medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Mobility-driven synthetic contact matrices: a scalable solution for real-time pandemic response modeling

Laura Di Domenico[1], Paolo Bosetti[2], Chiara E. Sabbatini[3], Lulla Opatowski[4,5], Vittoria Colizza[3,6,*]

[1]*Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland*
[2]*Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Université Paris Cité, CNRS UMR 2000, Paris, France*
[3]*Sorbonne Université, INSERM, Pierre Louis Institute of Epidemiology and Public Health, Paris, France*
[4]*Institut Pasteur, Epidemiology and Modelling of Antibiotic Evasion, Université Paris Cité, Paris, France*
[5]*UVSQ, INSERM, CESP, Anti-Infective Evasion and Pharmacoepidemiology Team, Université Paris-Saclay, Montigny-Le-Bretonneux, France*
[6]*Department of Biology, Georgetown University, Washington, District of Columbia, USA*

*corresponding author: vittoria.colizza@inserm.fr

Abstract

Non-pharmaceutical interventions during the COVID-19 pandemic significantly disrupted social mixing patterns, creating a need for updated mathematical models to guide an effective response. Accurately capturing evolving, age-specific social contacts has proven challenging. This study evaluates the effectiveness of mobility-driven synthetic contact matrices against survey-based empirical matrices in capturing the dynamics of COVID-19 observed in France from March 2020 to May 2022. Both matrices showed a gradual increase in average contacts following the first lockdown, with the closest agreement during school closures. However, when schools were open, empirical matrices recorded 3.4 times more contacts for individuals under 19 than synthetic matrices. The model parameterized with mobility-driven contact matrices provided the best fit to hospital admission data and captured hospitalization patterns for adolescents, adults, and seniors, whereas deviations remained for children across both models. Neither matrix allowed models to accurately reproduce serological trends in children in 2021, highlighting the challenges both approaches face in capturing disease-relevant contacts in children. These findings demonstrate the value of synthetic matrices as flexible, cost-effective tools for epidemic modeling, operationally ready in real-time. Routine collection of age-stratified mobility data is essential to improve pandemic response.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

Mathematical models of infectious disease transmission represented a critical tool to guide real-time public health response during the COVID-19 pandemic[1,2]. However, one of the main challenges was accurately integrating changes in human behavior into transmission models[3,4]. The shifts in mobility and contact patterns produced by unprecedented social distancing measures significantly impacted the spread of SARS-CoV-2[5–7]. How to best characterize these shifts and integrate them in real-time modeling remains an open issue.

A key factor in the transmission of respiratory diseases is the pattern of social contacts between age groups[8]. Starting from pioneering work in mid-2000s[9], an increasing number of studies used population-based surveys[10–12] to build static contact matrices describing age-stratified population-level mixing in European countries, as well as to generate synthetic contact matrices in other countries accounting for socio-demographical structures[13–15]. These pre-pandemic contact matrices were essential for early pandemic modeling[16–18]. But they became increasingly inadequate as the pandemic progressed and various interventions were implemented, such as school closures and remote working that affected age groups differently. Models relying on static matrices, or those assuming uniform rescaling of contacts[19–22], struggled to accurately estimate the impact of these interventions across ages, e.g. in terms of hospitalizations and cumulative infections.

Ideally, time-varying contact matrices can be constructed from repeated social contact surveys, but such surveys are resource-intensive and often difficult to implement in real-time. During the COVID-19 pandemic, the CoMix project conducted repeated social contact surveys in representative samples of the populations of over 20 countries in Europe[23]. Yet, only the UK continuously collected data throughout the pandemic through weekly waves[24], and only three countries (the UK, Belgium, the Netherlands) covered also the first pandemic wave, the one reporting the largest shifts in behavior. In France, survey data covering the first wave were available through SocialCov[25], a project collecting contact data from a convenient sample recruited online. While surveys with frequent waves, such as weekly data collection, can provide valuable information on social mixing patterns, they still present considerable challenges for real-time analysis. Processing raw survey data to construct accurate contact matrices in real-time is resource-intensive, requiring efficient data cleaning, aggregation, and interpretation to reflect dynamic behavioral shifts. When survey data were unavailable, transmission models had to rely on alternative proxies to estimate changes in social mixing.

Mobility data[26–29] proved essential[30], providing insights into movement flows and location-specific activity in response to restrictions and recommendations. Early in the pandemic, members of our team developed a novel framework to generate mobility-based synthetic contact matrices for France, capturing shifts in contact patterns driven by the epidemic and governmental measures. Initially introduced to assess the impact of the first lockdown[18], this framework was later expanded to integrate various data sources throughout the pandemic[31]. The synthetic contact matrices were constructed by applying age-specific contact reductions to the pre-pandemic contact matrix[12] based on location and contact type. Google mobility data informed adjustments in workplace contacts, school attendance data shaped changes in the school contact layer, and pandemic survey data on physical contact avoidance reduced skin-to-skin interactions. These data streams enabled to produce weekly, real-time synthetic contact matrices throughout the pandemic[18,31–35].

Although mobility-based synthetic contact matrices offer a promising alternative for real-time modeling, their accuracy relative to empirical contact matrices remains unexplored. This study fills this critical gap by evaluating the effectiveness of these two approaches in modeling the pandemic dynamics in France from March 2020 to May 2022. Focusing on weekly mobility-based synthetic contact matrices that were used for pandemic response[32] and empirical matrices from seven waves of social contact surveys[25], we provide insights into the real-time parameterization of transmission models for future outbreak responses.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

RESULTS

Comparison of contact patterns over time

We derived weekly synthetic contact matrices using pre-pandemic empirical contact data (Fig. 1a) and behavioral data collected during the pandemic. These matrices incorporated age-specific contact reductions across different locations and contact types (Fig. 1b). Workplace contacts were adjusted using Google mobility data[26] related to workplaces (Fig. 1c), while reductions in physical contacts across all settings outside the household were informed by health protective behaviors from the French CoviPrev survey[36] (Fig. 1d). These data sources captured changes in behavior in response to public health measures, spontaneous adaptation to rising case numbers, and seasonal variations. Both indicators showed a general upward trend from the levels measured during the first lockdown, approaching pre-pandemic levels by the end of the pandemic crisis, although these levels were yet not fully reached in May 2022. Contacts in other settings (school, transport, leisure) were reduced in the matrix according to closure schedules of school and non-essential businesses (Tables S4 and S5). Detailed methodology for matrix construction is available in the Methods and Supplementary Information.

We compared the weekly synthetic contact matrices with empirical contact matrices from seven waves of the SocialCov survey[25] conducted during different phases of the pandemic in France (Fig. 2a). SocialCov recruited participants through the governmental app TousAntiCovid. Since the survey sample was not representative of the French population, it was adjusted using sampling with replacement in each wave to reproduce age and gender distribution in France.

In the first lockdown (spring 2020), synthetic matrices estimated 3.4 daily contacts, consistent with survey estimates of 3.6 (Fig. 2b). This corresponded to a 76% reduction in contacts compared to pre-pandemic values. Over time, the number of synthetic contacts exhibited an increasing trend, modulated by intermittent school closures and social distancing measures, reaching 10.4 contacts by May 2022 (age stratification shown in Fig. 2e,f). This number was highly correlated with the Normalcy Index (Pearson's coefficient = 0.86, p-value < 10^{-236}; Fig. 2c, S7), which reflects pandemic-driven behavioral shifts, and the Stringency Index (Fig. S7), quantifying the intensity of social distancing measures (see Methods). The same increasing trend was observed in the empirical SocialCov contacts, but following the first lockdown they were on average generally twice as high as synthetic estimates, with the ratio gradually decreasing from 2.1 in December 2020 to 1.5 in May 2022, except during summer 2021 when it fell to 1.2 (Fig. 2d). These differences were largely driven by discrepancies in children and adolescents. Indeed, comparisons of the matrices by age groups revealed good agreement in contact numbers among adults and seniors, with contact ratios ranging between 0.9 and 1.4 (Fig. 2d, 2e, 3a). However, contacts among individuals under 19yo were 3-4 times higher in survey-based matrices compared to synthetic matrices, except during school closures in April 2020 and August 2021 (Fig. 2d, 2f, 3b).

Normalized contact matrices, which disregard contact intensity, showed strong cosine similarity between survey-based and synthetic matrices, exceeding 97% during school closures and ranging between 84-90% otherwise (Fig. 3c). Age-assortativity analysis revealed that survey-based matrices were less assortative, indicating weaker preferential mixing within the same age groups compared to synthetic matrices (Fig. 3d). Additionally, young individuals under 19yo accounted for approximately 50% of average connectivity in survey-based matrices, compared to 25% in synthetic matrices (Fig. 3d).

Transmission dynamics, model selection and validation

The structural differences in contact matrices, which cannot be fully captured by a single scaling factor, lead to varying estimates of the reproductive number (Fig. S8). To assess the impact of these differences on transmission dynamics, we used an age-stratified compartmental model for COVID-19 (Fig. S1), fitting it to daily hospital admissions using either weekly synthetic contact matrices obtained from mobility data or survey-based matrices (Fig. 4a). For the latter, we extended the seven empirical contact matrices to periods beyond the survey waves based on assumed similarity of mixing conditions, e.g. we used the matrix estimated for August 2021 for all following periods with school holidays, but also in previous periods, i.e. in periods antecedent to data collection (see Methods and Fig. 4b). The model estimated a

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

baseline transmission rate per contact in the pre-lockdown phase and a time-varying multiplicative factor to adjust the transmission rate over time, accounting for shifts in the force of infection not captured by the contact matrices, nor by other relevant aspects explicitly included in the model (e.g., age-dependent susceptibility and severity, transmission advantages of variants, vaccine effectiveness; see Methods). In practice, this correcting factor acts as a global rescaling of the contact matrices. A correcting factor closer to 1 indicates better alignment between the contact matrices and actual disease-relevant contacts. We also considered a model using the static pre-pandemic contact matrix for comparison.

The model parameterized with synthetic contact matrices best fitted hospital admission data in terms of AIC (Table S8), also yielding lower mean absolute error compared to models using survey-based matrices or pre-pandemic matrices (Fig. 4c). Discrepancies were most pronounced during the Alpha and Omicron waves (Fig. 4c, S3). Hospitalization patterns by age, not used to fit the models, were also better captured by the synthetic matrices model for adolescents, adults and seniors (Fig. S5). For children, larger deviations between model predictions and observed hospitalizations were found in all models.

The median correcting factor over time was estimated at 1.16 (IQR 0.87–1.33) for the model using synthetic matrices and 0.79 (IQR 0.64–1.08) for the one survey-based matrices (Fig. 5a), suggesting that small adjustments were needed to accurately reflect shifts in the effective contact behavior relevant to transmission dynamics for both models. The correcting factors varied over time differently between the two models. For the model using synthetic matrices, the correcting factor was close to 1 during the Alpha, Delta, and Omicron BA.1 waves, while for the model using survey-based matrices, it was around 1 in 2020 and during the Omicron BA.2 wave (Fig. S4). The synthetic matrices showed the largest deviations from 1 during summer 2020 (Fig. 5b). In contrast, a model using a static pre-pandemic contact matrix would generally require larger corrections throughout the pandemic period (Fig. 5a).

We then compared model estimates of antibody-positive individuals (Methods) with age-specific serological data. Estimates for adults and seniors were similar across the three models and aligned well with serological data (Fig. 5e, f). However, for children and adolescents, the model informed by empirical contact matrices produced much higher estimates than the model using synthetic matrices, especially after summer 2020. These higher estimates matched the observed serological status of adolescents in June 2021 but were too high in February 2021, where the model using synthetic matrices better captured the serological status (Fig. 5d). For children, neither model successfully reproduced the observed trend in 2021: the model parameterized with empirical matrices overestimated the serological status, while the model with synthetic matrices underestimated it (Fig. 5c).

Sensitivity analyses confirmed overall these findings. Varying the specification of survey-based contact matrices (i.e., incorporating weekend effects or using May 2022 as the pre-pandemic matrix) did not improve the model fit with respect to a model using synthetic contact matrices (Fig. S9). Additionally, assuming lower relative susceptibility for children and adolescents (70% with respect to adults) also for the variants increased the correcting factors across all models (with a 5-6% relative increase in the median correcting factor with respect to the main analysis), but did not allow to fully capture serological data (Fig. S10). The synthetic model still outperformed the survey-based matrices model in terms of AIC and mean absolute error (Table S9, Fig. S10).

DISCUSSION

Effective modeling for real-time outbreak response requires continuously updated contacts that accurately capture shifts in behavior[19,37,38]. Traditional surveys, while strongly valuable, can be difficult and costly to implement. Moreover, delays from data collection to contact matrix construction and integration into models can potentially hamper timely decision-making[39]. In response to these challenges during the COVID-19 pandemic, synthetic contact matrices based on mobility data were used for the first time as an alternative[18]. These matrices offered the ability to update contact patterns in real time, enabling more agile modeling of transmission dynamics[31–35,40,41]. Our study demonstrates that mobility-based synthetic matrices performed well in capturing the dynamic changes in contact behavior across

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

age groups throughout the pandemic, allowing accurate estimates of hospitalizations and infections across most age groups, and thus offer a valuable alternative to a model informed with empirical survey-based matrices.

The transmission model informed by weekly synthetic contact matrices better reproduced the COVID-19 epidemic in France from March 2020 to May 2022, achieving the lowest AIC and mean absolute error compared to models based on empirical matrices or static pre-pandemic matrices. This finding extends previous modeling results limited to the post-first-lockdown phase[31]. The parameterization with synthetic matrices required only a small correction (i.e., a time-varying multiplicative factor close to 1) to reproduce observed hospitalizations, indicating the ability of these matrices to reflect temporal changes in contact behavior. Similar results were obtained with the model parameterizated with survey-based matrices, but adjustments differed over time between the two models. For example, the synthetic matrices struggled to capture contact patterns during summer 2020—the first summer of the COVID crisis—when contacts were highly impacted and not easily reflected by workplace presence or school calendars due to the summer holidays. While survey-based matrices performed slightly better during this period, they were retrospectively informed using data collected in summer 2021, highlighting the challenge of real-time application when survey waves are infrequent; higher survey frequency of surveys in critical periods would likely improve performance.

Both models accurately captured hospitalization and infection rates among adults and seniors. However, the model using empirical contact matrices estimated higher infection rates in children and adolescents compared to the synthetic matrix model. Neither model fully captured the serological status of children in 2021.

This discrepancy between synthetic and survey-based matrices suggests that contact patterns established by young individuals and relevant to transmission are particularly challenging to capture. The synthetic matrices applied uniform reductions in physical contacts across all age groups, using the CoviPrev survey[36], which however was limited to adults only. This may have led to an underestimation of school contacts, as children likely had less opportunity to avoid physical interactions in structured environments like classrooms. Indeed, contact estimates from the two methods reconcile during periods of school closure. Conversely, the survey-based matrices likely overestimated these contacts, leading then to inflated infection estimates for children and adolescents. This may stem from several factors. First, safety measures like mask-wearing in schools and staggered schedules likely mitigated the risk of transmission, meaning the same level of contact did not necessarily result in higher transmission. Second, the data collection method could contribute to the overestimation of contacts in the survey-based matrices for SocialCov. The SocialCov survey used convenience sampling rather than quota sampling, even though data were then adjusted for representativeness. Moreover, the pre-pandemic surveys required detailed listings of individual contacts, while the pandemic-era SocialCov surveys used aggregated contact reports by age group, which may have made it easier for respondents, especially parents reporting for children, to overestimate contacts. Indeed, previous work has shown that aggregated formats tend to report higher contact numbers compared to detailed individual listings[42]. Overall, the smaller number of contacts generated synthetically in the younger age groups better captured the disease dynamics of hospitalizations when integrated into the transmission model.

Contact patterns for adults and seniors, as well as infection rates, were consistent across both synthetic and empirical approaches. Our approach to build synthetic matrices assumed that reductions in workplace attendance, inferred from mobility data, corresponded to reductions in contact rates, using an approximation between a fully density-dependent and a fully frequency-dependent transmission scheme. This assumption aligns with recent findings that workplace mobility reductions significantly impact contact patterns[40]. The strong agreement between model predictions and age-stratified surveillance and serological data for adults and seniors supports the use of mobility data as an effective approach for modeling workplace-related contacts during a pandemic. Previous studies observed a strong correlation between mobility data and COVID-19 spread in the early stages of the outbreak[43–45], which weakened over time as behavior changes and preventive measures (e.g., masking) became widespread[46–49]. This diminishing correlation exposed the limitations of a simplistic use of mobility data in

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

predicting pandemic trends[47]. However, these limitations critically depend on the mobility metrics chosen, their integration into transmission models, and the specific epidemiological objectives. Although linear relationships with transmission rates did not persist over time, our findings demonstrate that workplace attendance data from mobility sources can be informative when effectively integrated into a synthetic modeling framework. This approach provides a valuable alternative to empirical contact data, contingent on selecting appropriate mobility sources and methods to generate synthetic contacts from these sources. In our model, we specifically used workplace-related mobility to non-linearly adjust the work layer of the contact matrix, allowing it to reflect real-time mobility restrictions.

A key advantage of synthetic contact matrices is their ability to be updated weekly, a flexibility generally lacking in survey-based matrices[24]. For survey-based matrices without continuous follow-up, assumptions are required to extend contact patterns between survey waves. We tested various matrix specifications, and the model parameterized with mobility-based synthetic matrices consistently outperformed those with survey-based matrices. However, we did not explore optimal methods for extending empirical contact matrices, as these methods may be context-specific and and warrant further research. Additionally, infrequent survey waves may require the use of data from different periods (e.g., using data from summer 2021 to inform summer 2020), limiting the applicability of these surveys in real-time and introducing further assumptions for temporal extensions. Even with frequent survey waves, such as weekly, the process of generating contact matrices from raw survey data remains resource-intensive, posing a challenge for real-time modeling. Automated tools and streamlined methods are essential to transform survey data into actionable insights efficiently[23,50].

On the other hand, empirical matrices offer higher age resolution, which is theoretically achievable with synthetic matrices but critically depends on the availability of age-stratified mobility data. In some cases, cellphone mobility data can provide such age-specific stratification[51]. Google mobility data, though not age-stratified, effectively reflects adult behavior through workplace attendance trends. For school-related contacts, we used data from the Ministry of Education during the reopening period after the first lockdown, when school attendance was voluntary. Afterward, we relied on the school calendar, including reactive closures due to non-pharmaceutical interventions. The availability of real-time or near-real-time data remains a challenge for generating synthetic matrices, especially in terms of age-specific detail, and particularly in younger age groups.

Our study has a set of limitations. First, our model assumed distinct daily contacts, ignoring the repetition of contacts, which could underestimate transmission risks[52]. However, this assumption was intrinsic to the modeling framework chosen and could be improved only moving to an agent-based framework. Second, we considered age-specific susceptibility for the original strain[53,54], and assumed homogeneous susceptibility across age groups for the variants due to limited evidence and following other works[22,55]. However, sensitivity analyses considering age-specific susceptibility also for variants showed that our best model, informed by synthetic contact matrices, continued to outperform other models in terms of AIC and error. Third, due to a lack of age-specific estimates, we assumed that the average time to seroreversion for young individuals was the same as that for adults. This assumption may limit the accuracy of comparisons between model predictions and serological data for children and adolescents. Finally, the mobility and behavioral data used to construct synthetic matrices were not age-stratified, although the proxies employed (e.g., workplace attendance, school calendar) were indirectly age-specific. Despite these limitations, our results demonstrate that mobility-informed synthetic contact matrices provide a robust, adaptable approach for modeling transmission dynamics during pandemics, and they offer a real-time alternative to static or infrequently updated empirical contact data, with strong implications for pandemic preparedness and response. We did not include here French social contact data from the COMIX survey[23]. These survey data for France refer to a rather restricted period (December 2020 to April 2021) that does not cover the first wave, and data for minors were collected only for 2 waves out of 7 adult waves. Future work could extend this analysis to other countries.

This study highlights the potential of mobility-based synthetic contact matrices to accurately model changes in contact behavior and epidemic dynamics during a pandemic. Compared to infrequent survey-based matrices often unavailable in near real time, synthetic matrices captured well the time-varying

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

nature of contacts, leading to good predictions of hospitalizations and infection rates, particularly when real-time data were critical. The findings advocate for greater integration of non-traditional digital data sources, such as mobility, into epidemiological modeling frameworks. These data streams are more flexible, scalable, and cost-effective than empirical surveys, making them valuable tools for real-time outbreak monitoring and response. As the world prepares for future pandemics, our study underscores the importance of leveraging real-time data to inform public health interventions and improve crisis management.

METHODS

Pre-pandemic baseline contact matrix. We considered four age groups: children [0-10], adolescents [11-18], adults [19-64] and seniors with 65+ years old. We used pre-pandemic contact data collected from a large-scale survey in France in 2012[12], distinguishing between contacts engaged during regular weekdays, weekends or school holidays. We derived a social contact matrix corrected by reciprocity, and broken down by location (home, school, work, transport, leisure, other) and type of contact (skin-to-skin or non-physical contact at short distance), using the Social Contact Rates (SOCRATES) Data Tool[50]. The original survey collected Supplementary Professional Contacts (SPC), i.e. participants with more than 20 daily professional contacts were asked not to report them but rather to provide their total number and age distribution. We included SPC only in the elements of the work matrix involving adults and seniors. This baseline matrix was then adapted to the French population in 2020 using demography data, applying an appropriate density correction (following Ref.[15]). Through the density correction, the original matrix M_{ij} (whose elements represent the average number of contacts an individual in age group i establishes with individuals in age group j) is projected to the demographic structure of 2020 by defining a new matrix $M'_{ij} = M_{ij} * (N/N_j) * (N'_j/N')$, where N, N_j, refer to the total population and the population in age group j, respectively, in the year of the survey, while N', N'_j refer to the population in 2020. This density correction preserves reciprocity, so that the projected matrix fulfills the condition $M'_{ij}N'_i = M'_{ji}N'_j$. We considered the contact matrix estimated for a regular weekday. We did not model explicitly the weekday/weekend effect, which is absorbed in the correcting factor estimated when fitting the transmission model. However, we accounted for the impact of school holidays. We used the pre-pandemic contact data collected during spring school holidays to model the synthetic pandemic contact matrix during spring, winter and Christmas holidays. We made some assumptions when modeling summer holidays in the lack of available pre-pandemic contact data. See the section below and the Supplementary Information for further details.

Construction of synthetic contact matrices. We built time-varying synthetic contact matrices on a weekly basis (except for lockdown periods) from March 2020 to May 2022. Matrices were obtained applying reductions to the layers of the pre-pandemic contact matrix (location and type of contact) to simulate the social mixing conditions experienced during the pandemic. In particular, we parametrized the matrices to account for (i) impact of adoption of telework in reducing contacts at work and on transports, (ii) impact of full or partial school closure (or remote learning) and impact of school holidays on school-related contacts, (iii) impact of social-distancing measures on contacts associated to non-essential activities, (iv) impact of adoption of health preventive behaviors such as avoiding physical contacts. We used proxy contact data collected during the pandemic to inform the parameterization of the synthetic contact matrices. In particular, we used Google mobility data[26] related to workplaces to adjust the number of contacts in the work matrix over time. Google measures the change in the number of visitors to a specific location with respect to a pre-pandemic baseline; the mobility change related to workplaces can thus be interpreted as an effective reduction in attendance at work. We assumed that such reduction in attendance produces a reduction of contact rates that is in between the frequency-dependent and the density-dependent assumption[56]. We also used data on voluntary school attendance in the exit phase of the first lockdown, and the calendar of school holidays to adjust school-related contacts. Finally, we used data from the CoviPrev survey[36] on declared avoidance of physical contacts during the pandemic to reduce the proportion of skin-to-skin contacts. The framework is fully detailed in Section 4.2 of the

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Information. The resulting average number of contacts over time was tested for correlation with the Normalcy Index and the Stringency Index. The Normalcy Index[57] is a measure of the impact of the pandemic on human behavior, integrating multiple daily indicators of human activities in a score from 0 to 100, with 100 representing the pre-pandemic level. The Stringency Index[58] is a composite measure of nine response metrics (e.g. as school closure, restrictions on public gatherings) to quantify the strictness of government policies for epidemic control, in a scale from 0 to 100, with 0 indicating absence of measures.

SocialCov contact surveys. Seven surveys were conducted in France to collect data on contact behavior. Survey participants were recruited online and contact matrices were adjusted to the French population for representativeness. SocialCov recruited participants through convenience sampling through the governmental app TousAntiCovid. In particular, the survey was promoted via the news channel of the app which invited individuals aged 18 and above to complete the questionnaire. Since the survey sample was not representative of the French population, synthetic populations were generated for each campaign to better reflect the age and gender distribution in France, using sampling with replacement from the SocialCov participant pool. The first survey was conducted during the first lockdown (matrix LD)[25]. The survey was then implemented in 6 additional waves, in December 9-22, 2020 (M1), January 10-21, 2021 (M2), March 2-10, 2021 (M3), August 12-24, 2021 (M4), December 6-17, 2021 (M5), and May 20-29, 2022(M6). Contacts were defined as either a physical contact (such as a kiss or a handshake) or a close contact (such as face-to-face conversation at less than 1 m distance). Contacts for each individual were truncated at 50 to reduce the impact of outliers. We used contacts reported on a weekday to allow comparison with the synthetic contact matrices that were built based on a regular weekday pre-pandemic matrix. In a sensitivity analysis, we used matrices weighted by weekday and weekend. Participants in the survey reported contacts aggregated by age group, and were used to produce age-stratified 10x10 contact matrices. We aligned the age groups of the contact survey with the four age groups used for the synthetic matrices to allow comparison (see Section 4.3 of the Supplementary Information).

Comparison of contact patterns. From the set of weekly synthetic contact matrices, we extracted the ones with the closest matching period to the survey waves and made direct comparison between the two sets of matrices (Table S6). We summarized the information contained in each contact matrix through different metrics in order to make the comparisons. We computed (i) the average number of contacts, overall and by age class, (ii) a measure of matrix correlation based on the cosine similarity, (iii) the proportion of average connectivity due to young individuals (<19 y.o., i.e. including children and adolescents) and (iv) an index for age-assortativity. These quantities are mathematically defined in the Supplementary Information (Section 5). The degree of assortativity measures the extent to which contacts occur between individuals who share their characteristics (in our case, the relevant characteristic is age). Contact matrices estimated from empirical data usually show some assortativity with age, i.e. they have a strong diagonal component, as individual tend to mix with individuals with similar age. By quantifying the degree of assortativity though an index, we compared the mobility-based synthetic matrices and the survey-based empirical matrices to understand whether contact patterns are more or less assortative. We also compared the two sets of matrices by computing the ratio of their largest eigenvalues and the ratio of basic reproductive number, using the next-generation matrix approach[59,60] .

Transmission model. We integrated the two sets of contact matrices into an age-stratified transmission model to simulate the unfolding of the COVID-19 pandemic in France from early 2020 to May 2022. The model has been presented in depth in a previous work[32]. We used a stochastic age-structured two-strain transmission model with vaccination, parameterized using French data on demography[61], age profile[61], and vaccine uptake[62]. Transmission dynamics follows a compartmental scheme illustrated in Fig. S1, which accounts for latency period, pre-symptomatic transmission, asymptomatic and symptomatic infections with different degrees of severity, and individuals affected by severe symptoms requiring hospitalization. Contact rates for each disease stage are adjusted to model the effect of spontaneous change of behavior due to severe illness and the impact of testing and self-isolation. The model

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

reproduces the co-circulation of two strains, and was applied to describe the Wuhan-Alpha period (February 2020 – May 2021), the Alpha-Delta period (June 2021 – August 2021), and the Delta-Omicron period (September 2021 – May 2022). Epidemiological parameter values and sources for the Wuhan strain are reported in Table S1. The model is parametrized with age-dependent susceptibility and disease severity. For the Wuhan strain, children and adolescents have a relative susceptibility of 70% with respect to adults. Variant-dependent parameters include the generation time, the transmission advantage, the infection-hospitalization ratio (Section 2 of the Supplementary Information). The model is further stratified by (i) vaccine dose, to build vaccine coverage in the population over time according to data on vaccine doses administered in France[62], and (ii) time since vaccination, to model steps of waning in vaccine effectiveness. The model accounts for possible re-infection with Omicron after a prior infection, with waning in protection against re-infection. We distinguished between different levels of protection conferred by vaccine-only, natural (infection-only) and hybrid immunity (Section 3 of the Supplementary Information). In parallel to disease stage progression, we also modeled seropositivity to compare model results with seroprevalence data from Sante publique France measuring the presence of IgG-type antibodies. Following the modeling approach adopted in Ref.[21], we assumed that upon becoming infectious (i.e. while exiting the E compartment in Fig. S1), infected individuals also enter in a compartment AB_{pre} (in parallel to I_p) which represents the pre-seropositivity compartment. Then, individuals move to the seroposivity compartment AB^+ after seroconversion, and finally they move to the seronegativity compartment AB^- after seroreversion. We informed the average time spent in AB_{pre} and AB^+ based on estimates from the literature for IgG-type antibodies. We used 12 days for seroconversion[63] for all age groups, and 200 days for seroreversion for adults[64–67]. Given the evidence of slower seroreversion for more severe infections, we used 400 days for seniors[68]. In the lack of estimates specific for children and adolescents, we assumed the same value as adults. We then compared the proportion of AB^+ over time predicted by the model with French national seroprevalence estimates[69] collected by Sante publique France, available by age group, and corrected by test sensitivity.

Inference framework. The model was fitted to daily hospital admission data since the start of the pandemic (February 2020) up to May 22, 2022. We used a maximum likelihood approach to fit a step-wise transmission rate per contact (see Section 6 of the Supplementary Material for full details). More specifically, in the pre-lockdown phase (February – March 2020), we fitted the starting date of the epidemic and the baseline transmission per contact β_{pre-LD}. Then, we fitted a correcting factor α_{phase} of the transmission rate in subsequent time-windows, each one representing a different pandemic phase, based on epidemic activity, behavior and interventions implemented (e.g. pre-lockdown, lockdown, exit phase, summer, curfew). As changes in the variant's transmissibility are explicitly modeled through the transmission advantage, the parameter α_{phase} is meant to absorb other factors potentially affecting the transmission, that are not captured by the time-varying contact matrices, e.g. mask usage or outdoor/indoor activity, or misspecification of contacts. This correcting factor can be thus interpreted as a correction of the contact matrices used in the model. The product of the fitted α_{phase} at time t with the corresponding contact matrix represents the effective contact rate needed to reproduce the observed epidemic dynamic. The closer the correcting factor to 1, the larger the ability of the model to incorporate the relevant changes in the effective contact rates through the time-varying contact matrices, and to reproduce the epidemic dynamic.

Model comparison. We compared the outcomes of three models. The first model is informed with weekly mobility-based synthetic contact matrices, as done in our previous works during the pandemic[18,31–35]. The second model is informed with empirical contact matrices estimated from survey data (SocialCov[25]) collected during the pandemic. The contact matrices available from 7 distinct surveys were extended beyond the survey period in order to cover the time span of the model, as pictured in Fig. 4b. We used the matrix estimated for summer 2021 as a proxy for school holidays, as this was the only survey wave fully occurring in a period with schools closed (excluding the lockdown). Both models used the empirical pre-pandemic contact matrix in the period prior to the implementation of the lockdown. The two models with time-varying contact matrices (either mobility-based or survey-based) were compared to a third reference model integrating the same static pre-pandemic contact matrix

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

throughout the whole pandemic period. To identify the model which better reproduced the change of behavior and the resulting epidemic dynamic, we compared the model outcomes in terms of (i) Akaike Information Criterion (AIC) as a measure of goodness of fit, (ii) mean absolute error of the model trajectory of daily hospital admissions with respect to the data, (iii) distribution of the correcting factor, (iv) age-specific model estimates of the proportion of antibody-positive individuals over time (accounting for seroconversion and seroreversion) in comparison with seroprevalence data, and (v) age-specific model estimates of hospital admissions compared to age-stratified data. We carried out sensitivity analyses on the susceptibility of young individuals with SARS-CoV-2 variants, and on the integration of the survey-based contact matrices into the transmission model, including (i) the effect of weekend and (ii) the empirical matrix of May 2022 as a pre-pandemic matrix. Further details are provided in Section 9 of the Supplementary Information.

ACKNOWLEDGMENTS

This study was partially funded by: ANR grant DATAREDUX (ANR-19-CE46-0008-03); EU Horizon 2020 grant MOOD (H2020-874850); Horizon Europe grants VERDI (101045989) and ESCAPE (101095619).

AUTHOR CONTRIBUTIONS

V.C. and L.D.D. conceived and designed the study. V.C, L.D.D. and C.E.S. developed the framework for the synthetic contact matrices. L.D.D. and C.E.S analysed the data to build the synthetic contact matrices. P.B. and L.O. collected and analysed the data from the SocialCov survey. L.D.D. developed the code for the comparison. L.D.D. and C.E.S. performed the numerical simulations. L.D.D. analysed the results. L.D.D., P.B., L.O. and V.C interpreted the results. L.D.D. drafted the article. All authors contributed to and approved the final version of the Article.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Code of the transmission model is publicly available at https://github.com/EPIcx-lab/COVID-19/tree/master/mobility_driven_synthetic_contact_matrices, along with the mobility-driven synthetic contact matrices. SocialCov survey data are available from the co-authors P.B. and L.O. upon request. All other data used in the analyses are available online at the cited references.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Becker, A. D. *et al.* Development and dissemination of infectious disease dynamic transmission models during the COVID-19 pandemic: what can we learn from other pathogens and how can we move forward? *Lancet Digit. Health* **3**, e41–e50 (2021).
2. Jit, M. *et al.* Reflections On Epidemiological Modeling To Inform Policy During The COVID-19 Pandemic In Western Europe, 2020–23. *Health Aff. (Millwood)* **42**, 1630–1636 (2023).
3. Funk, S. *et al.* Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics* **10**, 21–25 (2015).
4. Bedson, J. *et al.* A review and agenda for integrated disease models including social and behavioural factors. *Nat. Hum. Behav.* **5**, 834–846 (2021).
5. Zhang, J. *et al.* Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
6. Jarvis, C. I. *et al.* Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med.* **18**, 124 (2020).
7. Lucchini, L. *et al.* Living in a pandemic: changes in mobility routines, social activity and adherence to COVID-19 protective measures. *Sci. Rep.* **11**, 24452 (2021).
8. Wallinga, J., Teunis, P. & Kretzschmar, M. Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. *Am. J. Epidemiol.* **164**, 936–944 (2006).
9. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
10. Hoang, T. *et al.* A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections. *Epidemiol. Camb. Mass* **30**, 723–736 (2019).
11. Mousa, A. *et al.* Social contact patterns and implications for infectious disease transmission – a systematic review and meta-analysis of contact surveys. *eLife* **10**, e70294 (2021).
12. Béraud, G. *et al.* The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases. *PLOS ONE* **10**, e0133203 (2015).
13. Prem, K., Cook, A. R. & Jit, M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLOS Comput. Biol.* **13**, e1005697 (2017).
14. Prem, K. *et al.* Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era. *PLOS Comput. Biol.* **17**, e1009098 (2021).
15. Arregui, S., Aleta, A., Sanz, J. & Moreno, Y. Projecting social contact matrices to different demographic structures. *PLOS Comput. Biol.* **14**, e1006638 (2018).
16. Prem, K. *et al.* The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, e261–e270 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

17. Davies, N. G. *et al.* Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health* **5**, e375–e385 (2020).
18. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Med.* **18**, 240 (2020).
19. Munday, J. D., Abbott, S., Meakin, S. & Funk, S. Evaluating the use of social contact data to produce age-specific short-term forecasts of SARS-CoV-2 incidence in England. *PLOS Comput. Biol.* **19**, e1011453 (2023).
20. Hauser, A. *et al.* Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe. *PLOS Med.* **17**, e1003189 (2020).
21. Knock, E. S. *et al.* Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Sci. Transl. Med.* **13**, eabg4262 (2021).
22. Marziano, V. *et al.* Estimating SARS-CoV-2 infections and associated changes in COVID-19 severity and fatality. *Influenza Other Respir. Viruses* **17**, e13181 (2023).
23. Verelst, F. *et al.* SOCRATES-CoMix: a platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries. *BMC Med.* **19**, 254 (2021).
24. Munday, J. D. *et al.* Estimating the impact of reopening schools on the reproduction number of SARS-CoV-2 in England, using weekly contact survey data. *BMC Med.* **19**, 233 (2021).
25. Bosetti, P. *et al.* Lockdown impact on age-specific contact patterns and behaviours, France, April 2020. *Eurosurveillance* **26**, (2021).
26. Google. COVID-19 Community Mobility Report. *COVID-19 Community Mobility Report* https://www.google.com/covid19/mobility?hl=en.
27. Cot, C., Cacciapaglia, G. & Sannino, F. Mining Google and Apple mobility data: temporal anatomy for COVID-19 social distancing. *Sci. Rep.* **11**, 4150 (2021).
28. Pepe, E. *et al.* COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Sci. Data* **7**, 230 (2020).
29. Besley, T. & Dray, S. Pandemic responsiveness: Evidence from social distancing and lockdown policy during COVID-19. *PLOS ONE* **17**, e0267611 (2022).
30. Oliver, N. *et al.* Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci. Adv.* **6**, eabc0764 (2020).
31. Pullano, G. *et al.* Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature* **590**, 134–139 (2021).
32. Di Domenico, L., Goldberg, Y. & Colizza, V. Planning and adjusting the COVID-19 booster vaccination campaign to reduce disease burden. *Infect. Dis. Model.* **10**, 150–162 (2025).
33. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Modelling safe protocols for reopening schools during the COVID-19 pandemic in France. *Nat. Commun.* **12**, 1073 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

34. Di Domenico, L., Sabbatini, C. E., Pullano, G., Lévy-Bruhl, D. & Colizza, V. Impact of January 2021 curfew measures on SARS-CoV-2 B.1.1.7 circulation in France. *Eurosurveillance* **26**, 2100272 (2021).
35. Di Domenico, L. *et al.* Adherence and sustainability of interventions informing optimal control against the COVID-19 pandemic. *Commun. Med.* **1**, 1–13 (2021).
36. Santé publique France. CoviPrev : une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie de COVID-19. https://www.santepubliquefrance.fr/etudes-et-enquetes/coviprev-une-enquete-pour-suivre-l-evolution-des-comportements-et-de-la-sante-mentale-pendant-l-epidemie-de-covid-19 (2020).
37. Tran Kiem, C. *et al.* SARS-CoV-2 transmission across age groups in France and implications for control. *Nat. Commun.* **12**, 6895 (2021).
38. Taube, J. C., Susswein, Z., Colizza, V. & Bansal, S. Respiratory disease contact patterns in the US are stable but heterogeneous. 2024.04.26.24306450 Preprint at https://doi.org/10.1101/2024.04.26.24306450 (2024).
39. Shadbolt, N. *et al.* The challenges of data in future pandemics. *Epidemics* **40**, 100612 (2022).
40. Davies, N. G. *et al.* Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: a modelling study. *Lancet Infect. Dis.* **0**, (2020).
41. Barnard, R. C., Davies, N. G., Jit, M. & Edmunds, W. J. Modelling the medium-term dynamics of SARS-CoV-2 transmission in England in the Omicron era. *Nat. Commun.* **13**, 4879 (2022).
42. Backer, J. A. *et al.* Dynamics of non-household contacts during the COVID-19 pandemic in 2020 and 2021 in the Netherlands | Scientific Reports. *Sci. Rep.* **13**, 5166 (2023).
43. 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).
44. Kraemer, M. U. G. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
45. Jia, J. S. *et al.* Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394 (2020).
46. Kishore, N. *et al.* Evaluating the reliability of mobility metrics from aggregated mobile phone data as proxies for SARS-CoV-2 transmission in the USA: a population-based study. *Lancet Digit. Health* **4**, e27–e36 (2022).
47. Delussu, F., Tizzoni, M. & Gauvin, L. The limits of human mobility traces to predict the spread of COVID-19: A transfer entropy approach. *PNAS Nexus* **2**, pgad302 (2023).
48. Jewell, S. *et al.* It's complicated: characterizing the time-varying relationship between cell phone mobility and COVID-19 spread in the US. *Npj Digit. Med.* **4**, 1–11 (2021).
49. Badr, H. S. & Gardner, L. M. Limitations of using mobile phone data to model COVID-19 transmission in the USA. *Lancet Infect. Dis.* **21**, e113 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

50. Willem, L. *et al.* SOCRATES: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for COVID-19. *BMC Res. Notes* **13**, 293 (2020).
51. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *Lancet Digit. Health* **2**, e638–e649 (2020).
52. Loedy, N., Wallinga, J., Hens, N. & Torneri, A. Repetition in social contacts: implications in modelling the transmission of respiratory infectious diseases in pre-pandemic and pandemic settings. *Proc. Biol. Sci.* **291**, 20241296 (2024).
53. Davies, N. G. *et al.* Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211 (2020).
54. Viner, R. M. *et al.* Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. *JAMA Pediatr.* **175**, 143–156 (2021).
55. Hogan, A. B. *et al.* Long-term vaccination strategies to mitigate the impact of SARS-CoV-2 transmission: A modelling study. *PLOS Med.* **20**, e1004195 (2023).
56. Hu, H., Nigmatulina, K. & Eckhoff, P. The scaling of contact rates with population density for the infectious disease models. *Math. Biosci.* **244**, 125–134 (2013).
57. The global normalcy index. *The Economist* https://www.economist.com/graphic-detail/tracking-the-return-to-normalcy-after-covid-19 (2021).
58. Hale, T. *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **5**, 529–538 (2021).
59. Diekmann, O., Heesterbeek, J. A. P. & Roberts, M. G. The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface* **7**, 873–885 (2010).
60. Angeli, L. *et al.* Who acquires infection from whom? A sensitivity analysis of transmission dynamics during the early phase of the COVID-19 pandemic in Belgium. *J. Theor. Biol.* **581**, 111721 (2024).
61. INSEE. Estimations de population - Pyramide des âges régionales et départementales. https://www.insee.fr/fr/statistiques/3696315 (2020).
62. Santé publique France. Données relatives aux personnes vaccinées contre la Covid-19 (VAC-SI). https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-personnes-vaccinees-contre-la-covid-19-1/ (2020).
63. Borremans, B. *et al.* Quantifying antibody kinetics and RNA detection during early-phase SARS-CoV-2 infection by time since symptom onset. *eLife* **9**, e60122 (2020).
64. Gallian, P. *et al.* SARS-CoV-2 IgG seroprevalence surveys in blood donors before the vaccination campaign, France 2020-2021. *iScience* **26**, 106222 (2023).
65. Brazeau, N. F. *et al.* Estimating the COVID-19 infection fatality ratio accounting for seroreversion using statistical modelling. *Commun. Med.* **2**, 1–13 (2022).
66. Chen, S., Flegg, J. A., White, L. J. & Aguas, R. Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys. *PLOS Comput. Biol.* **17**, e1009436 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

67. Menges, D. *et al.* Heterogenous humoral and cellular immune responses with distinct trajectories post-SARS-CoV-2 infection in a population-based cohort. *Nat. Commun.* **13**, 4855 (2022).
68. Peghin, M. *et al.* The Fall in Antibody Response to SARS-CoV-2: a Longitudinal Study of Asymptomatic to Critically Ill Patients Up to 10 Months after Recovery. *J. Clin. Microbiol.* **59**, e01138-21 (2021).
69. Le Vu, S. *et al.* Prevalence of SARS-CoV-2 antibodies in France: results from nationwide serological surveillance. *Nat. Commun.* **12**, 3025 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

FIGURES

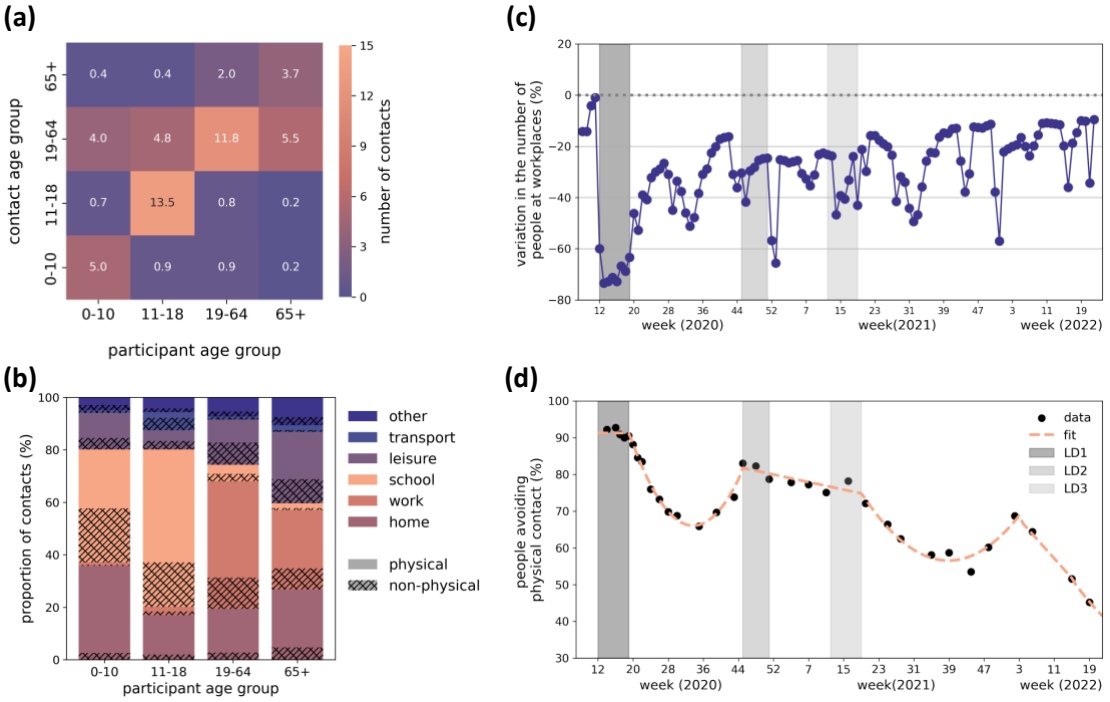

Figure 1. Input data for synthetic contact matrices. (a) Pre-pandemic empirical contact matrix M estimated for France, for a regular weekday from Ref.[12]. The element M_{ij} represents the average number of contacts that one individual in participant age group i (columns) engages with individuals in the contact age group j (rows). **(b)** Breakdown of the total number of contacts for each age group by location (color) and type (pattern), where physical means skin-to-skin contact, non-physical otherwise. **(c)** Google mobility data related to workplaces[26]. The plot shows the weekly average of the daily variation in mobility, excluding weekends. The mobility variation is computed by Google as the variation in the number of people visiting workplaces with respect to a pre-pandemic baseline. **(d)** Survey data (black dots) and piece-wise polynomial fit (dashed line) of the proportion of people declaring to avoid physical contacts over time, from the French survey CoviPrev[36]. In panels **(c)** and **(d)**, vertical grey bars indicate the periods of the three national lockdowns in France.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

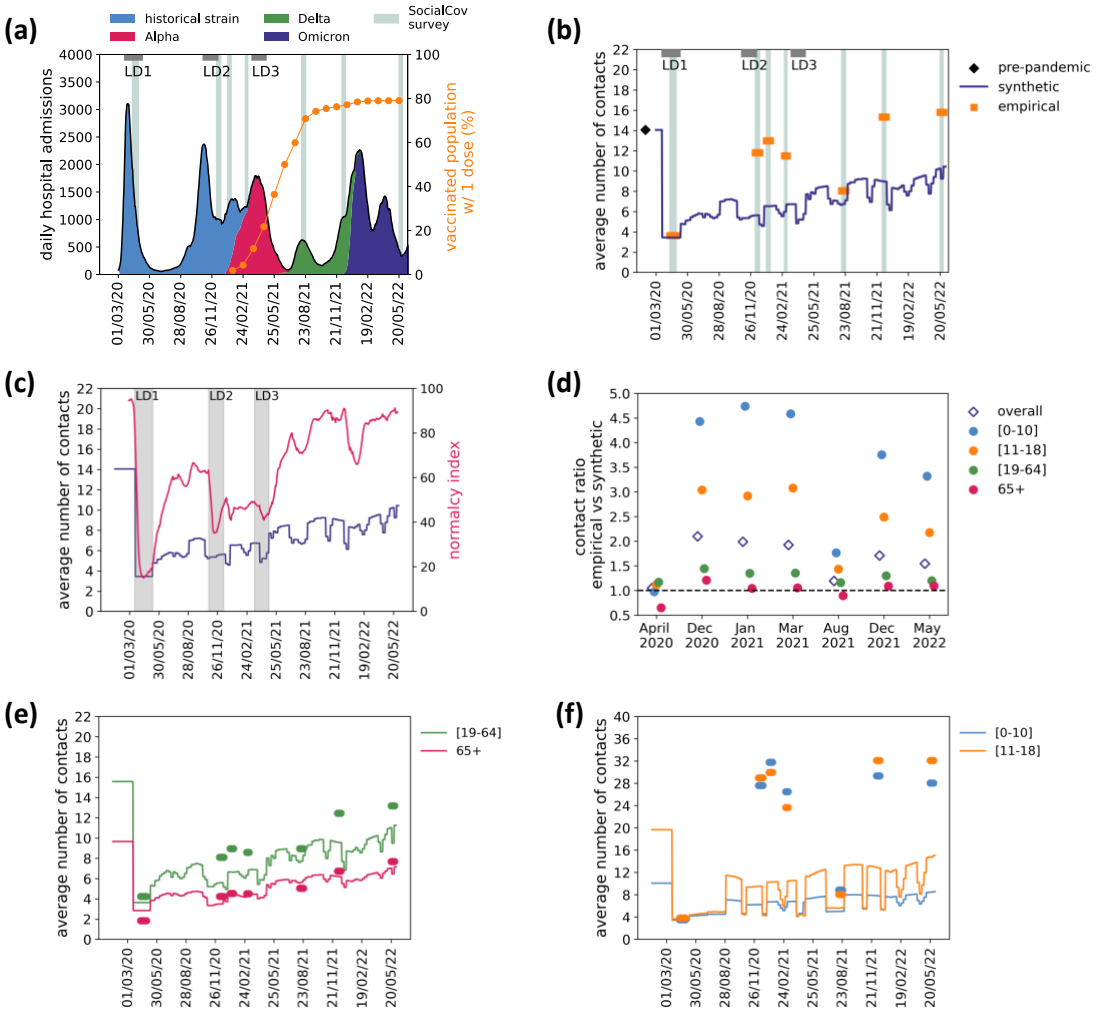

Figure 2. Synthetic and empirical contacts over time. (a) Timeline of the COVID-19 pandemic in France. Trajectory of daily hospital admissions (left y-axis) is shown in black, with colored areas indicating the frequency of SARS-CoV-2 circulating variants. Proportion of vaccinated population with 1 dose (right y-axis) is shown in orange. The grey horizontal bars annotated with LD indicate the periods of the three national lockdowns. Data collection periods for the empirical contacts (SocialCov surveys) are indicated with vertical shaded areas. **(b)** Average number of contacts over time, in the synthetic (dark blue line) and empirical contact matrices (orange dots). The value of the pre-pandemic empirical contact matrix used for baseline is shown in black. **(c)** Average number of contacts over time in the synthetic matrix (dark blue line, left y-axis) shown in comparison with the Normalcy Index (pink line, right y-axis), vertical grey bars indicate the periods of the three national lockdowns. **(d)** Ratio of the number of contacts estimated in the empirical contact matrices with respect to the synthetic matrices, broken down by survey wave (x-axis) and by age group (filled dots) or overall (void diamonds). **(e-f)** Average number of contacts over time by age groups (adults and seniors in panel e, children and adolescents in panel f), in the synthetic (lines) and in the empirical contact matrices (dots), color-coded as in panel (d).

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

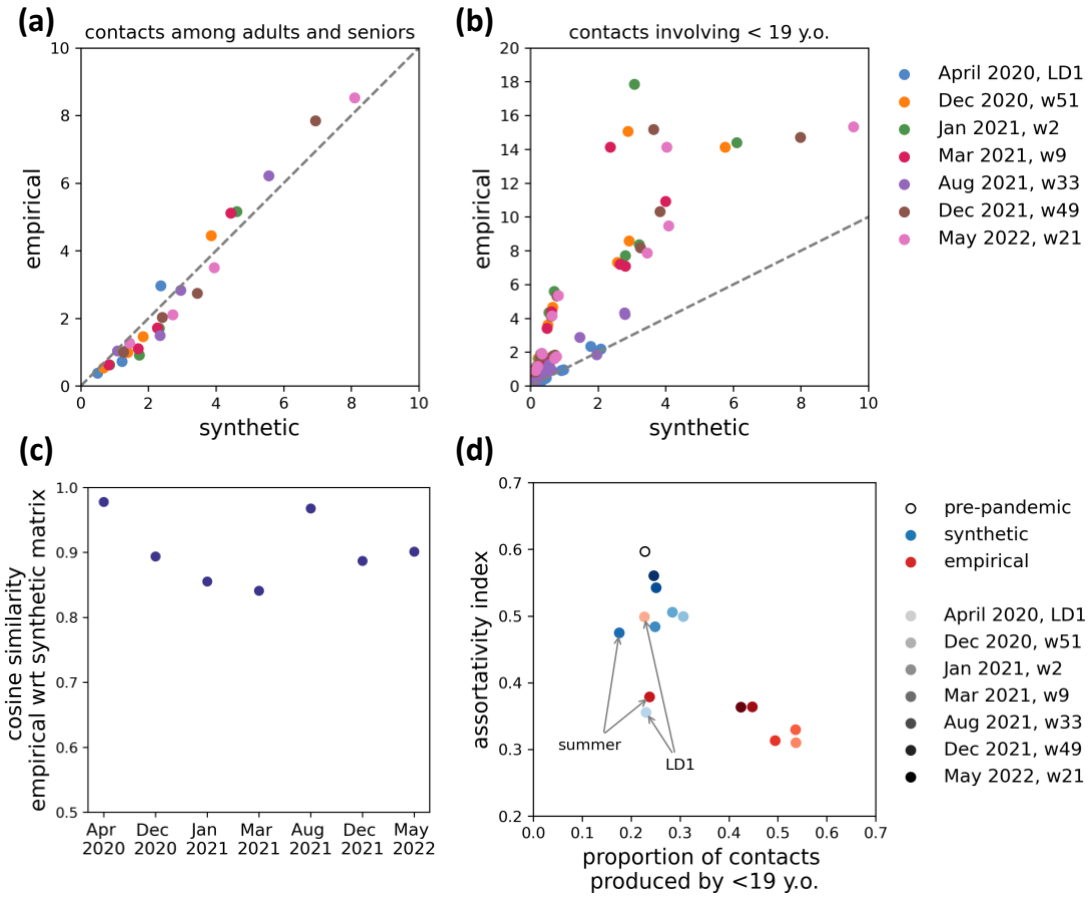

Figure 3. Comparison of contact patterns. (a) Contacts among adults and seniors, i.e. matrix elements M_{ij} with $i, j \in$ {[19-64], 65+}, in the synthetic (x-axis) vs the empirical matrices (y-axis). Colors indicate the seven waves. **(b)** As in panel **(a)**, showing the rest of the elements of the matrix involving individuals in the younger age groups. **(c)** Cosine similarity (invariant to contact intensity) between the survey-based and the corresponding synthetic contact matrices. **(d)** Proportion of the overall connectivity produced by young individuals (<19 y.o.) vs age-assortativity index, in the empirical (reds) and synthetic matrices (blues), for the seven survey periods. The value for the pre-pandemic contact matrix (void black dot) is shown for reference.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

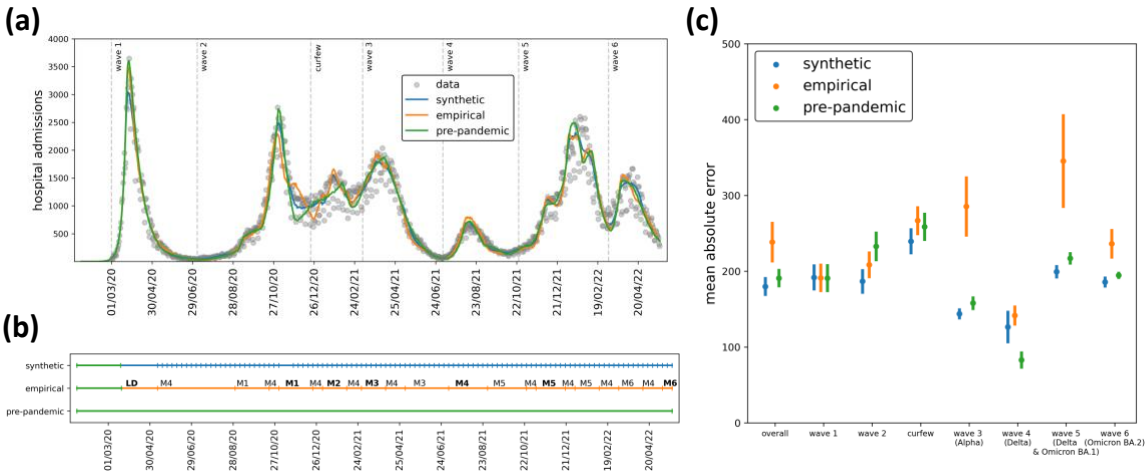

Figure 4. Model fit. (a) Model median trajectory of daily hospital admissions (median values out of 100 independent stochastic simulations), obtained by fitting the transmission model using weekly synthetic contact matrices (synthetic, blue), extended survey-based contact matrices (empirical, orange) and a constant pre-pandemic matrix (pre-pandemic, green). Data used for the fit are displayed with grey dots. Uncertainty around the median trajectory is displayed in Fig. S2. The vertical dashed lines define the epidemic phases used in panel (c). **(b)** Sequence of contact matrices used in the transmission model. Each row corresponds to one of the three models, i.e. the model informed with synthetic mobility-based contact matrices (first row), with survey-based empirical contact matrices (second row), or with a static pre-pandemic contact matrix (third row); each tick indicates a change in the contact matrix used in the model. Colors indicate the source of the matrix, i.e. pre-pandemic (green), synthetic (blue) and empirical (orange). The empirical contact matrices estimated from the 7 pandemic survey waves are denoted with LD, M1, M2 up to M6, following the notation in Methods and Table S6. They have been extended beyond the survey period to cover the whole study period. They are highlighted in bold in the periods that overlap the survey wave. **(c)** Mean absolute error (MAE) of daily model predictions with respect to the daily observed data, on the overall period (March 2020 – May 2022) and broken down by epidemic phase (epidemic waves and in-between periods). For each stochastic run, we computed the MAE as the sum of absolute differences between daily model predictions and the observed daily data, divided by the number of days in the epidemic phase under consideration. Dots and lines represent the average MAE and 95% confidence interval computed across 100 stochastic runs.

medRxiv preprint doi: https://doi.org/10.1101/2024.12.12.24318903; this version posted December 13, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

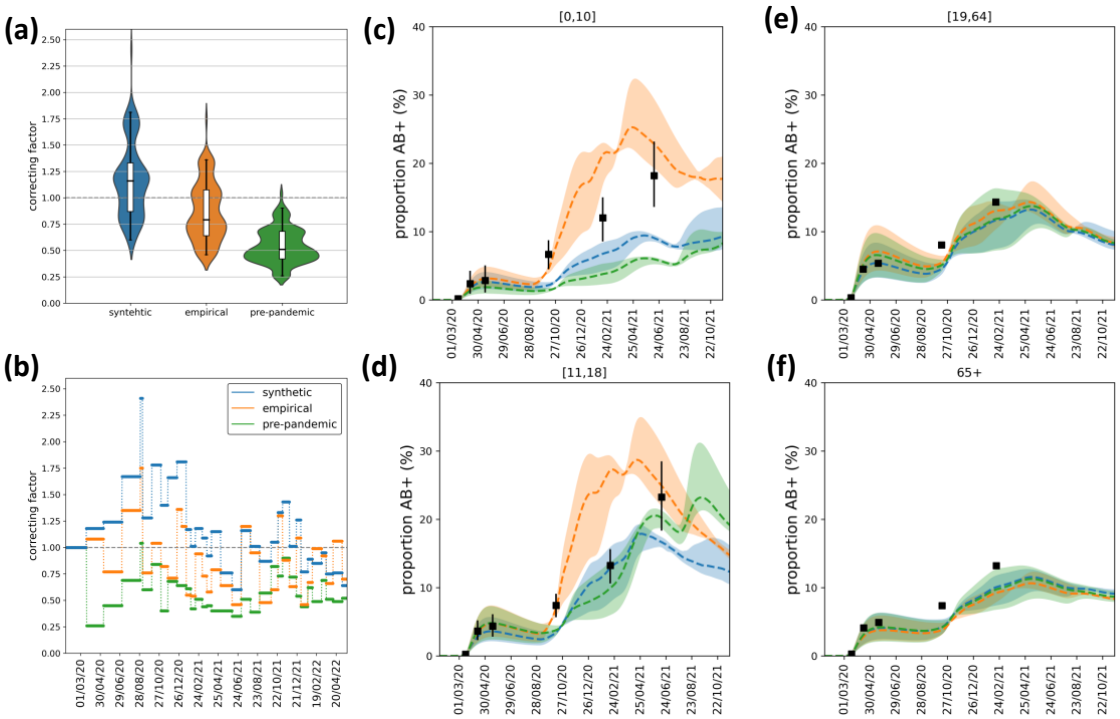

Figure 5. Model comparison. (a) Distribution of the correcting factor, fitted with the model using synthetic matrices (synthetic, blue), survey-based matrices (empirical, orange), and a constant pre-pandemic matrix (pre-pandemic, green). The box plot indicates median (line), interquartile range (box), and quantiles 2.5% and 97.5% (whiskers). The distribution by pandemic phase is shown in Fig. S4. **(b)** Correcting factor over time, fitted with the model using synthetic matrices, survey-based matrices, and a constant pre-pandemic matrix. **(c-f)** Proportion of antibody-positive population over time, estimated with the three models, by age class (in (c) children [0,10], in (d) adolescents [11-18], in (e) adults [19-64], in (f) seniors 65+). Dashed lines and shaded areas indicate the median and 95% probability ranges, respectively, computed across 100 stochastic simulations. Black symbols indicate estimates from serological data.