=377)
Age Group								
0-19	110 (22%)	96 (21%)	72 (18%)	77 (20%)	46 (16%)	68 (19%)	69 (19%)	72 (19%)
20-64	279 (55%)	255 (55%)	228 (56%)	208 (53%)	156 (54%)	186 (51%)	203 (55%)	209 (55%)
65-74	97 (19%)	92 (20%)	90 (22%)	91 (23%)	75 (26%)	95 (26%)	84 (23%)	82 (22%)
75+	22 (4%)	18 (4%)	16 (4%)	15 (4%)	10 (3%)	17 (5%)	16 (4%)	14 (4%)
Don't know	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Prefer not to answer	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Sex of Participants								
Female	285 (56%)	265 (58%)	223 (55%)	216 (55%)	156 (54%)	199 (54%)	199 (54%)	198 (53%)
Male	222 (44%)	195 (42%)	183 (45%)	175 (45%)	131 (46%)	167 (46%)	172 (46%)	178 (47%)
Missing	1 (0.2%)	1 (0.2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.3%)	1 (0.3%)
Household Size								
1	133 (26%)	119 (26%)	112 (28%)	113 (29%)	104 (36%)	109 (30%)	113 (30%)	108 (29%)
2	235 (46%)	219 (48%)	196 (48%)	185 (47%)	126 (44%)	165 (45%)	170 (46%)	167 (44%)
3	90 (18%)	79 (17%)	63 (16%)	58 (15%)	41 (14%)	65 (18%)	66 (18%)	71 (19%)
4+	50 (10%)	44 (10%)	35 (9%)	35 (9%)	16 (6%)	27 (7%)	23 (6%)	31 (8%)
Region of Participants								
Densely populated area	114 (22%)	98 (21%)	88 (22%)	82 (21%)	59 (21%)	83 (23%)	77 (21%)	85 (23%)
Intermediate density area	239 (47%)	217 (47%)	192 (47%)	193 (49%)	138 (48%)	171 (47%)	185 (50%)	182 (48%)
Thinly populated area	155 (31%)	146 (32%)	126 (31%)	116 (30%)	90 (31%)	112 (31%)	110 (30%)	110 (29%)

(d) Brandenburg

Figure 5: Overall contacts by place

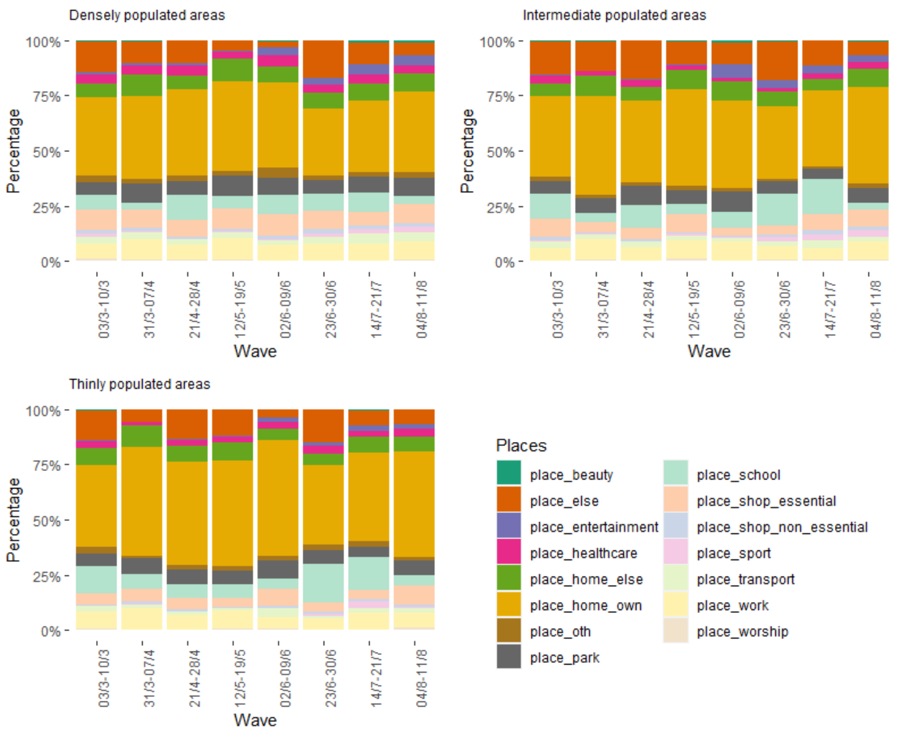


Table 3: Summary statistics by wave and population density area

	Densely populated area	Intermediate density area	Thinly populated area	Densely populated area	Intermediate density area	Thinly populated area	Densely populated area	Intermediate density area
	(N=1407)	(N=888)	(N=689)	(N=1242)	(N=817)	(N=636)	(N=1115)	(N=735)
All contacts								
Mean (SD)	3.87 (13.5)	5.54 (18.4)	5.49 (11.7)	3.15 (9.09)	4.03 (9.67)	5.01 (25.3)	3.56 (14.1)	4.31 (10.7)
Median (Min, Max)	1.00 [0, 335]	2.00 [0, 342]	2.00 [0, 135]	1.00 [0, 200]	2.00 [0, 147]	2.00 [0, 559]	1.00 [0, 313]	2.00 [0, 147]
Median (P25, P75)	1.00 [1.00, 3.00]	2.00 [1.00, 5.00]	2.00 [1.00, 5.00]	1.00 [1.00, 3.00]	2.00 [1.00, 4.00]	2.00 [1.00, 4.00]	1.00 [0, 3.00]	2.00 [1.00, 4.00]
Household contacts								
Mean (SD)	0.861 (1.02)	1.21 (1.11)	1.29 (1.13)	0.849 (0.977)	1.14 (1.07)	1.27 (1.18)	0.808 (0.937)	1.11 (1.02)
Median (Min, Max)	1.00 [0, 9.00]	1.00 [0, 6.00]	1.00 [0, 5.00]	1.00 [0, 7.00]	1.00 [0, 6.00]	1.00 [0, 10.0]	1.00 [0, 6.00]	1.00 [0, 6.00]
Median (P25, P75)	1.00 [0, 1.00]	1.00 [0, 2.00]	1.00 [0, 2.00]	1.00 [0, 1.00]	1.00 [0, 2.00]	1.00 [0, 2.00]	1.00 [0, 1.00]	1.00 [0, 2.00]
Non household contacts								
Mean (SD)	3.01 (13.5)	4.33 (18.4)	4.21 (11.6)	2.30 (9.03)	2.89 (9.61)	3.74 (25.3)	2.75 (14.1)	3.20 (10.6)
Median (Min, Max)	0 [0, 335]	1.00 [0, 342]	0 [0, 135]	0 [0, 197]	0 [0, 146]	0 [0, 558]	0 [0, 312]	0 [0, 146]
Median (P25, P75)	0 [0, 2.00]	1.00 [0, 3.00]	0 [0, 4.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]
All contacts truncated at 50								
Mean (SD)	3.68 (10.4)	5.20 (13.0)	5.34 (10.4)	3.05 (7.48)	3.92 (8.39)	4.33 (13.9)	3.31 (10.1)	4.27 (10.3)
Median (Min, Max)	1.00 [0, 225]	2.00 [0, 242]	2.00 [0, 127]	1.00 [0, 150]	2.00 [0, 100]	2.00 [0, 223]	1.00 [0, 201]	2.00 [0, 147]
Median (P25, P75)	1.00 [1.00, 3.00]	2.00 [1.00, 5.00]	2.00 [1.00, 5.00]	1.00 [1.00, 3.00]	2.00 [1.00, 4.00]	2.00 [1.00, 4.00]	1.00 [0, 3.00]	2.00 [1.00, 4.00]
Non household contacts truncated at 50								
Mean (SD)	2.82 (10.4)	3.99 (13.0)	4.05 (10.2)	2.20 (7.42)	2.78 (8.32)	3.06 (13.9)	2.50 (10.0)	3.16 (10.2)
Median (Min, Max)	0 [0, 225]	1.00 [0, 242]	0 [0, 126]	0 [0, 147]	0 [0, 100]	0 [0, 223]	0 [0, 201]	0 [0, 146]
Median (P25, P75)	0 [0, 2.00]	1.00 [0, 3.00]	0 [0, 4.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]

	Thinly populated area	Densely populated area	Intermediate density area	Thinly populated area	Densely populated area	Intermediate density area	Thinly populated area	Densely populated area
	(N=569)	(N=1001)	(N=686)	(N=503)	(N=820)	(N=525)	(N=408)	(N=1040)
All contacts								
Mean (SD)	3.71 (14.2)	2.43 (6.10)	3.21 (10.5)	5.75 (52.5)	2.47 (4.96)	3.36 (8.66)	3.33 (7.06)	3.05 (8.78)
Median (Min, Max)	2.00 [0, 314]	1.00 [0, 124]	2.00 [0, 241]	2.00 [0, 1172]	1.00 [0, 60.0]	2.00 [0, 130]	2.00 [0, 71.0]	1.00 [0, 156]
Median (P25, P75)	2.00 [1.00, 3.00]	1.00 [0, 2.00]	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]	1.00 [0, 3.00]	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]	1.00 [0, 3.00]
Household contacts								
Mean (SD)	1.24 (1.14)	0.775 (0.887)	1.09 (1.00)	1.19 (1.07)	0.649 (0.854)	1.02 (0.983)	1.04 (0.997)	0.797 (0.888)
Median (Min, Max)	1.00 [0, 6.00]	1.00 [0, 5.00]	1.00 [0, 6.00]	1.00 [0, 5.00]	0 [0, 5.00]	1.00 [0, 4.00]	1.00 [0, 5.00]	1.00 [0, 5.00]
Median (P25, P75)	1.00 [0, 2.00]	1.00 [0, 1.00]	1.00 [0, 2.00]	1.00 [0, 2.00]	0 [0, 1.00]	1.00 [0, 2.00]	1.00 [0, 2.00]	1.00 [0, 1.00]
Non household contacts								
Mean (SD)	2.47 (14.2)	1.65 (5.96)	2.12 (10.5)	4.56 (52.6)	1.82 (4.91)	2.34 (8.60)	2.29 (6.95)	2.25 (8.69)
Median (Min, Max)	0 [0, 314]	0 [0, 122]	0 [0, 240]	0 [0, 1172]	0 [0, 60.0]	0 [0, 130]	0 [0, 70.0]	0 [0, 156]
Median (P25, P75)	0 [0, 2.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]
All contacts truncated at 50								
Mean (SD)	3.36 (7.26)	2.39 (5.53)	3.14 (8.92)	3.77 (10.0)	2.47 (4.96)	3.36 (8.66)	3.28 (6.65)	3.05 (8.78)
Median (Min, Max)	2.00 [0, 114]	1.00 [0, 93.0]	2.00 [0, 191]	2.00 [0, 172]	1.00 [0, 60.0]	2.00 [0, 130]	2.00 [0, 51.0]	1.00 [0, 156]
Median (P25, P75)	2.00 [1.00, 3.00]	1.00 [0, 2.00]	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]	1.00 [0, 3.00]	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]	1.00 [0, 3.00]
Non household contacts truncated at 50								
Mean (SD)	2.12 (7.17)	1.62 (5.39)	2.04 (8.84)	2.57 (10.1)	1.82 (4.91)	2.34 (8.60)	2.24 (6.53)	2.25 (8.69)
Median (Min, Max)	0 [0, 114]	0 [0, 91.0]	0 [0, 190]	0 [0, 172]	0 [0, 60.0]	0 [0, 130]	0 [0, 50.0]	0 [0, 156]
Median (P25, P75)	0 [0, 2.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]

	Intermediate density area	Thinly populated area	Densely populated area	Intermediate density area	Thinly populated area	Densely populated area	Intermediate density area	Thinly populated area
	(N=671)	(N=498)	(N=1035)	(N=684)	(N=481)	(N=1023)	(N=677)	(N=494)
All contacts								
Mean (SD)	5.01 (17.8)	3.59 (6.44)	2.95 (11.6)	3.54 (7.99)	3.77 (10.7)	2.88 (7.41)	2.91 (7.35)	2.97 (5.43)
Median (Min, Max)	2.00 [0, 326]	2.00 [0, 71.0]	1.00 [0, 305]	1.00 [0, 126]	2.00 [0, 181]	1.00 [0, 120]	2.00 [0, 126]	2.00 [0, 57.0]
Median (P25, P75)	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]	1.00 [0, 3.00]	1.00 [1.00, 3.00]	2.00 [1.00, 3.00]	1.00 [0, 3.00]	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]
Household contacts								
Mean (SD)	1.04 (0.954)	1.20 (1.10)	0.708 (0.832)	1.05 (0.971)	1.12 (1.04)	0.730 (0.854)	0.991 (0.973)	1.09 (1.02)
Median (Min, Max)	1.00 [0, 5.00]	1.00 [0, 5.00]	1.00 [0, 5.00]	1.00 [0, 6.00]	1.00 [0, 6.00]	1.00 [0, 4.00]	1.00 [0, 6.00]	1.00 [0, 6.00]
Median (P25, P75)	1.00 [0, 1.00]	1.00 [0, 2.00]	1.00 [0, 1.00]	1.00 [0, 1.25]	1.00 [0, 2.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 2.00]
Non household contacts								
Mean (SD)	3.97 (17.7)	2.39 (6.21)	2.25 (11.5)	2.49 (7.83)	2.65 (10.6)	2.15 (7.38)	1.92 (7.28)	1.88 (5.36)
Median (Min, Max)	0 [0, 326]	0 [0, 70.0]	0 [0, 302]	0 [0, 125]	0 [0, 181]	0 [0, 120]	0 [0, 124]	0 [0, 54.0]
Median (P25, P75)	0 [0, 2.00]	0 [0, 2.00]	0 [0, 1.50]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]
All contacts truncated at 50								
Mean (SD)	4.61 (12.7)	3.59 (6.44)	2.76 (7.60)	3.50 (7.47)	3.56 (8.18)	2.84 (6.78)	2.91 (7.35)	2.97 (5.43)
Median (Min, Max)	2.00 [0, 176]	2.00 [0, 71.0]	1.00 [0, 155]	1.00 [0, 101]	2.00 [0, 131]	1.00 [0, 105]	2.00 [0, 126]	2.00 [0, 57.0]
Median (P25, P75)	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]	1.00 [0, 3.00]	1.00 [1.00, 3.00]	2.00 [1.00, 3.00]	1.00 [0, 3.00]	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]
Non household contacts truncated at 50								
Mean (SD)	3.57 (12.6)	2.39 (6.21)	2.05 (7.50)	2.45 (7.30)	2.44 (8.06)	2.11 (6.74)	1.92 (7.28)	1.88 (5.36)
Median (Min, Max)	0 [0, 176]	0 [0, 70.0]	0 [0, 152]	0 [0, 100]	0 [0, 131]	0 [0, 105]	0 [0, 124]	0 [0, 54.0]
Median (P25, P75)	0 [0, 2.00]	0 [0, 2.00]	0 [0, 1.50]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]

Table 4: Descriptives by population density area

	Densely populated area (N=8683)	**Intermediate density area** (N=5683)	**Thinly populated area** (N=4278)
Age Group			
0-14	972 (11%)	866 (15%)	698 (16%)
15-19	275 (3%)	365 (6%)	268 (6%)
20-24	307 (4%)	226 (4%)	168 (4%)
25-34	949 (11%)	458 (8%)	367 (9%)
35-44	802 (9%)	491 (9%)	385 (9%)
45-54	1213 (14%)	807 (14%)	622 (15%)
55-64	1885 (22%)	1201 (21%)	887 (21%)
65-74	1907 (22%)	1024 (18%)	755 (18%)
75+	371 (4%)	242 (4%)	127 (3%)
Don't know	1 (0%)	0 (0%)	0 (0%)
Prefer not to answer	1 (0%)	3 (0%)	1 (0%)
Sex of Participants			
Female	4296 (50%)	2962 (52%)	2449 (57%)
Male	4371 (50%)	2712 (48%)	1820 (43%)
Missing	16 (0.2%)	9 (0.2%)	9 (0.2%)
Household Size			
1	3388 (39%)	1362 (24%)	933 (22%)
2	3508 (40%)	2440 (43%)	1796 (42%)
3	1040 (12%)	1092 (19%)	755 (18%)
4+	747 (9%)	789 (14%)	794 (19%)
Waves			
3/3 - 10/3 (Wave1)	1407 (16%)	888 (16%)	689 (16%)
31/3 - 7/4 (Wave2)	1242 (14%)	817 (14%)	636 (15%)
21/4 - 28/4 (Wave3)	1115 (13%)	735 (13%)	569 (13%)
12/5 - 19/5 (Wave4)	1001 (12%)	686 (12%)	503 (12%)
2/6 - 9/6 (Wave5)	820 (9%)	525 (9%)	408 (10%)
23/6 - 30/6 (Wave6)	1040 (12%)	671 (12%)	498 (12%)
14/7 - 21/7 (Wave7)	1035 (12%)	684 (12%)	481 (11%)
4/8 -11/8 (Wave8)	1023 (12%)	677 (12%)	494 (12%)