Determinants of SARS-CoV-2 transmission to guide vaccination strategy in a city

Sarah C. Brüningk[1,2,*], Juliane Klatt[1,2,*], Madlen Stange[2,3,5,*], Alfredo Mari[2,3,5], Myrta Brunner[3,4], Tim-Christoph Roloff[2,3,5], Helena M.B. Seth-Smith[2,3,5], Michael Schweitzer[3,5], Karoline Leuzinger[6,15], Kirstine K. Søgaard[3,5], Diana Albertos Torres[3,5], Alexander Gensch[3], Ann-Kathrin Schlotterbeck[3], Christian H. Nickel[7], Nicole Ritz[9], Ulrich Heininger[9], Julia Bielicki[9], Katharina Rentsch[10], Simon Fuchs[12], Roland Bingisser[7], Martin Siegemund[11], Hans Pargger[11], Diana Ciardo[13], Olivier Dubuis[13], Andreas Buser[14], Sarah Tschudin-Sutter[8], Manuel Battegay[8], Rita Schneider-Sliwa[4], Karsten M. Borgwardt[1,2,+], Hans H. Hirsch[6,8,15,+], Adrian Egli[3,5,+,†]

[1]*Machine Learning & Computational Biology, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland*

[2]*Swiss Institute for Bioinformatics (SIB), Lausanne, Switzerland*

[3]*Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland*

[4]*Human Geography, Department of Environmental Sciences, University of Basel, Basel, Switzerland*

[5]*Clinical Bacteriology and Mycology, University Hospital Basel & University of Basel, Basel, Switzerland*

[6]*Clinical Virology, University Hospital Basel & University of Basel, Basel, Switzerland*

[7]*Emergency Department, University Hospital Basel, Basel, Switzerland*

[8]*Infectious Diseases and Hospital Epidemiology, University Hospital Basel & University of Basel,*

Basel, Switzerland

[9]*Paediatric Infectious Diseases and Vaccinology, University of Basel Children's Hospital Basel and University of Basel, Basel, Switzerland*

[10]*Laboratory Medicine, University Hospital Basel, Basel, Switzerland*

[11]*Intensive Care Medicine, University Hospital Basel, Basel, Switzerland*

[12]*Health Services for the City of Basel, Basel, Switzerland*

[13]*Viollier AG, Allschwil, Switzerland*

[14]*Regional Blood Transfusion Service, Swiss Red Cross, Basel, Switzerland*

[15]*Transplantation & Clinical Virology, Department Biomedicine, University of Basel, Basel, Switzerland*

**These authors contributed equally to this work*

+These authors share senior authorship

†corresponding author

Transmission chains within cities provide an important contribution to case burden and economic impact during the ongoing COVID-19 pandemic, and should be a major focus for preventive measures to achieve containment. Here, at very high spatio-temporal resolution, we analysed determinants of SARS-CoV-2 transmission in a medium-sized European city. We combined detailed epidemiological, mobility, and socioeconomic data-sets with whole genome sequencing during the first SARS-CoV-2 wave. Both phylogenetic clustering and compartmental modelling analysis were performed based on the dominating viral variant (B.1-C15324T; 60% of all cases). Here we show that transmissions on the city population level

are driven by the socioeconomically weaker and highly mobile groups. Simulated vaccination scenarios showed that vaccination of a third of the population at 90% efficacy prioritising the latter groups would induce a stronger preventive effect compared to vaccinating exclusively senior population groups first. Our analysis accounts for both social interaction and mobility on the basis of molecularly related cases, thereby providing high confidence estimates of the underlying epidemic dynamics that may readily be translatable to other municipal areas.

Introduction

The COVID-19 pandemic has affected the lives of billions of people globally. Efforts to understand the transmission of SARS-CoV-2 have been undertaken at different scales including at global level[1–3], across continents (Europe and North America[4]), and within countries (such as Austria[5], Brazil[6], France[7], Iceland [8], South Africa[9], and Thailand[10]). The ultimate goal of COVID-19 molecular epidemiological studies is to understand and contain the pandemic. In order to reach this global goal, local interventions are effective to cut transmission chains in family, social, and small communities networks[11,12] and work best in well-defined (phylogenetically or epidemiologically) clusters. However, most infections are acquired from unknown sources and transmitted cryptically along the transmission chain and are the real drivers of the pandemic. Hence, it is essential to identify the key determinants and transmission routes at the city-level, which will ultimately guide the planning of vaccination campaigns. For this, spatially and temporally highly resolved and diverse data are required including detailed information on transmission chains, and on determinants such as demographic, socioeconomic, and interaction patterns that influence both population mobility

and social interaction behaviour[13]. Molecular surveillance and epidemiology are indispensable to fully understand determinants driving transmission. However, such data are difficult to obtain. Studies on SARS-CoV-2 that have focused on selected, larger cities (>300k inhabitants) outside of Europe (Beijing[14], Boston[15], Houston[16], and New York City[17–19] either lack resolution in terms of sequenced positive cases (incomplete transmission chains), or have a paucity of socioeconomic or mobility data that naturally stratify a city (incomplete determinants). Of those studies describing the distribution of cases together with changes in mobility within cities, none to date rigorously study socioeconomic differences between city quarters as determinants of transmission [16,18–20]. An integrated model considering multiple factors (epidemiological, geographic , mobility , socioeconomic, transmission dynamics information) could add profound insights into the determinants underpinning transmission, rather than solely evaluating transmission dynamics. Understanding these determinants will be important for optimizing the delivery of vaccines to those areas and individuals that drive disease transmission and help to maximize the impact of an initially limited amount of vaccine doses.

In this study, we aim to guide vaccination strategies through the identification of determinants of SARS-CoV-2 transmission within a medium-sized European city. Medium-sized cites of less than 300k inhabitants encompass a considerable amount of the urban population (41%)[21] and hence play an important role for the spread of the pandemic. Basel-City (175,350 inhabitants) serves here as a representative example of these. As in most other European cities, urban quarters within Basel-City differ in demographics, socioeconomic factors, housing structure, and mobility. We combine and complement epidemiological, mobility, and socioeconomic datasets with whole genome se-

quencing (WGS) data – all at very high spatial (statistical blocks: a city block partitioned by e.g. streets, rivers) and temporal (day-by-day) resolution for each case reported within Basel-City during the first COVID-19 wave – and serology data to account for the fraction of unreported cases. Specifically, we employ phylogenetic cluster analysis[22] based on transmissions within the dominant virus variant (B.1-C15324T) that was responsible for 60% of the infections in Basel-City[23] during the first wave with an ordinary differential equation (ODE) susceptible-exposed-infected-recovered (SEIR) model. This model includes a city-wide mobility network of real-world measured data to identify routes of transmission. Dynamic variations in the social interaction behaviour of the population of Basel were also included to estimate the temporal change in the effective reproduction numbers. The model fit was performed for partitions of the city based on different socioeconomic determinants (living space per person, share of 1-person households, median income, and share of senior residents). Finally, we applied the SEIR-model for the simulation of different mobility reduction and vaccination scenarios within the city, to demonstrate the impact of a targeted vaccination strategy accounting for the identified determinants.

Results

SARS-CoV-2 spread and clustering within the city. Over the time course of the first wave from 25th February, 2020 until 22nd April, 2020, the University Hospital Basel performed 7073 PCR tests (Basel-City residents only) in context of the obligatory screening of all in-coming patients irrespective of the reason for their hospital visit, as well as from an associated voluntary walk-in test center, hence samples include symptomatic (mild and severe) and asymptomatic cases. Of

these, 750 were positive for SARS-CoV-2 (10.6% positivity rate) (Figure 1A, B) and only 0.7% of the sequenced cases were nosocomial infections. We successfully sequenced the whole-genomes of 411 samples. Epidemiological (occupation in a health service job, resident of a care home, contact to positive cases, onset of symptoms, place of infection) and demographic (age, address, occupation, income) were gathered for most of these cases.

We inferred the lineage identity[24] and observed 29 viral lineages in Basel-City (Figure S8), with 247 genomes (60.0%) belonging to a viral variant that we henceforth refer to as B.1-C15324T[23] (Figure 1A, B). Applying a genetic divergence threshold, a total of 128 phylogenetic clusters, and 70 clusters that belong to lineage B.1-C15324T (Figure 1C) were determined. Mapping this data across the 19 city quarters, we found that most within- and among-quarter transmission clusters were spread randomly, except for significant within-quarter transmissions in four quarters: Riehen, Bruderholz, Am Ring, and Iselin. The results indicate that a significant number of clusters are unique to these latter quarters. Only few quarters shared more clusters between each other than expected by chance as indicated by the coloured links (Figure 1D). Breaking up the statistical blocks within the city into tertiles of socioeconomic determinants (1: low, 2: intermediate, 3: high, N/A: no available data or censored for privacy reasons), we observe that the viruses shared among blocks with either the highest living space per person, lowest share of 1-person households, highest median income, or highest seniority are more likely to belong to the same phylogenetic cluster (Figure 1E). This indicates that largely retired or rather affluent people from Riehen, Bruderholz, Am Ring, and Iselin transmitted the virus in their social networks either in the same quarter or in the same socioeconomic rank. In contrast, positive cases that belong to lower socioeconomic

tertiles (either lower income, less living space, or younger age) are less likely to be members of the same tight phylogenetic cluster indicating cryptic transmission from and to those cases across all socioeconomic levels and across the city.

Spatio-temporal variation of mobility and social interaction patterns. We studied both the spatial and temporal variation in mobility based on a high-resolution, multi-modal traffic model obtained from the traffic department of Basel-City. All following results are presented for a partition of the set of housing blocks into tertiles of low (T1), medium (T2) and high (T3) median income if not declared otherwise (see supplementary material for other socioeconomic partitions). Figure 2A (and Figures S1-S3) depicts this partition, which serves to compute both the intra-tertile and inter-tertile mobility from the Basel-City traffic model[25], and the corresponding mobility graph. Importantly, the statistical blocks per tertile do not form a single, geographically connected, entity. We observe that mobility varies by transport modality and tertile (Figure 2A inset). For example, within T1 the share of private motorized traffic and mobility in general is more pronounced than in the tertiles of higher median income T2 and T3.

The full mobility matrix serves as input to our SEIR model (see methods, eq. (1-7)) to simulate the contribution of mobility to SARS-CoV-2 transmission. The summed edge weights of the mobility graph account for this mobility contribution and the obtained results are shown in Figure 2B for each partition. We observe that low and median income populations are more mobile than their wealthier counterparts. Moreover, there is little mobility within areas with a low share of 1-person households, a result of the predominantly peripheral location of the relevant statistical blocks (see

Figure S5). For living space per person or percentage of senior citizens, mobility was comparable between tertiles with a trend towards higher mobility within the younger population groups.

Dynamic changes in mobility were assessed by agglomerating normalized traffic counts for public and private transport modalities (Figure 2C). There was a clear drop in mobility for both public and private transport modes around the onset of the national lockdown date (12th March, 2020). The decrease was more pronounced for public transport, resulting in a weighted average mobility drop of approximately 50% (Figure 2D). Figure 2D also shows the dynamic change in social interaction contribution to B.1-C15324T case numbers obtained from the estimation of the effective reproductive number using a Kalman filter[26,27]. Despite noticeable fluctuation, social interaction contribution decreased on average over time. This data also reflects variation in case reporting which affected the estimated effective reproductive number. Importantly, since the B.1-C15324T was eventually eradicated, a final social interaction contribution of zero was expected. These results, in terms of spatial and temporal variation, serve as input to the following SEIR-model.

Spatio-temporal spread of the epidemic and its socioeconomic determinants. Using a SEIR epidemiological model considering both sequenced and non-sequenced/unreported cases, as well as mobility and social interactions, the dynamic spread of the first wave was fitted for each of the socioeconomic tertiles with the relevant mobility graphs. A bootstrap approach was used to account for stochastic data uncertainty while we accounted for systematic errors due to the fraction of unreported cases by carrying out a survey for serum SARS-CoV-2 specific IgG amongst samples from Basel residents from that time period. Unreported cases appeared to be a driving force of

the transmission, as 1.9% of 2,019 residents showed detectable antibody levels, corresponding to 77.2% unreported cases overall, and 87.5% for the sequenced B.1-C15324T strain. Figures 3A-C (Figures S4-S6A-C) show the SEIR-model fit to data for each tertile. The corresponding dynamic change of the effective reproductive number (R_{eff}) is given in Figures 3D-F (Figures S4-S6D-F). Independent of the underlying partition, the model provided adequate fits (RMSE < 4.5) and we observe a drop in R_{eff} following the dynamic changes in mobility and social interaction. Importantly, there was a significant difference (2% achieved significance) in R_{eff} between statistical blocks of the highest and lowest median income. The significance level was here scored based on a comparison to 99 random partitioning of the statistical blocks (see Table S1). For all socioeconomic partitions these differences are summarized as histograms in Figures 3G-J. We found that blocks with a higher median income (2% achieved significance), or higher living space per person (1% achieved significance), or lower share of 1-person households (2% achieved significance) had a significantly lower R_{eff} (< 1.7) relative to the maximum R_{eff} observed in the relevant partition. A partitioning based on the share of senior residents did not result in significant differences in R_{eff} (45% achieved significance). Differences in R_{eff} are due to two factors: the effective mobility contribution (Figure 2B), and the modelled reproductive number (R, eq.(8)). In particular, the tertile with the lowest share of 1-person households (T1) showed less mobility compared to T2 and T3, leading to significant differences in R_{eff}. In contrast, mobility in the T1 and T3 tertiles of living space per person were similar (Figure 2B), yet differences in R_{eff} were significant, indicating that the transmission was not dominated by mobility alone.

Impact of mobility changes and modelling of future vaccination scenarios. We simulated the developments of the first wave of the epidemic under the assumption of different mobility scenarios and modelled two future vaccination strategies. All estimated parameters (including R_{eff}, see methods for details) were employed for this scenario building. Figure 4A (Figures S4-S6G) displays the results for mobility scenarios as observed with up to 50% mobility reduction (scenario MO), 100% mobility (scenario M1), and no mobility (scenario M2). Peak case numbers (April 12th) would have been approximately three times higher in the case of no reduction in mobility (M1). However, the decrease in peak case numbers assuming zero mobility (M2) as modelled, was not as pronounced. Mobility reduction hence played a vital role for the containment of SARS-CoV-2 during the first wave.

To study the optimized delivery of vaccines we used data-driven estimates for number of vaccines and their efficacy (70%[28] and 90%[29,30]) in the early 2021 period. Scenarios were compared based on absolute case numbers and on the time to reach 50% of intensive care unit (ICU) capacity as quantifiable endpoint to judge the degree of burden on the healthcare system. We model the scenario of a single exposed individual being introduced into the Basel-City population. Figure 4B (Figures S4-S6H) shows the results for an outbreak scenario (denoted as V1) in which a specific fraction of the population (33% or 66%) will be vaccinated at either 70% or 90% efficacy. As expected, we show that higher vaccine efficacy or higher population fraction vaccinated reduces the slope and plateau (i.e. overall reduction of case burden) of the epidemic curve. Based on these results, vaccination of a third of the city's population at 90% efficacy would cause a decrease in the slope of the epidemic curve resulting in approximately 36 day delay to reach 50% ICU capacity.

It should be noted, that vaccination of the population at random is an artificial scenario, applied here only to demonstrate the impact of vaccination efficacy relative to the population fraction vaccinated. This scenario will serve in the following as a baseline comparison for two more realistic vaccination strategies given in Figure 4C, and D (Figures S4-S6I). Figure 4C shows a scenario where selected demographics with lowest income, that had less options to socially distance and hence were more likely to be exposed to and/or transmit the virus (reflected by a higher R_{eff}, scenario V2), were prioritized for vaccination. With this strategy, the slope of the epidemic curve would be reduced compared to randomly vaccinating the same number of subjects from the whole population leading to a substantial further delay of approximately 30 days to reach the 50% ICU capacity mark (see S9 for more detailed illustration regarding the development of ICU occupancy under the different vaccination scenarios), and an overall reduction of the case burden for the whole city of up to 6.9 folds relative to V1.

Figure 4D shows a scenario where priority was given to the population group with the highest share of senior residents (aged > 64), which had lower mobility than the rest of the population (Figure 2B) but constitutes 60% of ICU cases (scenario V3). We observe that scenario V3 resulted in a steeper epidemic curve and would yield 50% ICU capacity at a similar time as a random vaccination strategy. However, the total number of cases at this time would be approximately double in scenario V3 compared to V1.

Discussion

This analysis evaluates complementary aspects of the spread of SARS-CoV-2 within a medium-sized (<500k residents) city, including local transmission analysed by phylogenetic tree inference and clustering, and the overall spread described by a compartmental SEIR-model. We harness these rich and detailed data sets for the optimization of vaccination strategies within a city. The main strength of this study lies within the high degree of diverse and detailed data available, which included on a per case basis, whole genome sequencing of the virus, the residential address, socioeconomic data, symptom onset (if any), estimated place and time of infection as well as contact tracing information. This was complemented by high resolution, both spatially and temporally, mobility data as well as serology data for the estimation of the fraction of unreported cases, in one comprehensive study. Despite the small size of Basel-City compared to previous city-models that focused on large metropolises[20], we benefit from high data density and quality with respect to the number of residents, and address here a representative example of a European, medium-sized city for which studies are currently lacking.

By basing our modelling efforts only on sequenced data, despite reducing the number of included cases, transmission dynamic analysis did not need to account for potential new introductions. This is a strong advantage of our approach over others[15,31–33] since a continuum model, such as a SEIR-model, is based on the assumption of inherently related cases which is difficult to control in the absence of sequencing information. Another strength of our modelling strategy was the accounting for unreported cases which were the driving force of infection outside the observed clusters. Our 77% estimate of unreported cases overall (not limited to B.1-C15324T) fall within the range of

previous reports within Europe[8,34].

The use of mobility and socioeconomic data in our models is also unique among studies published to date. Our analysis is based on information regularly collected and analysed by the statistical office of Basel city, providing a high spatial and temporal resolution network of the inner city mobility patterns. In contrast to the evaluation of mobile phone data which has frequently been used by others[19,20,35,36] our mobility information is not restricted to groups with a device but represents long-term analysis of different traffic streams. Such data is not subjected to privacy legislation and is hence expected to be more readily available for other cities, making our analysis transferable. We do not hold information on the duration and specific location of individuals, but a continuum estimate of population mixing that aligns well with the concept of a compartmental ODE model. We enrich this information with time-variable data on the degree of capacity utilization for the specific study period.

Based on these data, we evaluated the effective reproductive number for different socioeconomic groups of the population, rather than at a purely geographical level. We found that socioeconomic brackets characterized by low median income and a smaller living space per person, were associated with a significantly larger effective reproductive number than the socioeconomically stronger groups. This observation is in line with previous results[35], suggesting that population groups from a weaker socioeconomic background (here low median income) are more mobile and at higher risk for SARS-CoV-2 infection/transmission. This SEIR-model analysis complements our findings based on the analysis of phylogenetic clusters. We observed clusters predominantly within higher socioeconomic groups, implying that these are spread within the same social network. It

is likely that those individuals are retired, or have had the ability to work from home during the first wave, a pattern that has been observed also in other cities[37]. In contrast, infections within lower socioeconomic groups may result from multiple sources. This observation aligns with the possibility that low socioeconomic status may relate to jobs requiring higher personal contact, and compulsory mobility[38], which has been shown to increase the risk of infection by 76%[39].

Mobility and the reduction thereof has been suggested as a proxy to evaluate the reduction of the spread of SARS-CoV-2 [6,32,36]; however, it is clear that in addition to a change in mobility patterns there has also been a change in the hygiene and social interaction behaviour of the population, encouraged by national campaigns (e.g. in Switzerland[35]). In our SEIR-model we separate these two contributions allowing for an easier translation of our model for scenario building, such as for vaccination. Assuming limited availability of vaccines in the early phase after licensing and market introduction, implying that immunizing the required percentage of population to achieve herd immunity (49% for a perfectly effective vaccine[40] and a basic reproduction number of 1.95 corresponding to the WHO estimate) is infeasible, it is essential to strategically optimize vaccine distribution. According to our model the largest reduction in case numbers would occur in a scenario where priority is given to those population groups that cannot effectively reduce their contacts or that are highly mobile corresponding to low median income groups. The overall reduction in case burden by choosing a targeted distribution is estimated reach up to 6.9 folds compared to a random vaccine distribution strategy. Also, the significant delay of transmission dynamics provides additional time to increase overall vaccination coverage throughout the population. By making a well-informed choice for the distribution of vaccines, an over proportional effect can

be achieved. We showed that the senior citizen population was not driving the transmission of SARS-CoV-2 during the first wave. Hence, although vaccinating high-risk groups would reduce the number of hospitalized and ICU patients in the short term, the spread of the pandemic would be more effectively contained by vaccinating the transmission drivers. A vaccine would be most effective if applied in areas where it is not possible to trace transmission chains, as is the case for mobile and socioeconomically weaker groups. In contrast, we observed that cases within the more senior population groups clustered, indicating that contact tracing strategies would be efficient for this group where transmission chains were detected and contained. By restricting vaccination to risk groups only, a larger fraction of the general population will be exposed to SARS-CoV-2 implying that contact and travel restrictions remain vital to contain transmission. Such measures come at great economic cost[41]. Also, reliable estimations on the efficacy of vaccines in different age groups is yet to be established and it remains uncertain if the vaccines will reach the reported [42,43] high efficacy also in the senior population. We did not account for this effect in our SEIR-model and performed vaccination scenario building accounting only for parameters based on data-driven estimates. However, in the case of a less effective vaccine in senior population groups this would intensify the differences between the simulated scenarios in favour of vaccination of the transmission drivers. Our vaccine scenario simulations are based on a simple, yet data-driven model that is currently specific to Basel-City and we further did not account for vaccination ramp-up, groups with co-morbidity, or differences in vaccine efficacy directly as has been suggested by others [44]. Despite these simplifications our results are in line with previous suggestions[44] and, ideally after independent validation for a different city, provide valuable insight into how transmission dy-

namics could be harnessed to guide vaccination strategy. In conclusion, high-resolution city-level epidemiological studies are essential for understanding factors affecting pandemic transmission chains and delivering effective, thereby supporting tailored public health information campaigns and vaccination distribution strategies at the municipal level. We here provided an example of such an analysis within a European city and suggest that the findings and modelling approaches presented may be readily translated to other cities.

References

1. Yuan, F., Wang, L., Fang, Y. & Wang, L. Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity. *Transboundary and emerging diseases* (2020). URL `http://www.ncbi.nlm.nih.gov/pubmed/33207070`.

2. Nabil, B., Sabrina, B. & Abdelhakim, B. Transmission route and introduction of pandemic SARS-CoV-2 between China, Italy, and Spain. *Journal of medical virology* (2020). URL `http://www.ncbi.nlm.nih.gov/pubmed/32697346` `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7404595`.

3. Hadfield, J. *et al.* NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018). URL `https://pubmed.ncbi.nlm.nih.gov/29790939/`.

4. Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and North America. *Science (New York, N.Y.)* **370**, 564–570 (2020). URL `https://science.sciencemag.org/content/370/6516/564.full`.

5. Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Science translational medicine* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33229462.

6. Candido, D. S. *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science (New York, N.Y.)* **369**, 1255–1260 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32703910 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7402630.

7. Salje, H. *et al.* Estimating the burden of SARS-CoV-2 in France. *Science (New York, N.Y.)* **369**, 208–211 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32404476 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7223792.

8. Gudbjartsson, D. F. *et al.* Spread of SARS-CoV-2 in the Icelandic Population. *New England Journal of Medicine* **382**, 2302–2315 (2020).

9. Post, L. A. *et al.* A SARS-CoV-2 Surveillance System in Sub-Saharan Africa: Modeling Study for Persistence and Transmission to Inform Policy. *Journal of medical Internet research* **22**, e24248 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33211026 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7683024.

10. Puenpa, J. *et al.* Molecular epidemiology of the first wave of severe acute respiratory syndrome coronavirus 2 infection in Thailand in 2020. *Scientific Reports* **10** (2020). URL https://pubmed.ncbi.nlm.nih.gov/33024144/.

11. Koo, J. R. *et al.* Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *The Lancet Infectious Diseases* **20**, 678–688 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S1473309920301626.

12. Mansilla Domínguez, J. M. *et al.* Risk Perception of COVID-19 Community Transmission among the Spanish Population. *International journal of environmental research and public health* **17** (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33276532.

13. Neiderud, C.-J. How urbanization affects the epidemiology of emerging infectious diseases. *Infection ecology & epidemiology* **5**, 27060 (2015). URL http://www.ncbi.nlm.nih.gov/pubmed/26112265 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4481042.

14. Du, P. *et al.* Genomic surveillance of COVID-19 cases in Beijing. *Nature communications* **11**, 5503 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33127911 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7603498.

15. Lemieux, J. *et al.* Phylogenetic analysis of SARS-CoV-2 in the Boston area highlights the role of recurrent importation and superspreading events. *medRxiv : the preprint server for health sciences* **12**, 15 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32869040 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7457619.

16. Long, S. W. *et al.* Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area.

medRxiv (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33024977 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7536878.

17. Maurano, M. T. *et al.* Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City region. *Genome research* **30**, 1781–1788 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33093069.

18. Bushman, D. *et al.* Detection and Genetic Characterization of Community-Based SARS-CoV-2 Infections - New York City, March 2020. *MMWR. Morbidity and mortality weekly report* **69**, 918–922 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32678072 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7366849.

19. Kissler, S. M. *et al.* Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City. *Nature communications* **11**, 4674 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32938924 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7494926.

20. Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* (2020). URL http://www.nature.com/articles/s41586-020-2923-3.

21. United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision* (United Nations, New York, 2019). URL https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

22. Ragonnet-Cronin, M. *et al.* Automated analysis of phylogenetic clusters. *BMC Bioinformatics* **14**, 317 (2013).

23. Stange, M. *et al.* SARS-CoV-2 outbreak in a tri-national urban area is dominated by a B.1 lineage variant linked to mass gathering events. *medRxiv* 2020.09.01.20186155 (2020). URL https://doi.org/10.1101/2020.09.01.20186155.

24. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* **5**, 1403–1407 (2020). URL http://www.nature.com/articles/s41564-020-0770-5.

25. Bau- und Verkehrsdepartement Basel-Stadt Mobilität / Mobilitätsstrategie. Gesamtverkehrsmodell der Region Basel, Basismodell: "Ist-Zustand 2016" (2020).

26. Kalman, R. E. *et al.* Contributions to the theory of optimal control. *Bol. soc. mat. mexicana* **5**, 102–119 (1960).

27. Welch, G., Bishop, G. *et al.* An introduction to the kalman filter (1995).

28. Voysey, M. *et al.* Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *The Lancet* (2020). URL https://doi.org/10.1016/S0140-6736(20)32661-1.

29. Walsh, E. E. *et al.* Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. *The New England journal of medicine*

(2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33053279 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7583697.

30. Mulligan, M. J. *et al.* Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature* **586**, 589–593 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32785213.

31. Alm, E. *et al.* Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Eurosurveillance* **25**, 2001410 (2020). URL https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.32

32. Douglas, J. *et al.* Phylodynamics reveals the role of human travel and contact tracing in controlling COVID-19 in four island nations. *medRxiv* (2020).

33. Murall, C. L. *et al.* Genomic epidemiology of early introductions of sars-cov-2 into the canadian province of québec. URL https://virological.org/t/genomic-epidemiology-of-early-introductions-of-

34. Burki, T. Mass testing for COVID-19. *The Lancet Microbe* **1**, e317 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S2666524720302056.

35. Jay, J. *et al.* Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. *Nature human behaviour* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33144713.

36. Rader, B. *et al.* Crowding and the shape of COVID-19 epidemics. *Nature Medicine* **26**, 1829–1834 (2020). URL nature.com/articles/s41591-020-1104-0#citeas.

37. Wilson, C. These graphs show how covid-19 is ravaging new york city's low-income neighborhoods. *Time* (2020).

38. Reeves, R. & Rothwell, J. Class and covid: How the less affluent face double risks. brookings (march 27) (2020).

39. Rodríguez-Barranco, M. *et al.* The spread of SARS-CoV-2 in Spain: Hygiene habits, sociodemographic profile, mobility patterns and comorbidities. *Environmental Research* **192**, 110223 (2021).

40. Fine, P., Eames, K. & Heymann, D. L. "herd immunity": a rough guide. *Clinical infectious diseases* **52**, 911–916 (2011).

41. Dorn, F. *et al.* The Economic Costs of the Coronavirus Shutdown for Selected European Countries: A Scenario Calculation. Tech. Rep., ifo Institute - Leibniz Institute for Economic Research at the University of Munich, Munich (2020). URL https://www.econpol.eu/sites/default/files/2020-04/EconPol_Policy_Brief_25

42. AstraZeneca. AZD1222 vaccine met primary efficacy endpoint in preventing COVID-19 (2020).

43. Mahase, E. Covid-19: Moderna vaccine is nearly 95% effective, trial involving high risk and elderly people shows. *BMJ: British Medical Journal (Online)* **371** (2020).

44. Bubar, K. M. *et al.* Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *medRxiv* 2020.09.08.20190629 (2020). URL https://www.medrxiv.org/content/10.1101/2020.09.08.20190629v1.

45. Präsidialdepartement des Kantons Basel-Stadt. Statistisches Amt (2020). URL https://www.statistik.bs.ch/.

46. Leuzinger, K. *et al.* Epidemiology of Severe Acute Respiratory Syndrome Coronavirus 2 Emergence Amidst Community-Acquired Respiratory Viruses. *The Journal of infectious diseases* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33180912.

47. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **1**, 33–46 (2017). URL https://pubmed.ncbi.nlm.nih.gov/31565258/.

48. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, 2–4 (2017). URL https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2017.22.13

49. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution* **33**, 1635–1638 (2016). URL https://academic.oup.com/mbe/article/33/6/1635/2579822.

50. Egli, A. *et al.* High-resolution influenza mapping of a city reveals socioeconomic determinants of transmission within and between urban quarters. *bioRxiv* (2020). URL https://www.biorxiv.org/content/early/2020/04/04/2020.04.03.023135. https://www.biorxiv.org/content/early/2020/04/04/2020.04.03.023135.full.p

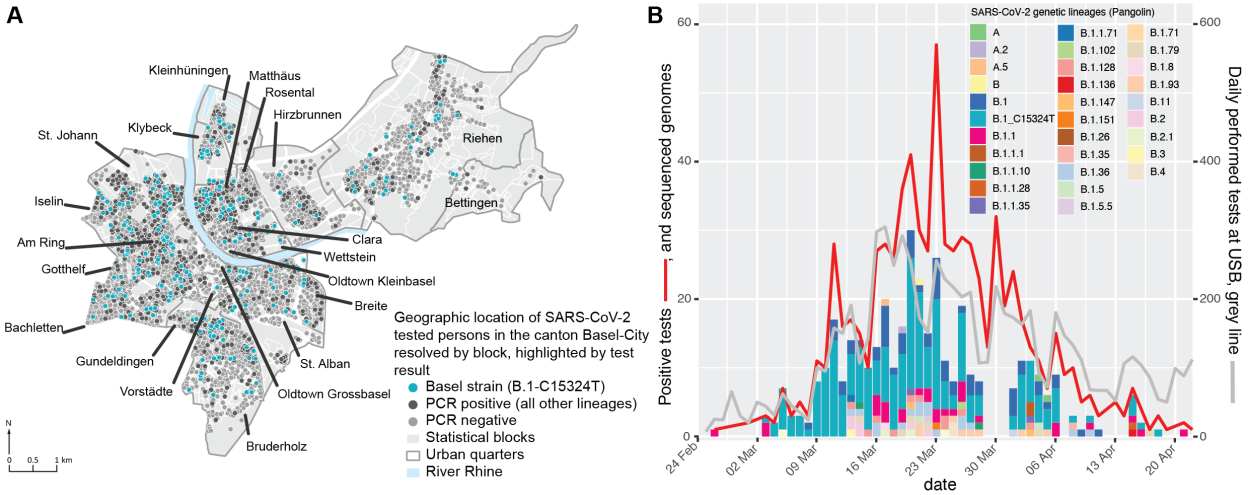
51. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639–45 (2009).

URL http://www.ncbi.nlm.nih.gov/pubmed/19541911 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2752132.

52. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020). URL https://www.sciencemag.org/lookup/doi/10.1126/science.aba9757.

53. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020). URL https://www.sciencemag.org/lookup/doi/10.1126/science.abb3221.

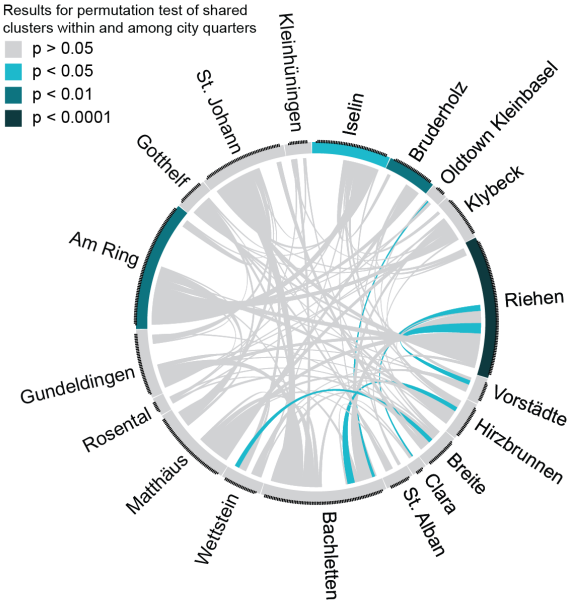
54. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26**, 672–675 (2020). URL https://www.nature.com/articles/s41591-020-0869-5.



C Table.Number of phylogentic clusters in Basel-City quarters within lineage B.1-C15324T (N = 70) and among all intodruced lineages (N = 128). Please note that clusters can be shared among quarters and therefore to not add up to N.

	N clusters in B.1-C15324T	N clusters in Basel-City	N infected persons in B.1-C15324T	N infected persons Basel-City
Oldtown Grossbasel	0	1	0	1
Oldtown Kleinbasel	1	3	1	3
Am Ring	11	19	25	34
Bachletten	15	22	22	34
Breite	6	10	7	13
Bruderholz	6	7	12	15
Clara	2	6	2	7
Gotthelf	8	12	9	15
Gundeldingen	13	25	18	41
Hirzbrunnen	5	9	5	9
Iselin	12	19	19	28
Kleinhüningen	4	5	7	11
Klybeck	8	8	11	11
Matthäus	12	22	19	33
Riehen	15	18	28	35
Rosental	3	11	4	15
St. Alban	6	9	11	19
St. Johann	14	27	23	42
Vorstädte	5	9	6	14
Wettstein	6	6	8	8
w/o adress	8	11	10	23

D Results for permutation test of shared clusters within and among city quarters



E Table. P-values as determined by permutation analysis for within-tertile shared clusters within B.1-C15324T by different socioeconomic factors. All among-tertile transmissions are non-significant indicating that no significant amounf of clusters are shared between different socioeconomic groups.

tertile	1	2	3	N/A
Living space per person	0.240	0.300	0.009	0.999
1-person households	0.001	0.287	0.295	0.999
Median income	0.222	0.458	0.002	0.458
Seniority	0.223	0.068	0.012	1.000

Figure 1: SARS-CoV-2 transmission in quarters and among demographic and socioeconomic groups during the first COVID-10 wave in Basel-City. A) Spatial case/non-case distribution throughout the city with the most dominant SARS-CoV-2 variant (B.1-C15324T), the focus of this study, highlighted in turquoise; B) epidemiological curve for Basel-City and distribution of phylogenetic lineages (pangolin nomenclature) from 25th of February to 22nd of April, 2020; C) summary for inferred phylogenetic clusters within (i) all lineages and (ii) the major variant B.1-C15324T in the different city quarters. High number of infected people within a quarter with a low number of clusters indicates presence of large transmission clusters whereas large number of clusters and low number of people infected within quarter indicates random infections and cryptic transmission; D) visualisation of a significance test for transmission within and among quarters; E) results of a significance test for transmission between tertiles (1: low, 2: intermediate, 3: high, N/A: no available data or censored for privacy reasons) of different socioeconomic factors.

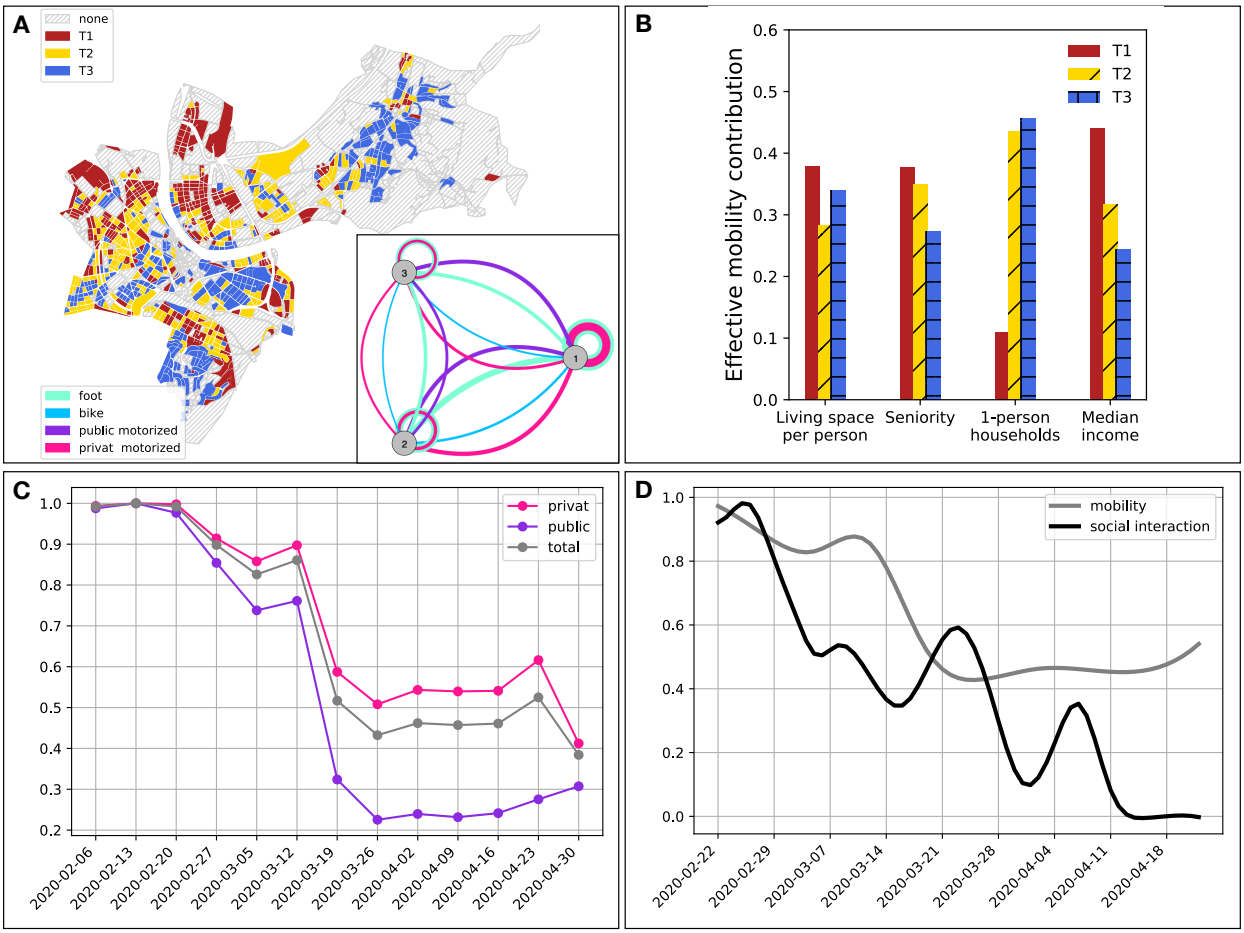


Figure 2: Spatio-temporal variation of mobility patterns within the Canton of Basel City. A: The Canton of Basel City and its delineation with respect to statistical blocks colored according to the partition into tertiles T1, T2, and T3 of increasing median income as provided by the canton's office for statistics. Inset: resulting mobility-graph, with nodes representing tertiles and edge-widths representing strength of effective connectedness through mobility by means of various modes of transport, as computed from the traffic-model provided by the traffic department of the Canton of Basel City. B: Relative mean contribution of mobility to a socioeconomic tertile's effective reproductive number associated with the major variant B.1-C15324T, by socioeconomic indicator. C: Normalized temporal development of private and public transport as well as their weighted sum during the first wave of the pandemic in the Canton of Basel City as obtained by the traffic department and public transport companies, respectively. D: Smoothed relative temporal development of social interaction and mobility contribution to the effective reproductive number associated with the major variant B.1-C15324T within the Canton of Basel City.

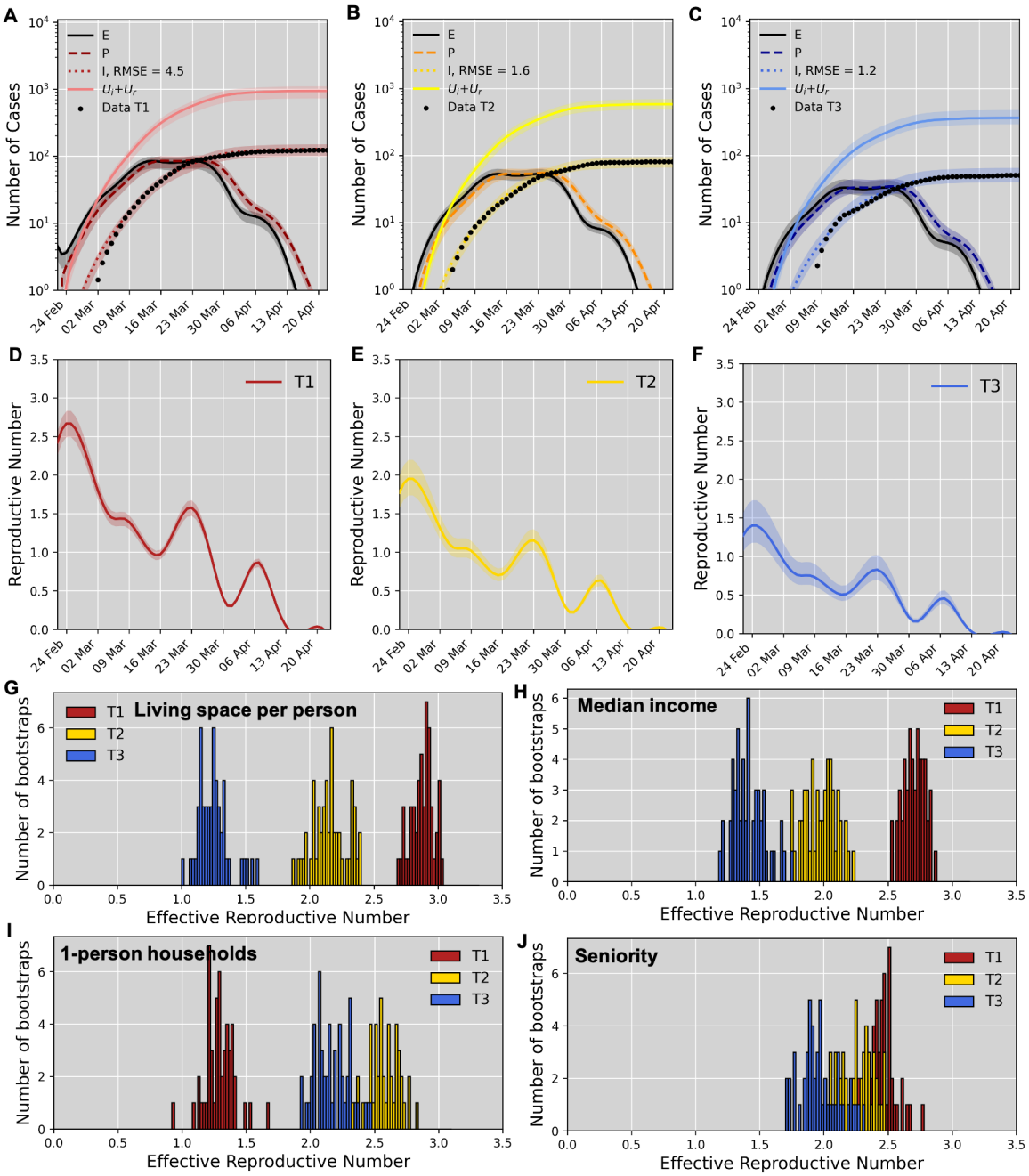


Figure 3: Model fit to the case number time-series. A-C) Fit results for a partition based on median income. Data points are shown together with model predictions based on undisturbed data (solid lines), and fifty bootstraps from disturbed data (bands) for the different tertiles T1 (low, A), T2 (intermediate, B) and T3 (high median income, C). D-F) The dynamic variation of the effective reproductive number for each of the tertiles shown in A-C. G-J) Histograms over all bootstraps for the effective, pre-lockdown reproductive number for each socioeconomic partition. Results are shown for partitions based on living space per person (G), median income (H), share of 1-person households (I), and share of senior residents (J).

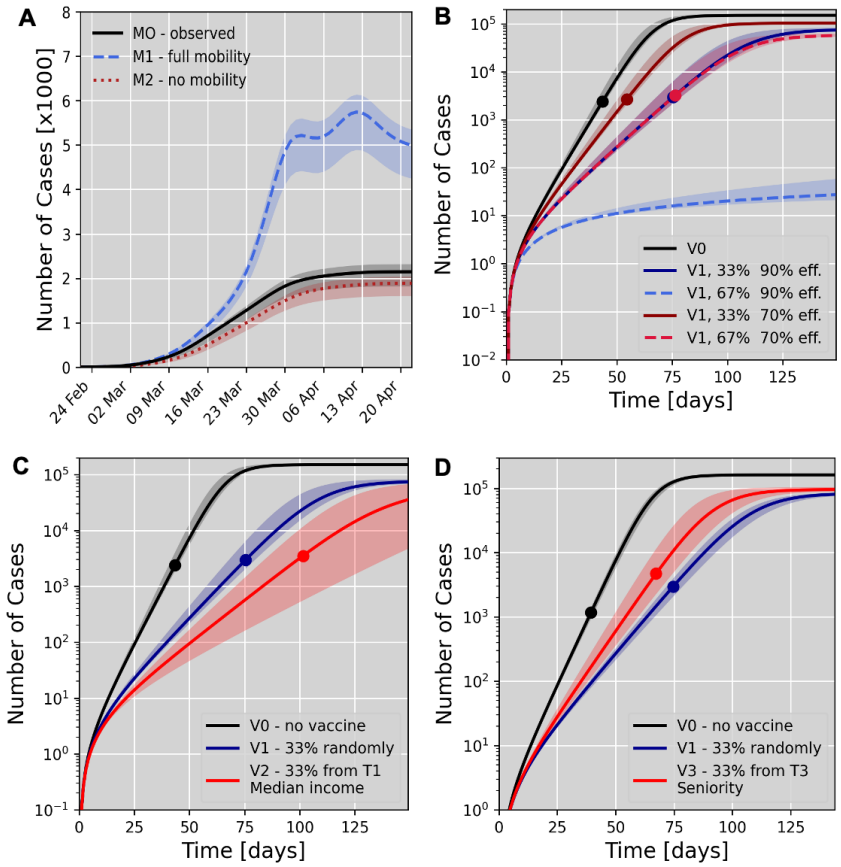


Figure 4: Scenario Simulations for a partition based on median income. A) Influence of the mobility pattern on the total number of infected cases during the first wave (sum of reported and unreported cases) modelling either no change in mobility (no lockdown scenario, M1), or full shut-down of all inner city mobility (M2). For comparison the observed scenario (MO) is shown. B) Simulation of future vaccination effects if a specific percentage of all citizens was randomly selected for vaccination at given efficacy (V1). We compare this to the scenario of no vaccine (V0). C) Simulation of future vaccination effects based on a partition according to median income. Scenario V2 models vaccination of 33% of all citizens selected from the tertile with the lowest median income (T1). A maximum 6.9 fold difference in cases was observed at day 99. D) Simulation of future vaccination effects based on a partition according to the share of senior residents. Scenario V3 models vaccination of 33% of all citizens selected from the tertile with the highest share of senior residents (T3). We model 90% vaccine efficacy and compare with scenarios V0 and V1. In C) and D) we model 90% vaccine efficacy and compare with scenarios V0 and V1. Dots indicate the time of reaching a 50% ICU occupancy.

Methods

Ethical statement. Ethical approval was given by the local ethical committee *Ethik Kommission Nordwest und Zentralschweiz* (EKNZ No. 2020-00769, to be found at https://ongoingprojects.swissethics.ch) and the project was registered at clinicaltrial.gov under NCT04351503.

City data. The Canton of Basel-City, an administrative area consisting of Basel, Bettingen, and Riehen, is located in the North-Western part of Switzerland, bordering Germany and France, and comprises 36.95 square kilometers in size. An estimated 34,000 commuters cross the borders from and to France and Germany daily[45], together with another 85,000 inner-Swiss commuters[45]. Basel-City is divided into 19 urban quarters and had a mean population of 175,350 inhabitants during the last five years (+/- 1,737 inhabitants) (Figure S7). For subsequent analyses, a total of 1,078 statistical (housing) blocks (a city block partitioned by e.g. streets, rivers) were identified within the city quarters. The University hospital Basel is a tertiary hospital with a capacity of 700 regular and 44 intensive care unit (ICU) beds.

PCR testing and whole-genome sequencing. PCR testing was available rapidly and frequent testing was established and supported by local guidelines by the end of February, 2020, before the first case arrived[46]. In total 7073 PCR tests from Basel-City residents were performed at the University hospital Basel (750 of which were SARS-CoV-2 positive) dating between 25th of February and 22nd of April, 2020. The ratio of negative to positive PCR tests changed during the local epidemic with a median of 10.6% positive PCR tests (Figure 1B). We successfully sequenced SARS-

CoV-2 whole genomes from 411 unique patients (54.8% of all cases). Of these, 247 (247/411, 60%) could be attributed to contain the C15324T mutation in the B.1 lineage (and therefore called B.1-C15324T) characteristic to the virus variant that originated in this tri-national area[23].

Geographic mapping and socioeconomic stratification. Each individual PCR test from (N = 7073), irrespective of the result, was linked to the patient's place of residence anonymized at the scale of statistical blocks in ArcMap 10.7 (by ESRI). To explore the statistical association between SARS-CoV-2 transmission and socioeconomic factors, we employed data provided by the Canton of Basel City's office for statistics for the year of 2017 (most recent data available), that specified the values for various socioeconomic indicators for each statistical block (except for those blocks where privacy legislation did not permit the sharing of such information). The indicators under study were (i) the living space (per capita in m^2), (ii) the share of 1-person private households, (iii) the median income (CHF), and (iv) the population seniority (percentage of senior citizens aged over 64 years per block). According to these socioeconomic indicators, blocks were allocated to one of three socioeconomic city tertiles (T1:$\leq$33rd percentile , T2: 33rd to 66th percentile T3:$>$66th percentile) where possible (e.g. Figure 2A). In general, sparsely populated blocks displayed a maximum of three positive cases and had to be excluded from analysis. All following analyses with respect to socioeconomic factors were based on these city partitions.

Phylogenetic inference and cluster analysis. Whole-SARS-CoV-2-genomes from Basel-City patients were assembled using our custom analysis pipeline COVGAP[23] (github.com/appliedmicrobiologyresearch CoV-2/). Global sequences and metadata were downloaded from GISAID[47,48] (as of October

22,2020; 155,278 consensus sequences). Sequences with more than 10 percent N's (27,013) and with incomplete dates (43,466) were removed. 84,799 sequences remained which were joined with Basel genomes. Filtering for the period of interest until April 22nd retained 39,913 genomes, of which 411 are from Basel-City residents dating from February 26th (first case) to April 22nd, 2020. To infer relatedness among the viral genomes and spread of SARS-CoV-2 in Basel-City, a time-calibrated phylogeny that was rooted to the first cases in Wuhan, China from December 2019, was inferred using a subset of the global genomes. For subsetting, we included 30 genomes per country and month, whereby all genomes from Basel-City were retained, totalling 3,495 genomes, using the nextstrain software v.2.0.0 (nextstrain.org) and augur v.8.0.0[3] as described in detail in [23].

The resulting global phylogeny was used to infer phylogenetic clusters in Basel-City. First, polytomies, which are caused by identical genomes in the tree were resolved using ETE3 v.3.1.1[49]. Cluster Picker v.1.2.3 [22] was then used to identify clusters in the resolved tree (options 0, 0, 4e-4, 5). Identified clusters were consolidated with epidemiological data (occupation in a health service job, resident of a care home, contact to positive cases, onset of symptoms, place of infection) to confirm the suitability of the divergence parameter. Cluster Matcher v.1.2.4[22] was then used to combine ancillary geographic (quarter), and socioeconomic (share of 1-person households, median income, living space, seniority) information that were subdivided into tertiles (1: low, 2: intermediate, 3: high, N/A: no available data or censored for privacy reasons) on identified clusters.

To test whether related genomes in Basel-City cluster according to a) quarter, b) living space per person, c) share of 1-person households, d) median income, or e) seniority a custom python-script for a random permutation test was performed[50] (github.com/appliedmicrobiologyresearch/Influenza-

2016-2017). The results for clustering within and among urban quarter and tertiles in socioeconomic determinants were visualized using circos v.0.69[51].

Serology. We determined SARS-CoV-2 antibody responses in a total of 2,019 serum samples collected from individuals between 25th of February and 22nd of May, 2020, to account for seroconversion. The cohorts included healthy blood donors and patients from the University Hospital Basel with residency in Basel-City, and Viollier AG Zurich, a lab medicine provider contracted for the SARS-CoV-2 PCR testing. We used SARS-CoV-2 serology testing to detect specific IgG and IgM antibodies (Cobas6800, Roche). This is an electro-chemiluminescence immunoassay (ECLIA) that targets the nucleocapsid. Serology information was used to estimate the fraction of unreported cases as follows: An estimated 1.88% (38/2,019) of the Basel-City population was infected with SARS-CoV-2. Of these 60% would be attributed to the B.1-C15324T strain, leading to a percentage of 88% of unreported/unsequenced cases to consider.

Mobility data. We employed the official traffic model provided by the traffic department of Basel-City [25]. The latter consists of the 2016 average A-to-B traffic on a grid of $\sim$1400 counting zones for four transport modalities: foot, bike, public motorized transport and private motorized transport. We further obtained weekly averages of pass-by traffic for the same count zones over the period of the first wave of the pandemic for the categories of combined foot and bike traffic, as well as private motorized traffic. Additionally, weekly public-transport passenger loads were provided by both the Swiss Federal Rail Company and the local public transport services. From these datasets, we computed the spatio-temporal variation of mobility within the city as follows.

Spatial variation was obtained by aggregating A-to-B traffic between the aforementioned counting zones, first to the statistical block level (by identifying the nearest block with respect to a zone's centroid), and second to tertile level via a statistical-block's association with a socioeconomic indicator tertile. This resulted in a there-by-three mobility matrix M_{jk} whose diagonal entries represent within-tertile mobility, while off-diagonal entries represent inter-tertile mobility. This matrix was normalized to one since only relative differences were relevant in our model. Temporal variation was obtained by computing the weighted sum of the private transport time-series provided by the traffic department and the public transport time-series provided the Swiss Federal Rail Company and the local public transport services. This sum was then normalized and smoothed with a uni-variate spline resulting the final time-series for temporal mobility variation employed in our model (denoted as $\alpha_{mob}(t)$), see Figure 2 D).

Dynamic changes in social interaction. SARS-CoV-2 transmission is contact-based. While the number of contacts potentially taking place within a day and a city is largely influenced by human mobility as estimated above, the risk of a contact becoming a transmission event is further determined by the precautions taken by the two individuals being in contact (such as washing hands, wearing masks, distance keeping). Both aspects together–mobility and risk-mitigating social behaviour within a (sub-)population–eventually result in an effective, time-dependent, reproductive number characterizing the virus's transmission within that (sub-)population. Hence, there are three relevant time-series: changes in the overall effective reproductive number, in mobility, and in social behaviour. While the computation of the temporal variations in mobility was described above,

the time-dependent effective reproductive number is obtained by means of applying a Kalman filter [26,27] to the piece-wise linearised time-series of daily confirmed cases of individuals having contracted the B.1-C15324T variant of SARS-CoV-2. Assuming a multiplicative model, the time-dependence of residual transmission risk stemming from lack of precaution in social interaction (denoted as $\alpha_{soc}(t)$), is obtained by point-wise division of the time-dependence of the effective reproductive number by the mobility time-series (depicted in Figure 2 D). Thus we are adhering to the logic that in the extreme case of zero mobility, no transmission can take place despite a finite risk of transmission rooted in a lack of precautions, while on the other hand in the case of zero risk of transmission due to perfect precautions, no transmission can take place despite non-zero mobility. Such logic dictates the choice of a multiplicative rather than additive model.

SEIR-model. Similar to previous work [20,52,53] the spread of the epidemic within the city of Basel was described using a compartmental two-arm SEIR model including both sequenced, and unsequencend/unreported cases. We accounted for susceptibles (S), exposed (E, incubation time T_{inc}), and pre-symptomatic yet infectious cases independent of reported/unreported status (P, reproductive number R). After a presymptomatic time T_{infP}, cases were separated according to the estimated proportion of reported and sequenced cases, p_{sq}, into either reported infectious (I), or unreported infectious (U_i, reproductive number R). Since our data did not include information on recovered patients, a 'recovered' compartment was not included following compartment I. It was assumed that reported cases remained isolated. The unreported compartment on the other hand transitions to recovery (U_r) after an infectious time T_{infU}. To allow for connectivity and exchange

between the defined partitions, cross contamination was included through the mobility matrix M_{jk} described above. Over the course of the studied period the Swiss government imposed a partial lockdown over the country resulting in notable temporal variation of mobility and social interaction patterns of the population of Basel-City which we account for with two time dependent weighting factors $\alpha_{mob}(t)$ and $\alpha_{soc}(t)$ introduced above. In summary this resulted in the following system of ODEs for the spread within each socioeconomic tertile j:

$$\frac{dS_j}{dt} = -R_j \alpha_{soc}(t) \alpha_{mob}(t) \frac{S_j}{N_j} \cdot \left[\sum_{k=1}^{3} M_{j,k} \cdot \left(\frac{P_k}{T_{\text{infP}}} + \frac{U_k}{T_{\text{infU}}} \right) \right] \tag{1}$$

$$\frac{dE_j}{dt} = R_j \alpha_{soc}(t) \alpha_{mob}(t) \frac{S_j}{N_j} \cdot \left[\sum_{k=1}^{3} M_{j,k} \cdot \left(\frac{P_k}{T_{\text{infP}}} + \frac{U_k}{T_{\text{infU}}} \right) \right] - \frac{E_j}{T_{\text{inc}}} \tag{2}$$

$$\frac{dP_j}{dt} = \frac{E_j}{T_{\text{inc}}} - \frac{P_j}{T_{\text{infP}}} \tag{3}$$

$$\frac{dI_j}{dt} = (1 - p_{sq}) \frac{P_j}{T_{\text{infP}}} \tag{4}$$

$$\frac{dUi_j}{dt} = p_{sq} \frac{P_j}{T_{\text{infP}}} - \frac{I_j}{T_{\text{infU}}} \tag{5}$$

$$\frac{dUr_j}{dt} = \frac{I_j}{T_{\text{infU}}} \tag{6}$$

For all compartments, the initial number of susceptibles was fixed to the relevant population. All other compartments were initialized as zeros, with the exception of a seed in the exposed compartment corresponding to the first reported cases. In summary, our model is based only on six free model parameters, including the respective reproductive number per tertile R_j (three parameters, range $[0, 20]$), the initial number of exposed in a single tertile (range $[0, 5]$), and the infectious times T_{infU} and T_{infP}(range $[1.5, 12]$ days). We assumed a latency period T_{inc} and infectious time prior to symptom onset T_{infP} of two days each [54].

Fitting procedure and evaluation. In total 247 cases within the time period from the 25th of February until the 22nd of April were included in this analysis. For all data a seven day moving window average was taken to account for reporting bias on weekends, and cumulative numbers of infected cases (compartment I) were calculated. Due to the loss of single sequencing plate, missing numbers on the 29th, 30th and 31st of March were imputed by assuming a constant ratio of the B.1-C15324T strain amongst the sequenced samples. Simulations were initialized on the 22nd of February, the estimated date of the occurrence of the initial exposed cases[23].

The ODE system was implemented in python (version 3.8.) using the scipy functions *odeint* to iteratively solve the system of equations and *minimize* (with L-BFGS algorithm, cut-off tolerance of 10^{-7}) for parameter fitting based on the average sum of squared differences between the logarithm of estimated and recorded cases. Points with cumulative case numbers below 15 were not scored in the cost-function since the continuum assumption of the model may not be satisfied for small case numbers. The fit was performed simultaneously for all four socioeconomic partitions to account for the shared parameters T_{infU} (obtained 1.8 days) and T_{infP} (obtained 2.1 days). Fit performance was evaluated based on root mean squared error (RMSE) between predicted and recorded case numbers. Data uncertainty and the corresponding influence on parameter estimation were assessed by bootstrapping. For each of 50 bootstraps each data point was randomly shifted according to numbers drawn from a normal distribution centered around zero and standard deviation 0.3. Results are shown as mean values over bootstraps and bootstrap uncertainty bands. We compare effective reproductive numbers corresponding to the normalization of R by the the

effective mobility contribution ($\sum_k M_{jk}$):

$$R_{eff,j} = \frac{R_j}{\sum_k M_{jk}} \tag{7}$$

Scenario simulation. The impact of mobility relative to social interaction was analysed by re-calculating the predicted epidemic trajectory under the constraint of constant intra-urban mobility ($\alpha_{mob}(t) = 1$, scenario M1) or fully restricted ($\alpha_{mob}(t) = 0$, scenario M2) mobility, corresponding to perfect isolation of the affected city areas. These scenarios were compared to the baseline of the actual reduction in mobility (scenario M0).

Vaccination scenarios were simulated as for both 90% and 70% effective vaccines resembling current vaccine candidate data [42,43]. This was achieved by moving the fraction of the vaccinated population from the susceptible to the recovered compartment U_r and calculating the spread of the pandemic with constant effective reproductive number and intra-city mobility. We accounted for a change in social interaction behaviour following vaccination by assigning a mean social interaction score of the vaccinated and not-vaccinated population ($\alpha_{soc,vacc}(t) = 1$, $\alpha_{soc,novacc}(t) = 0.5$). Mobility was modelled as 100% ($\alpha_{mob}(t) = 1$). Two scenarios were investigated and compared to the no vaccine scenario (V0): i) vaccination of a fixed population fraction (one or two thirds) randomly throughout the population (scenario V1), ii) vaccination of the corresponding number of individuals from the socioeconomic city areas presenting the highest reproductive number (scenario V2). These were compared to the simulation of no vaccination (scenario V0). In order gauge the benefit of a particular vaccination scenario, we calculated the time to reach 50% of intensive care unit (ICU) capacity. The University Hospital Basel has a total of 44 ICU beds. During the first

wave, 4.5% of reported SARS-CoV-2 positive cases were admitted to ICU, and their median length of ICU stay was 5.9 days (IQR, 1.5-12.9). If considering additional unreported cases (captured by serological testing), the percentage of patients requiring ICU admission was around 1%. Of all SARS-CoV-2 patients with ICU stay, 40% were younger than 64 years resulting in a probability of an under 64 year old infected case to be admitted to ICU of 0.5%.

Acknowledgements We greatly appreciate the input and data received from Construction- and Traffic department Canton Basel-City, Baselland Transport AG, Basler Verkehrs-Betriebe, Autobus AG, SBB AG, Statistical Office of the Canton of Basel-City, and want to specifically thank Björn Lietzke (Statistical Office of the Canton of Basel-City), Madeleine Imhof (Statistical Office of the Canton of Basel-City), Michael Redle (Construction- and Traffic department Canton Basel-City), Kathrin Grotrian (Construction- and Traffic department Canton Basel-City), Matthias Hofmann (Basler Verkehrs-Betrieb), Roman Stingelin (Autobus AG), Nadine Ruch (SBB AG) and Stefan Burtschi (Baselland Transport AG) for their support. We thank Christine Kiessling, Magdalena Schneider, Elisabeth Schultheiss, Clarisse Straub, and Rosa-Maria Vesco (University Hospital Basel) for excellent technical assistance with next generation sequencing. Computations were performed at sciCORE (http://scicore.unibas.ch) scientific computing facility at the University of Basel.

We thank all authors, who have shared their genomic data on GISAID, especially the Stadler Lab from ETH Zurich for sharing Swiss sequences. A full table (csv) outlining the originating and submitting labs is included as a supplementary file.

Author contributions AE, HHH, KB devised the project. SCB and JK developed and performed the mathematical modelling. MSt performed and interpreted phylogenetic analyses. SCB, JK and MSt led the

writing and revising of the report. AM provided the genome assembly pipeline. TR and HSS prepared viral RNA for sequencing. MyB and RSS provided geographical expertise. KKS collected clinical data. KR, DAT, AG, AKS, MSch analysed serology samples. DC and OD provided serology samples from Viollier AG. KL provided virological expertise. AB provided serology samples from the blood transfusion service. JB, STS and SF provided public health and epidemiological expertise. HP, MSi, CHN, RB, MB provided clinical expertise and valuable discussion on the results. NR, UH, JB provided epidemiological expertise. All author reviewed and edited the manuscript.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to Adrian Egli, adrian.egli@usb.ch, Clinical Bacteriology and Mycology, University Hospital Basel, Petersgraben, CH-4031 Basel, Switzerland, Phone: +41 61 556 57 49.

Data Availability The SEIR-model code used for this submission will be available on https://github.com/BorgwardtLab/BaselEpi.git. Code that was used for phylogenetic inference and calculation of significance of clusters in specified groups is available at https://github.com/appliedmicrobiologyresearch. SARS-CoV-2 whole genomes from Basel-City are available at gisaid.com and at European Nucleotide Archive (ENA) under accession number PRJEB39887.

Supplementary Material

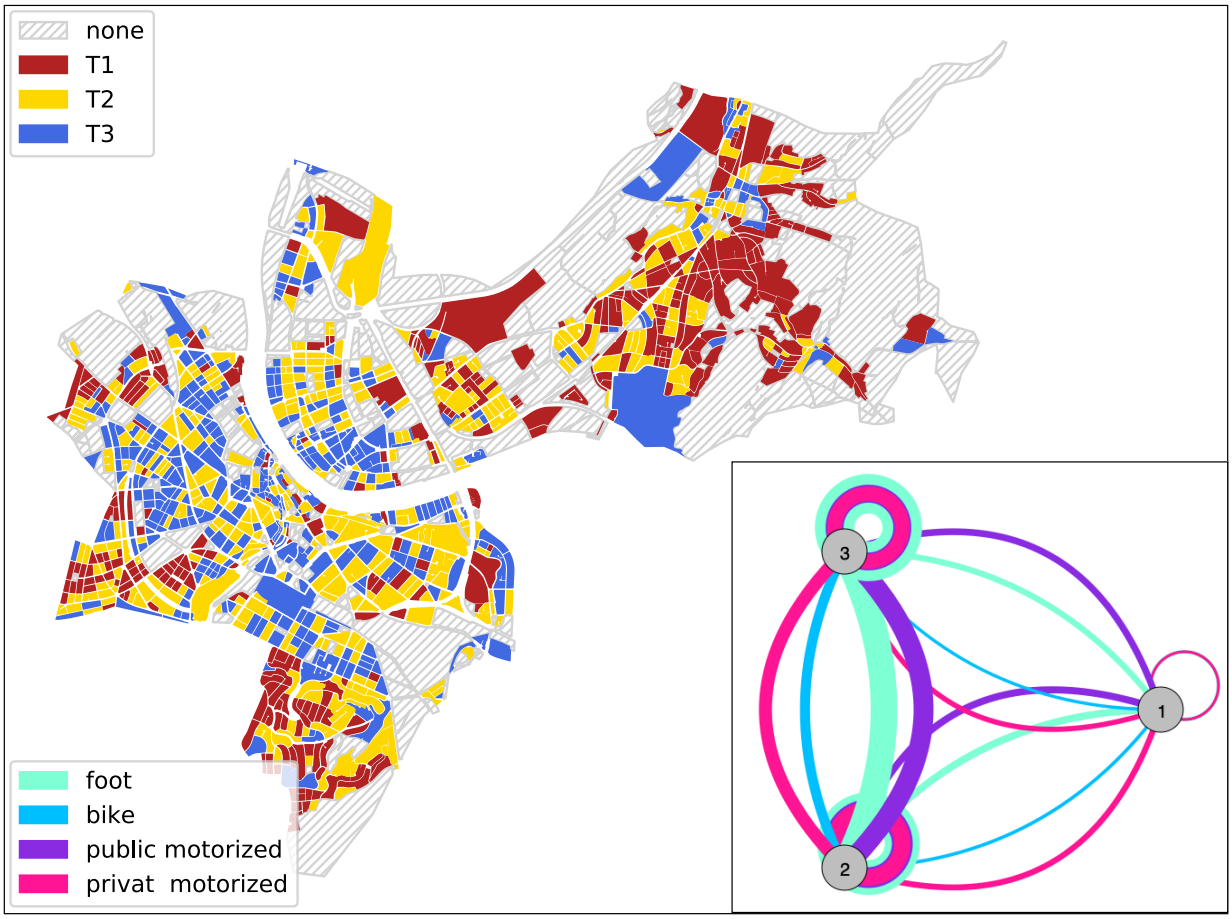


Figure S1: The Canton of Basel City and its delineation with respect to statistical blocks colored according to the partition into tertiles T1, T2, and T3 of increasing fraction of 1-person households per block as provided by the canton's office for statistics. Inset: resulting mobility-graph, with nodes representing tertiles and edges representing effective connectedness through mobility by means of various modes of transport (thicker/thinner edges indicating weaker/stronger connectedness), as computed from the traffic-model provided by the traffic department of the Canton of Basel City.

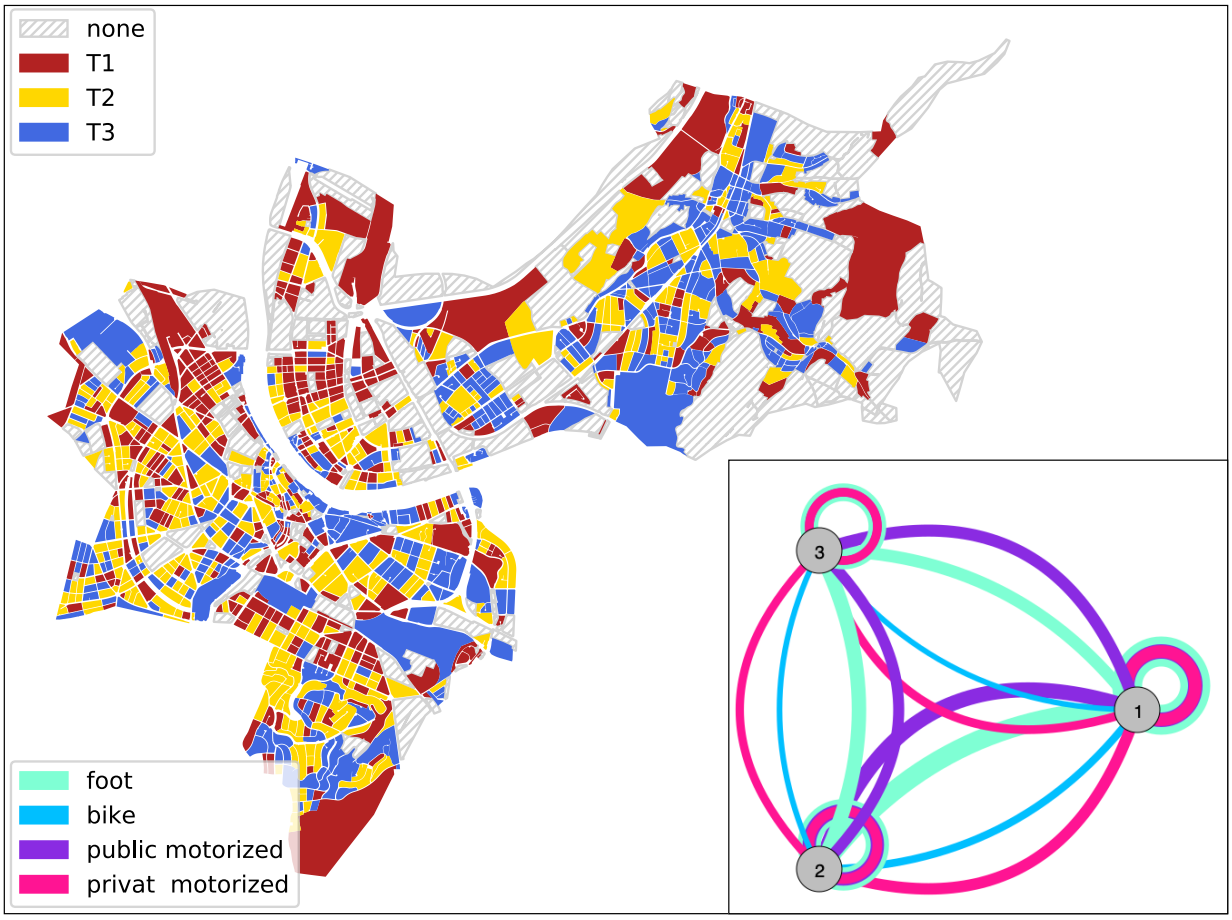


Figure S2: The Canton of Basel City and its delineation with respect to statistical blocks colored according to the partition into tertiles T1, T2, and T3 of increasing fraction of residents aged older than 64 per block as provided by the canton's office for statistics. Inset: resulting mobility-graph, with nodes representing tertiles and edges representing effective connectedness through mobility by means of various modes of transport (thicker/thinner edges indicating weaker/stronger connectedness), as computed from the traffic-model provided by the traffic department of the Canton of Basel City.

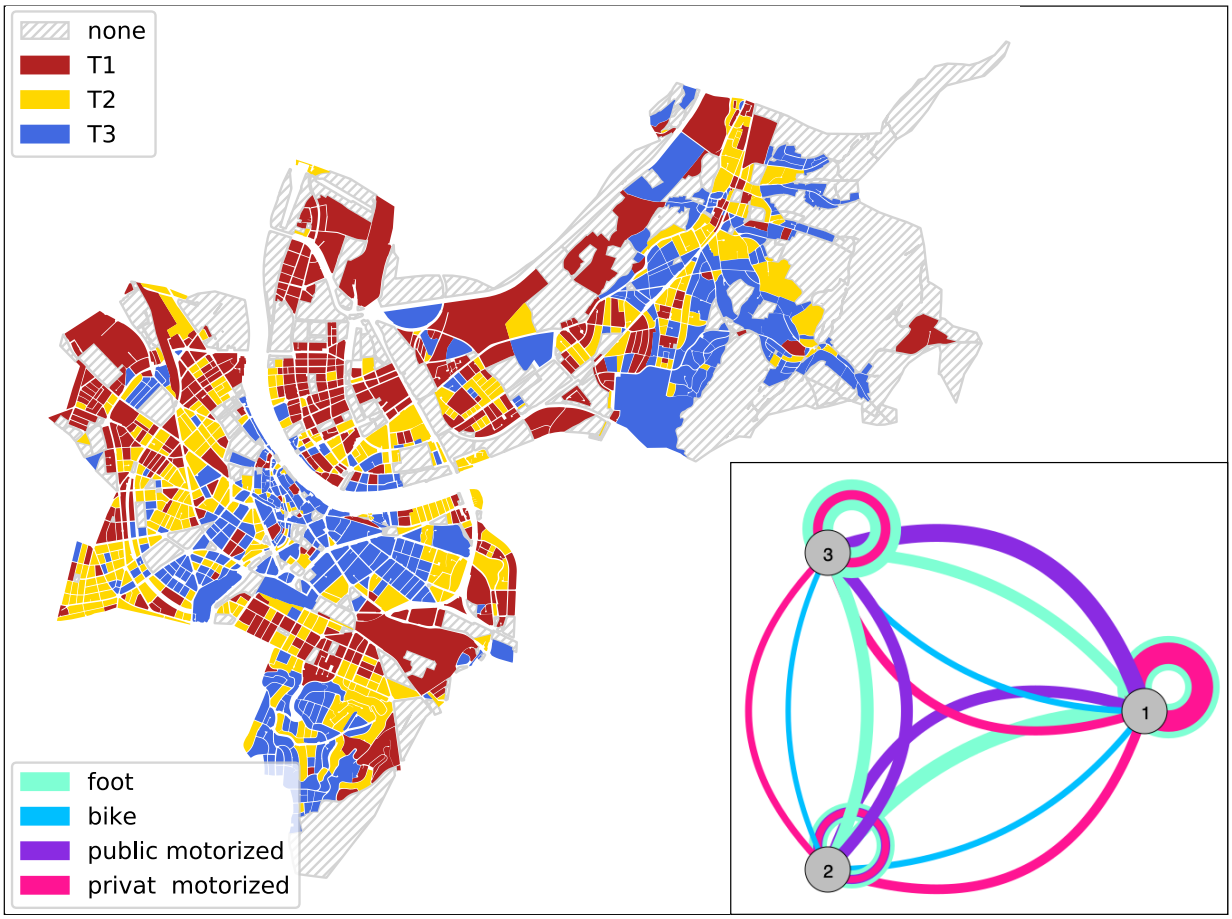


Figure S3: The Canton of Basel City and its delineation with respect to statistical blocks colored according to the partition into tertiles T1, T2, and T3 of increasing living space per person as provided by the canton's office for statistics. Inset: resulting mobility-graph, with nodes representing tertiles and edges representing effective connectedness through mobility by means of various modes of transport (thicker/thinner edges indicating weaker/stronger connectedness), as computed from the traffic-model provided by the traffic department of the Canton of Basel City.

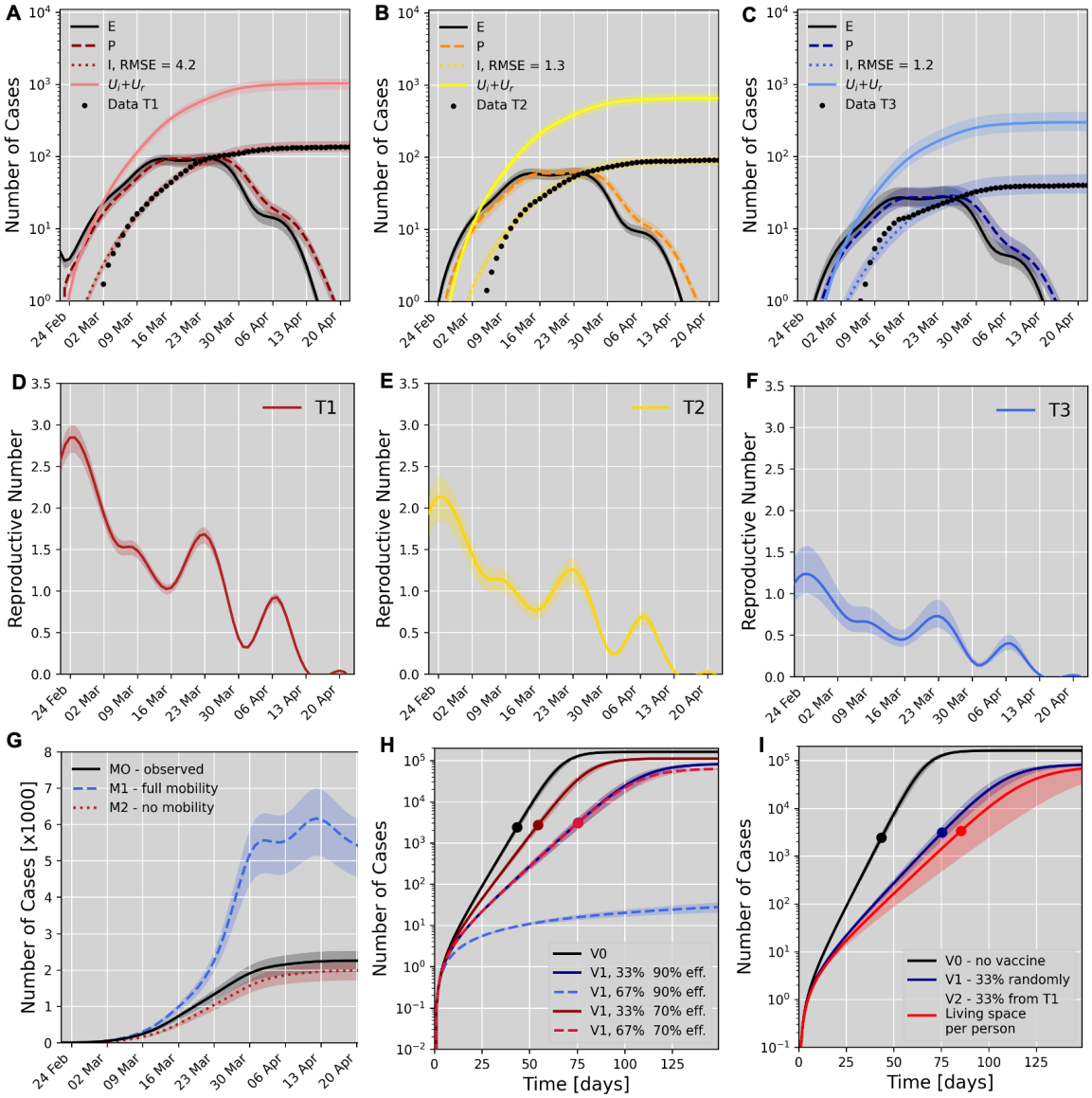


Figure S4: Data fit, reproductive number, and simulation of mobility and vaccination scenarios for a partition according to living space per person. A-C) Model fit to the case number time-series. Data points are shown together with model predictions based on undisturbed data (solid lines), and fifty bootstraps from disturbed data (bands) for the different tertitles T1(A), T2 (B) and T3 (C). D-F) The dynamic variation of the effective reproductive number for each of the tertiles shown in A-C). G) Influence of the mobility pattern on the total number of infected cases (sum of reported and unreported cases) assuming either no change in mobility (M1, 100% mobility), or full shut-down of all inner city mobility (M2, zero mobility). For comparison the observed scenario (M0) is shown. H) Prediction of vaccination effects if a specific percentage of all citizens was randomly selected for vaccination at given efficacy (scenario V1) compared to a simulation in the absence of any vaccination (V0). I) Prediction of vaccination effects if a specific percentage of all citizens was selected for vaccination from tertile T1 according to living space per person (scenario V2) together with the relevant scenarios V0 and V1. Dots indicate the time of reaching 50% ICU occupancy.

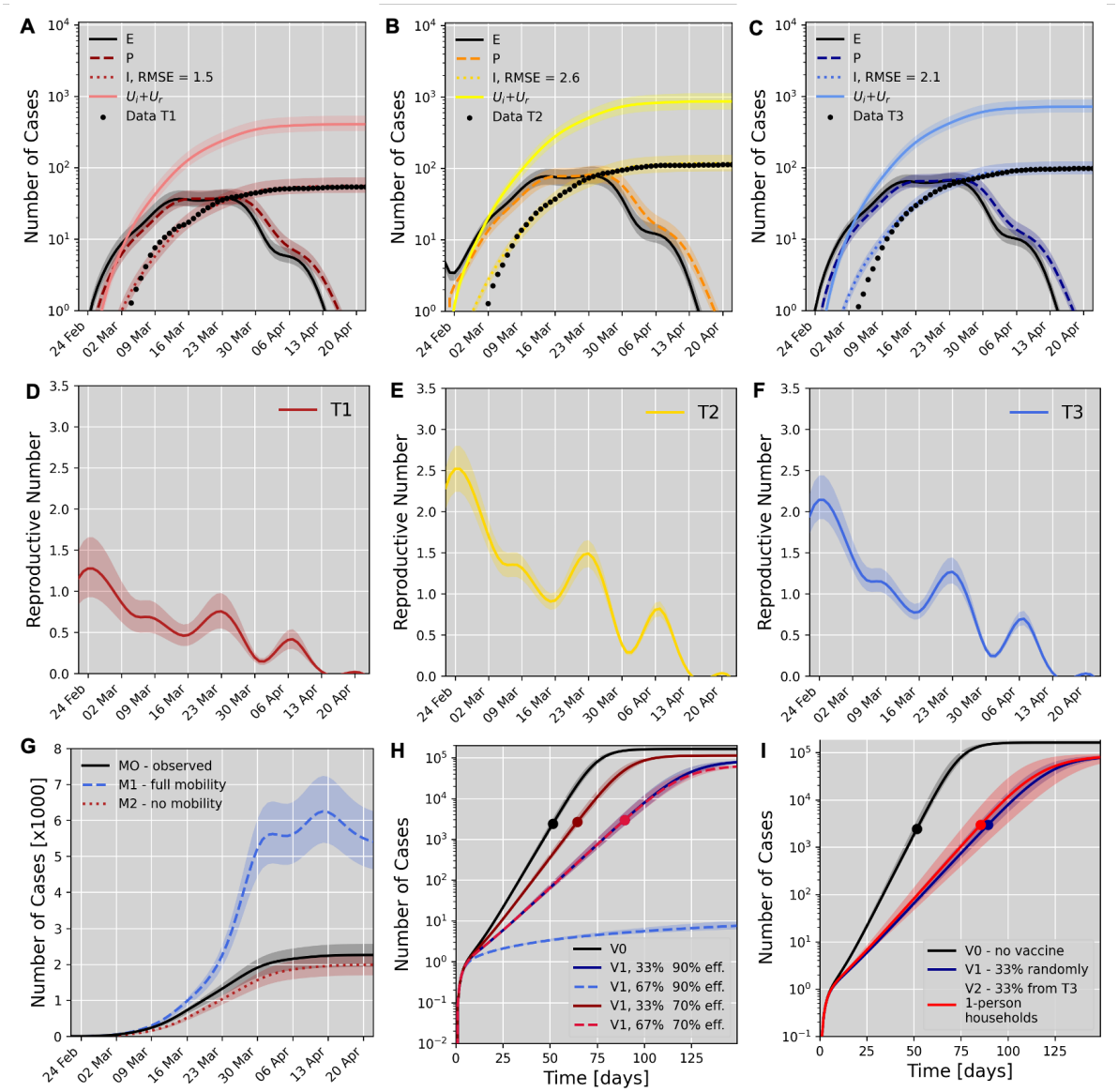


Figure S5: Data fit, reproductive number, and simulation of mobility and vaccination scenarios for a partition according to the share of 1-person households. A-C) Model fit to the case number time-series. Data points are shown together with model predictions based on undisturbed data (solid lines), and fifty bootstraps from disturbed data (bands) for the different tertitles T1(A), T2 (B) and T3 (C). D-F) The dynamic variation of the effective reproductive number for each of the tertiles shown in A-C). G) Influence of the mobility pattern on the total number of infected cases (sum of reported and unreported cases) assuming either no change in mobility (M1, 100% mobility), or full shut-down of all inner city mobility (M2, zero mobility). For comparison the observed scenario (M0) is shown. H) Prediction of vaccination effects if a specific percentage of all citizens was randomly selected for vaccination at given efficacy (scenario V1) compared to a simulation in the absence of any vaccination (V0). I) Prediction of vaccination effects if a specific percentage of all citizens was selected for vaccination from tertile T2 according to the share of 1-person households (scenario V2) together with the relevant scenarios V0 and V1. Dots indicate the time of reaching 50% ICU occupancy.

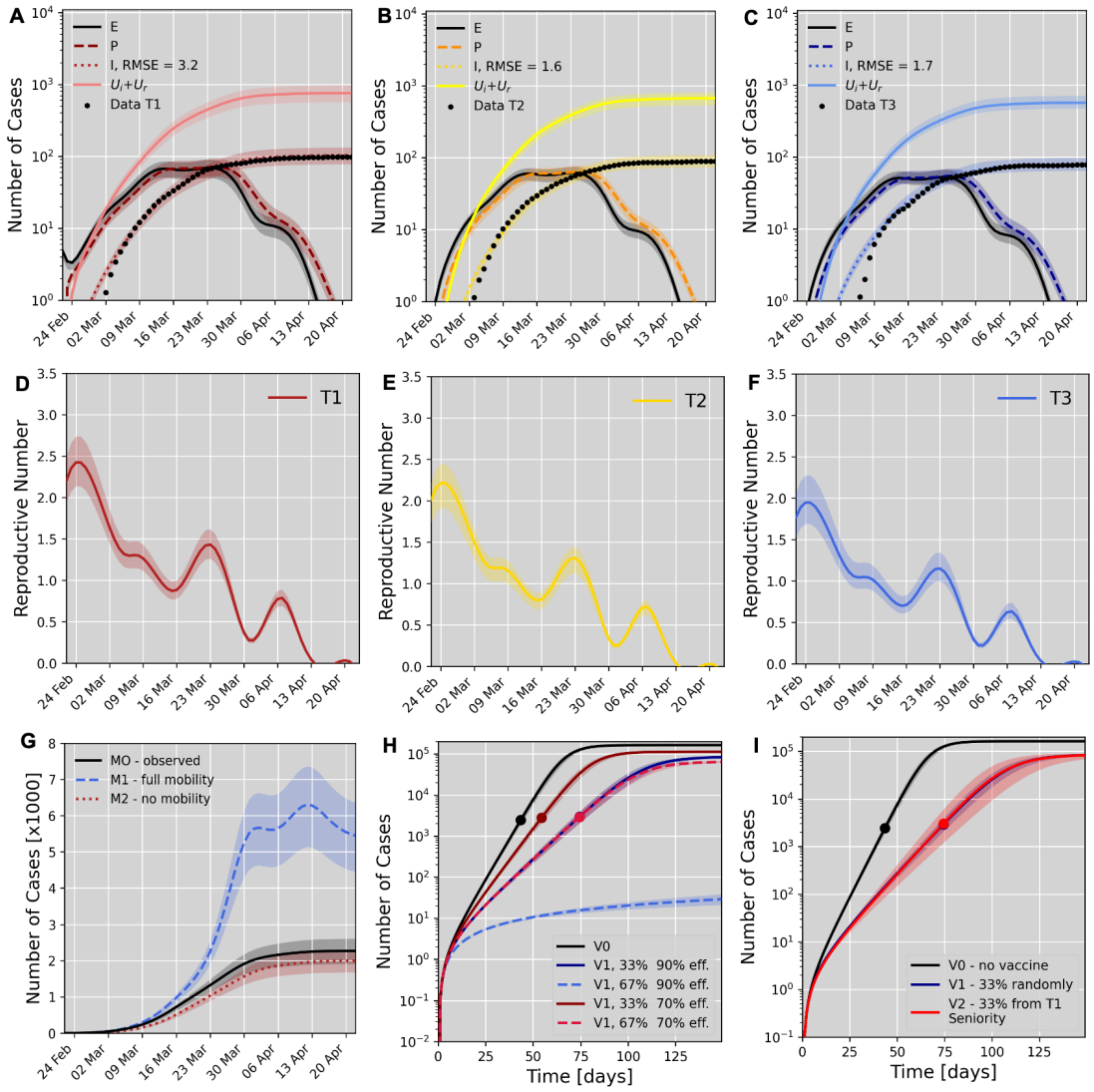


Figure S6: Data fit, reproductive number, and simulation of mobility and vaccination scenarios for a partition according to the share of senior residents. A-C) Model fit to the case number time-series. Data points are shown together with model predictions based on undisturbed data (solid lines), and fifty bootstraps from disturbed data (bands) for the different tertiles T1(A), T2 (B) and T3 (C). D-F) The dynamic variation of the effective reproductive number for each of the tertiles shown in A-C). G) Influence of the mobility pattern on the total number of infected cases (sum of reported and unreported cases) assuming either no change in mobility (M1, 100% mobility), or full shut-down of all inner city mobility (M2, zero mobility). For comparison the observed scenario (M0) is shown. H) Prediction of vaccination effects if a specific percentage of all citizens was randomly selected for vaccination at given efficacy (scenario V1) compared to a simulation in the absence of any vaccination (V0). I) Prediction of vaccination effects if a specific percentage of all citizens was selected for vaccination from tertile T1 according to the share of senior residents (scenario V2) together with the relevant scenarios V0 and V1. Dots indicate the time of reaching 50% ICU occupancy.

Figure S7: Urban quarters in Basel-City.

socioeconomic indicator	achieved significance level
living space per person	1%
median income	2%
fraction of 1-person households	2%
fraction of residents aged above 64	45%

Table S1: Achieved significance level (ALS) of maximum differences in R_{eff} associated with a partition of housing blocks according to various socioeconomic indicators. ALSs have been obtained by comparing these differences in R_{eff} with those obtained from 99 bootstrapping random partitions.

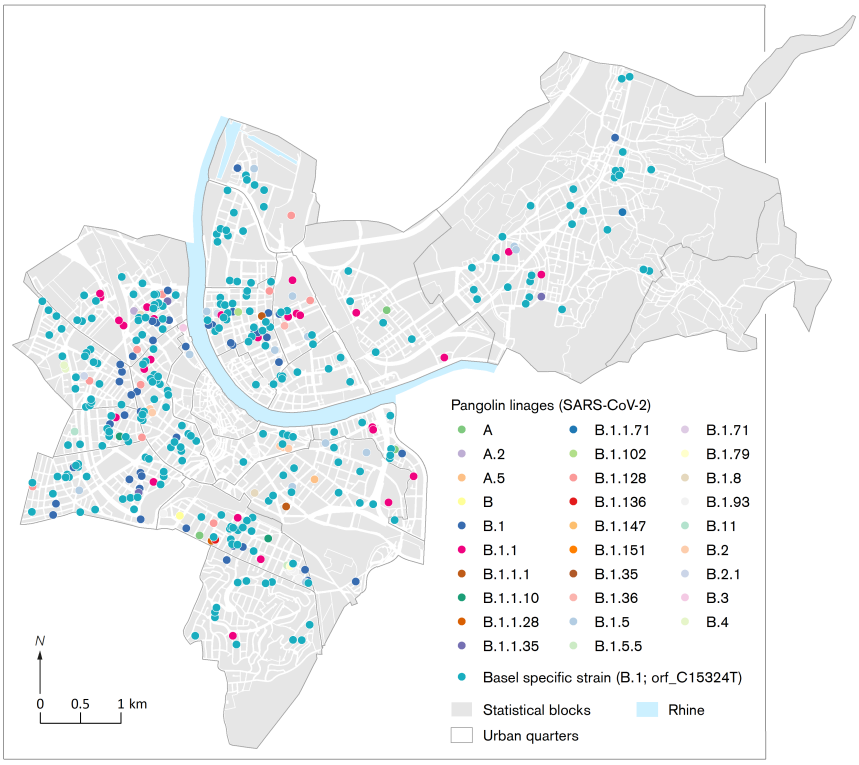


Figure S8: Lineage identity (pangolin) of PCR-confirmed COVID-19 cases from 26th of February until 22nd of April, 2020, in Basel-City with B.1-C15324T as dominant variant highlighted.

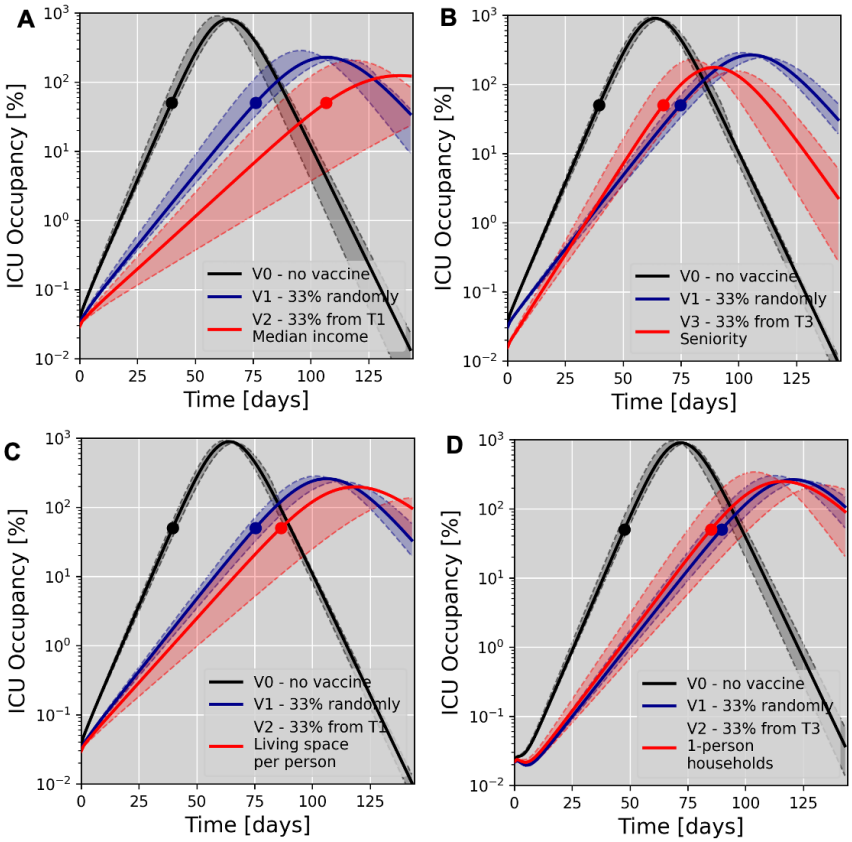


Figure S9: Modelling of ICU capacity for the three vaccination scenarios V0 (no vaccine), V1 (random vaccination of 33% of the population) and V2 or V3 (vaccination of 33% of the population prioritizing the specified tertile). Results are shown for all socioeconomic partitions including median income (A), share of senior residents (B), living space per person (C), and share of 1-person households (D). Dots indicate the time of reaching 50% ICU occupancy.