Assessing the effect of school closures on the spread of COVID-19 in Zurich

Maria Bekker-Nielsen Dunbar and Leonhard Held

on behalf of the SUSPend modelling consortium

Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich (UZH), Zurich, Switzerland.

Abstract. The effect of school closure on the spread of COVID-19 has been discussed intensively in the literature. To capture the interdependencies between children and adults we consider daily age-stratified incidence data and contact patterns between age groups which change over time based on social distancing policy indicators. We fit a multivariate time-series endemic-epidemic model to such data from the Canton of Zurich, Switzerland and use the model to predict the age-specific incidence in a counterfactual approach (with and without school closures). The results indicate a 21% median increase of incidence in the youngest age group (0-14), whereas the relative increase in the other age groups drops to values between 10% (15-24) and 1% (80+). We argue that our approach is more informative to policy makers than summarising the effect of school closures with time-dependent effective reproductive numbers, which are difficult to estimate due to the sparsity of incidence counts within the relevant age groups.

Keywords: COVID-19; endemic-epidemic modelling; surveillance data; social contacts; school closure

1. Introduction

Public health legislation authorises officials to order disease control measures such as closing schools. The usefulness of school closures has been seen in certain infectious disease outbreaks but not in others. School closure for infectious disease control may have additional health and wider societal effects as school is not just education but also has an important social function. Determining the usefulness of school closures is therefore of great value to policy makers. We know school closures have an effect on social mixing (Luca et al., 2018), though it would seem this effect might not be large for the early coronavirus outbreaks (European Centre for Disease Prevention and Control, 2020). For this reason, we are interested in examining the counterfactual scenario where school closure to combat coronavirus disease (COVID-19) was not introduced among school-aged children.

To examine true social distancing interventions implemented, we fit a multivariate endemic-epidemic model to the observed case data from the Canton of Zurich, Switzerland incorporating social contact patterns. Our model includes an age structure, which has been highlighted as important in models focusing on school closures (Jackson et al., 2014). We then predict from the fitted model given data until 17$^{\text{th}}$ March 2020 (the first day following declaration of state of emergency) with assumed time-varying contact weights implemented (Figure 1) and (the counterfactual scenario) with time-varying contact weights ignoring changes to contacts due to school closures. This affects the youngest age group (0-14 year olds) the age group that covers both compulsory and non-compulsory education. We then investigated the difference in the number of expected cases to evaluate the usefulness of school closures.

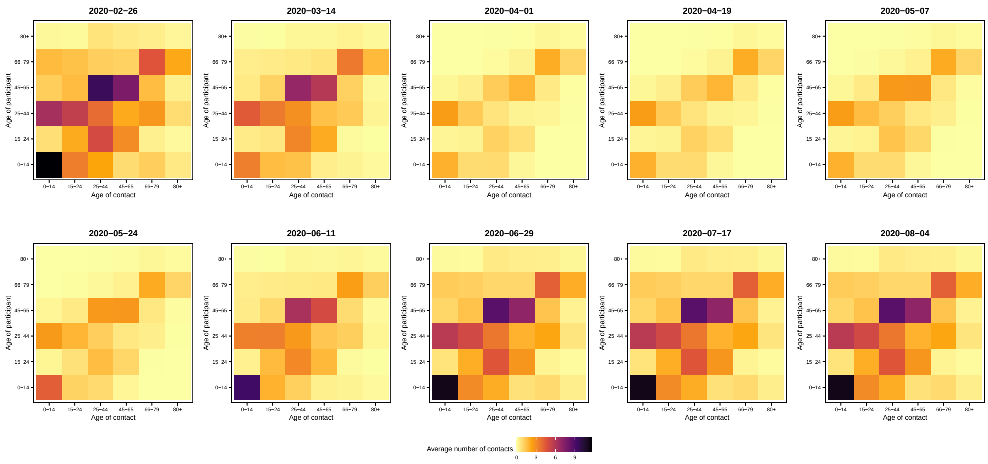

Figure 1. Snapshots of the time-varying contact matrix to reflect social distancing policies (this is the basis of our fitted model and prediction scenario A). Shown is the average number of contacts per day for individuals in the different age groups

2. Methods

Finkenstädt and Grenfell (2000) showcased how to formulate a time series susceptible-infected-recovered model through a case study of endemic measles infections and compared their methodology with established results from compartmental modelling, linking mathematical and statistical modelling. They highlighted a need to incorporate epidemic dynamics in statistical models; a need which was addressed by the endemic-epidemic (EE) modelling framework (introduced in Held et al., 2005).

2.1. Endemic-epidemic modelling

The EE framework is a time-series analysis-based method for infectious disease surveillance data. It can be derived from a mechanistic model of disease transmission (Höhle, 2016; Bauer and Wakefield, 2018; Wakefield et al., 2020), linking it with other modelling approaches. The multivariate formulation used in this work is the age-stratified EE model (Meyer and Held, 2017) where COVID-19 cases Y_{at} for age group a on day t are given by

$$Y_{at} \mid Y_{a,t-1}, \ldots, Y_{a,t-l} \sim \text{NegBin}(\lambda_{at}, \psi_a)$$
$$\lambda_{at} = \nu_{at} e_a + \phi_{at} \sum_{a'} c_{a,a',t} \sum_{l=1}^{l_{\max}} u_l Y_{a',t-l} \tag{1}$$

where λ_{at} is the mean and $\psi_a > 0$ the overdispersion parameter of a negative binomial distribution, where the limiting case $\psi_a \to 0$ represents the standard Poisson assumption. Overdispersion is sometimes termed k in the infectious disease literature when using negative binomial distributions to examine superspreading (e.g. Endo et al., 2019; Lloyd-Smith et al., 2020).

The mean λ_{at} is decomposed additively into an endemic (ν) and epidemic (ϕ) component. Age-specific proportions of population e_a enter as known offsets in the endemic component, whereas the epidemic component depends on contact weights $c_{a,a',t}$ representing transmission between age groups a and a' on day t, and u_l is the discrete-time serial interval distribution (Bracher and Held, 2020). We use a shifted (normalised) Poisson distribution with weights $u_l \propto \kappa^{l-1}/(l-1)! \exp(-\kappa), \kappa > 0$ with a maximum lag of $l_{\max} = 7$, as this has shown to be useful in other analyses of daily COVID-19 data (Grimée et al., 2021; Ssentongo et al., 2021). The transmission weights $c_{a,a',t}$ (entries of the contact matrix) are known but the serial interval lag distribution u_l (represented by the parameter κ) is not.

Various models were considered for the endemic and epidemic components of the model (see the supporting information for an overview of all models considered) and the best fitting model was determined based on the Bayesian information criterion (BIC). We always considered information on public holidays (as contacts may differ on those days) and daily testing rates to account for possible temporal changes in underascertainment. In addition to this, we also included daily temperature, linear time trends, and sine-cosine waves (a smooth non-linear trend not picked up by other parts of the model) as potential covariates in both components. The final model has 40 parameters and log-linear predictors given by:

$$\begin{aligned}\log(\nu_{at}) =& \beta_{\nu a} \mathbb{1}_{\{\text{age group } a\}}(a) + \beta_{\nu_{\text{day of the week}}} \mathbb{1}_{\{\text{weekday } t\}}(t) + \beta_{\nu_{\text{public holiday}}} \mathbb{1}_{\{t \text{ is a public holiday}\}}(t) \\ &+ \beta_{\nu_{\text{testing rate}}} T_t + \beta_{\nu_{\sin}} \sin(2\pi t/365) + \beta_{\nu_{\cos}} \cos(2\pi t/365)\end{aligned}$$

and

$$\begin{aligned}\log(\phi_{at}) =& \beta_{\phi a} \mathbb{1}_{\{\text{age group } a\}}(a) + \beta_{\phi_{\text{day of the week}}} \mathbb{1}_{\{\text{weekday } t\}}(t) + \beta_{\phi_{\text{public holiday}}} \mathbb{1}_{\{t \text{ is a public holiday}\}}(t) \\ &+ \beta_{\phi_{\text{testing rate}}} T_t + \beta_{\phi_{\text{time}}} t + \beta_{\phi_{\sin}} \sin(2\pi t/365) + \beta_{\phi_{\cos}} \cos(2\pi t/365)\end{aligned}$$

where T_t is the testing rate at time t. We calculate the amplitude and phase shift of each sinusoidal wave based on the sin/cos coefficients (Held and Paul, 2012), see the supporting information for details.

2.2. Time-varying contact matrices

Total contact matrices are made up of setting-specific contacts. We chose to use the Mistry et al. (2020) synthethic contact matrix compartments (Figure 2) in our work as

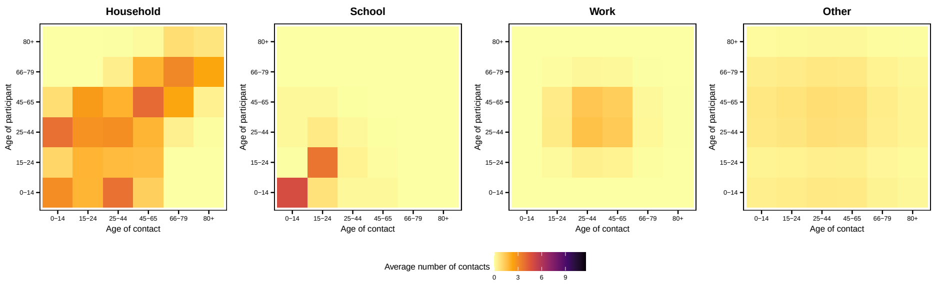

Figure 2. Synthetic contact matrix compartments which make up the first matrix in Figure 1 (the matrix before policy changes are applied)

they were the only ones we found provided with uncertainty estimates. Additionally, they were created with the European setting in mind, providing additional realism. We create Zurich-specific policy indicators following the methodology from Hale et al. (2020). As we are interested in reductions of contacts, we incorporate the policy indicators such that they take values between 0 and 1, where a higher value is a situation with less social distancing policy in place (Figure 3). This ensures that no metric used to incorporate policy changes increases contacts, thereby creating an artificially inflated baseline. See the supporting information for indicator construction details. The results of applying these indicators to the components of the contact matrix and combining using the Mistry et al. (2020) weights are showcased in Figure 1. We additionally reduced contacts in school settings on school holidays in both versions of the time-varying contact matrices as school holidays are known to cause a drop in contacts (Eames et al., 2012).

2.3. Counterfactual analysis

To examine the effect of school closures, we used a version of the time-varying contact matrix $c_{a,a',t}$ which does not have reductions applied to contacts in the school setting among the youngest age group, only regular school holidays. We predicted the course of the epidemic under two scenarios (as was; scenario A and adjusted to not have school closure; scenario B). The predictions are obtained using the methodology described in Held et al. (2017, Appendix A). To analyse the counterfactual scenario, we adjust the weights (time-varying contact matrix) in the EE model and calculate the predictive mean vector for the adjusted path forecast. That is, after fitting our EE model (described above) we predicted the epidemic from 17$^{\text{th}}$ March 2020 with the two options for time-varying matrices. To compare the two scenarios, we considered the absolute and relative increases in predicted cases between scenarios B and A. We expect both to be positive as scenario B is a deviation from the true non-pharmaceutical measures which were implemented and has a lower level of disease control in place.

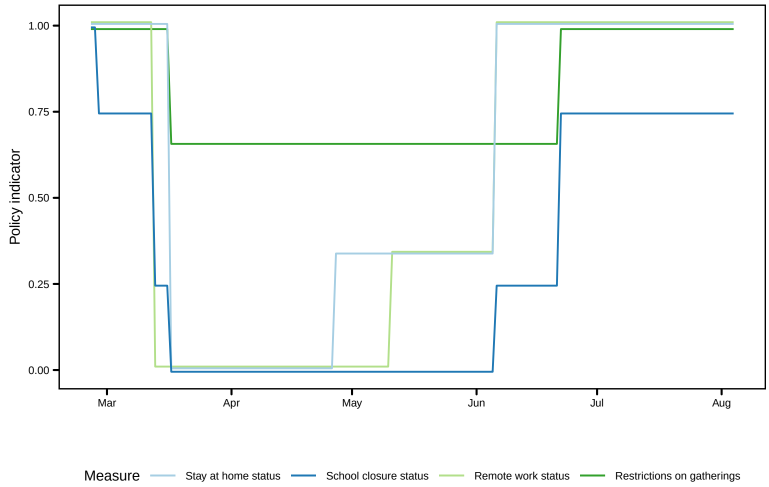

Figure 3. Policy indicators used to adjust contacts through the effect on the four components: household (affected by requirements to stay at home), school (affected by school closure and additional school holidays (not shown here)), work (affected by remove work), and other (affected by restrictions on gatherings). A slight jitter has been applied to ease comparison of step functions

2.4. Incorporating uncertainty

The importance of acknowledging uncertainty in COVID-19 modelling was highlighted by Davey Smith et al. (2020). We take into account both external (contact matrix weights) and internal (model coefficient estimates) parameter uncertainty in the counterfactual analysis. We used the weights reported by Mistry et al. (2020) to estimate a set of plausible values for the different contact matrix compartments to obtain total contacts across settings. We sampled the weights from normal distributions with the estimated means and standard errors of the reported weights used to create the time-varying contact matrix, allowing us to incorporate the corresponding external uncertainty. In particular, we simulated 1000 weights and created 1000 versions of the time-varying contact matrices.

Additionally, for each of $n = 1000$ versions of the EE model with those time-varying contact matrices, we sampled all entries in Table 1 using a 40-dimensional multivariate normal distribution based on the estimated parameter vector and the associated variance-covariance matrix. While Figure 4 and Table 1 show the results for using the Mistry et al. (2020) weights and the parameter estimates as given, our main results in Table 2 incorporate both external (contact matrices) and internal (model) uncertainties.

3. Results

The selected model has the lag distribution and fit shown in Figure 4. We see that the model captures the patterns observed in the data well, apart from the youngest age group, though this is likely an artefact of the low number of cases overall in that group. The

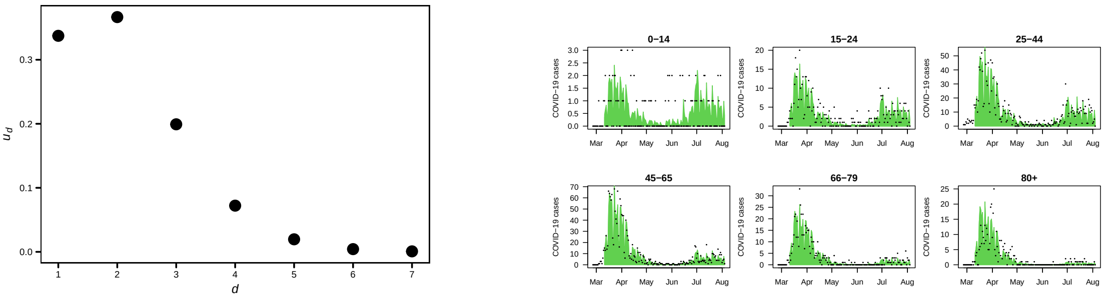

Figure 4. Lag distribution u_l (left) and model fit to observed case counts (right)

shape of the serial interval distribution (Figure 4, left) is right-skewed with a sharp peak early on. The peak found in this work is earlier than expected based on the literature and so we have conducted a sensitivity analysis with other estimates from EE models (Grimée et al., 2021; Ssentongo et al., 2021), see supporting information for details. The model coefficients are listed in Table 1 and the model fit is shown in Figure 4 (right). In Table 1, ν and ϕ denote the endemic and epidemic components, respectively. The coefficients (Table 1) show the expected pattern with a strong day of the week effect. We also see a seasonal pattern in particular in the epidemic component (amplitude and phase) which is counterbalanced by a positive time trend. There is considerable variation in the endemic and epidemic intercepts between age groups. Testing rate has a positive and significant effect in the epidemic component whereas the effect on the endemic component is less pronounced.

The result of the counterfactual analysis is given in Table 2. The distributions of our predicted counts are very skewed, which is why we provide median and percentiles as point and interval estimates rather than means and standard deviations. We see that the number of cases never increases more than 21 per cent and, as would be expected, that most of the increase is found among the youngest age group. The largest increase in case counts is found in the age group 25-44, who were not considered a vulnerable group at the time. Based on the median we see 261 additional cases (a 6% increase) with 5 additional cases in the oldest age group (which might be considered of most concern, those aged 80 and over). However the 90 per cent quantiles are considerably larger (1112 and 22 cases, respectively). Similar patterns of burden (which age groups have the highest case counts under the two scenarios) are seen in both scenarios. The large uncertainty seen is not so surprising as the data from Zurich is rather sparse. If we had instead applied the method to a larger population, such as the entire federation of Switzerland, precision would be increased.

We also compared the temporal dynamic in scenario B vs. A. An increase in cases is observed in age group 0-14 already in April but only later in the other age groups. The next age group to experience an increase in cases is 25-44, the parents of age group 0-14. Thus, the dynamic follows the contact patterns and underlines the importance of including both age-stratified data and contact matrices in models. See the supporting

Table 1. Coefficients of model visualised in Figure 4

Coefficient	Estimate	Standard error	Coefficient	Estimate	Standard error
$\beta_{\nu_{\text{day of the week Tuesday}}}$	0.421	0.099	$\beta_{\phi_{\text{day of the week Saturday}}}$	-0.484	0.073
$\beta_{\nu_{\text{day of the week Wednesday}}}$	-0.065	0.094	$\beta_{\phi_{\text{day of the week Sunday}}}$	-0.330	0.093
$\beta_{\nu_{\text{day of the week Thursday}}}$	0.122	0.086	$\beta_{\phi_{\text{public holiday}}}$	-1.194	0.261
$\beta_{\nu_{\text{day of the week Friday}}}$	0.033	0.087	$\beta_{\phi_{\text{test rate}}}$	0.013	0.002
$\beta_{\nu_{\text{day of the week Saturday}}}$	-0.899	0.124	$\beta_{\phi_{\text{time}}}$	0.222	0.028
$\beta_{\nu_{\text{day of the week Sunday}}}$	-0.719	0.150	$\beta_{\phi_{\text{amplitude}}}$	19.721	0.900
$\beta_{\nu_{\text{public holiday}}}$	-0.031	0.432	$\beta_{\phi_{\text{phase}}}$	1.799	0.088
$\beta_{\nu_{\text{test rate}}}$	-0.003	0.002	$\beta_{\phi 0-14}$	-4.307	0.946
$\beta_{\nu_{\text{amplitude}}}$	2.186	0.507	$\beta_{\phi 15-24}$	-2.268	0.872
$\beta_{\nu_{\text{phase}}}$	-0.275	0.123	$\beta_{\phi 25-44}$	-0.973	0.858
$\beta_{\nu 0-14}$	-5.432	0.523	$\beta_{\phi 45-65}$	-0.707	0.853
$\beta_{\nu 15-24}$	-4.562	0.491	$\beta_{\phi 66-79}$	-1.725	0.857
$\beta_{\nu 25-44}$	-3.654	0.489	$\beta_{\phi 80+}$	-1.904	0.847
$\beta_{\nu 45-65}$	-4.754	0.495	ψ_{0-14}	0.543	0.364
$\beta_{\nu 66-79}$	-4.649	0.504	ψ_{15-24}	0.044	0.050
$\beta_{\nu 80+}$	-4.627	0.559	ψ_{25-44}	0.063	0.024
$\beta_{\phi_{\text{day of the week Tuesday}}}$	0.226	0.057	ψ_{45-65}	0.004	0.013
$\beta_{\phi_{\text{day of the week Wednesday}}}$	-0.054	0.061	ψ_{66-79}	0.022	0.030
$\beta_{\phi_{\text{day of the week Thursday}}}$	-0.128	0.063	ψ_{80+}	0.288	0.102
$\beta_{\phi_{\text{day of the week Friday}}}$	0.208	0.059	$\log \kappa$	0.083	

Table 2. Distribution of predicted cases and comparative measures. P_{10} and P_{90} denote the 10 and 90 percentiles. Values are calculated based on the corresponding samples, including the differences B-A and the ratios B/A

	Scenario A			Scenario B			B - A			B / A		
Age	P_{10}	Median	P_{90}	P_{10}	Median	P_{90}	P_{10}	Median	P_{90}	P_{10}	Median	P_{90}
0-14	57	78	134	66	95	189	7.3	16.2	55.8	1.12	1.21	1.44
15-24	340	410	628	355	449	800	11.8	40.1	176.3	1.03	1.10	1.28
25-44	1187	1364	1936	1230	1482	2427	40.6	119.3	508.3	1.03	1.09	1.27
45-65	1409	1523	1862	1432	1585	2121	17.8	62.0	269.8	1.01	1.04	1.15
66-79	477	521	602	487	538	670	6.1	17.0	68.6	1.01	1.03	1.12
80+	312	353	405	317	360	420	1.6	4.8	21.7	1.00	1.01	1.06
Total (summed)	3844	4236	5530	3940	4504	6608	86.5	260.7	1112.1	1.02	1.06	1.21

information for an illustration and description of the patterns.

4. Conclusion

Understanding the effect of disease control interventions is useful for preparedness for future epidemics. We found it possible to utilise time-varying contact information in the EE modelling framework and fit an appropriate model to COVID-19 case surveillance data. We adjusted the contact matrix based on a timeline of COVID-19 events focused on Zurich. Our focus was school-based social distancing measures, as these might be considered a priority for policy makers when choosing exit strategies or phasing out of measures. In our counterfactual analysis, we did not assume interventions were applied equally to the entire population but we acknowledge that changes may affect other parts of the population through contact patterns between age groups and so the effects of school closure are not to be considered in isolation.

In this work we have assumed that the model parameters do not change when considering scenario B. However, some natural experiments are occurring where some schools are closed and other schools are kept open during the same time period in the same school district, see e.g. Berger et al. (2020) who examined whether increased testing can be used in conjunction with keeping schools open. This may allow us to re-evaluate that assumption in the future. Vlachos et al. (2021) have estimated the effect of school closures on the spread of SARS-CoV-2 virus (the agent causing COVID-19) among parents and teachers in Sweden, where lower-secondary schools (pupils aged 14-16) remained open during the first wave. Their results indicate that keeping lower-secondary schools open had minor consequences for the overall transmission, in line with our results. If information on mask usage is available it could be included in our model as a different type of reduction to contacts as a proxy for its effect on transmission events.

The estimation of age- and time-dependent multiplication factors to address additional causes of underreporting (Noufaily, 2020), such as the presence of asymptomatic COVID-19 infections are being considered in our ongoing work. We are examining the option of using reporting rates as the basis of such multiplication factors. The rates are based on adjusting case fatality rates for delays between hospitalisation and deaths in the vein of Russell et al. (2020) and Nishiura et al. (2009). Such an approach allows us to address underascertainment (capture asymptomatic cases which are not likely to be reported as well as those symptomatic cases that are reported). We expect underreporting to likely be more pronounced among children because they may have more asymptomatic cases (causing them to have fewer cases in the data). Our model will need suitable amendments once such multiplication factors are available.

Flaxman et al. (2020) studied the effect of non-pharmaceutical interventions on COVID-19 in Europe based on overall time-varying reproduction numbers (R_t). School closures are an intervention directly affecting only children and adolescents. Mortality in these age groups is notably very low, so our study aimed to quantify the indirect effects of school closures in other age groups based on age-stratified incidence data and appropriate contact information. We argue that risk communication strategies regarding disease control initiatives for COVID-19 should be based on age-specific (rather than overall) effective reproduction numbers, but those are difficult to estimate if data are sparse. This seems of particular importance if the interventions being considered are specific to certain groups of the population rather than overarching.

When a single summary indicator is reported to the public, uncertainty of its estimation should be included. For this reason we believe our work to be particularly useful as we capture a lot of the statistical uncertainty inherent to our work. While the effective reproduction number is an attractive summary measure due to the threshold property (> 1 indicates continued epidemic growth), some of the nuances behind its calculation will be lost to summarisation. These include the fact that R_t can be approximated in different ways depending on the data at hand and model considered. This echoes Becker (2015) and recent comments made by the European Union's Joint Research Centre (Annunziato and Asikainen, 2020) and Gostic et al. (2020), as well as the aforementioned comment on uncertainty by Davey Smith et al. (2020). Additional context should be provided as setting-specific disease control challenges in which the estimate was obtained need to be communicated as well as R_t, e.g. healthcare surge capacity, cases in nursing homes compared to cases in schools, and the impact of superspreading events although community transmission is low. We suggest other indicators, such as predicted case counts (as found here) are used alongside R_t to provide broader context as it is not fully informative in isolation. For additional discussion, see the supporting information.

Code and data

Code and publicly available data used in this manuscript can be accessed via `https://gitlab.switch.ch/suspend/COVID-19-school-ZH`. Project updates and associated analyses are available via `https://suspend.pages.switch.ch/project`. For more information and extended descriptions of our work, see the supporting information.

Acknowledgments

This work would not be possible without the financial support of the Swiss National Science Foundation (`https://data.snf.ch/covid-19/snsf/196247`) and the confidential data provided by Gesundheitsdirektion Zürich. They did not play any role in the formulation of this manuscript and the views expressed in this manuscript are those of the authors. The additional members of the SUSPend modelling consortium (in anti-alphabetical order by surname) are: Jon Wakefield, Sebastian Meyer, Felix Hofmann, Niel Hens, Mathilde Grimée, Deborah Chiavi, and Johannes Bracher (`https://suspend.pages.switch.ch/project`).

References

Annunziato, A. and T. Asikainen (2020). Effective reproduction number estimation from data series. JRC121343.

Bauer, C. and J. Wakefield (2018). Stratified spacetime infectious disease modelling, with an application to hand, foot and mouth disease in China. *J. R. Stat. Soc. C Appl. Stat 67*(5), 1379–1398.

Becker, N. G. (2015). *Modeling to Inform Infectious Disease Control*. Biostatistics series. Chapman & Hall/CRC.

Berger, U., C. Fritz, and G. Kauermann (2020). Schulschließungen oder Schulöffnung mit Testpflicht? Epidemiologisch-statistische Aspekte sprechen für Schulöffnungen mit verpflichtenden Tests. *CODAG Bericht 14*.

Bracher, J. and L. Held (2020). Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. *Int. J. Forecast.*

Davey Smith, G., M. Blastland, and M. Munafò (2020). Covid-19's known unknowns. *BMJ 371*.

Eames, K. T. D., N. L. Tilston, E. Brooks-Pollock, and W. J. Edmunds (2012). Measured dynamic social contact patterns explain the spread of H1N1v influenza. *PLOS Comput. Biol 8*(3), 1–8.

Endo, A., M. Uchida, A. J. Kucharski, and S. Funk (2019). Fine-scale family structure shapes influenza transmission risk in households: Insights from primary schools in Matsumoto city, 2014/15. *PLOS Comput. Biol. 15*(12), 1–18.

European Centre for Disease Prevention and Control (2020). COVID-19 in children and the role of school settings in COVID-19 transmission. https://www.ecdc.europa.eu/en/publications-data/children-and-school-settings-covid-19-transmission.

Finkenstädt, B. F. and B. T. Grenfell (2000). Time series modelling of childhood diseases: a dynamical systems approach. *J. R. Stat. Soc. C Appl. Stat 49*(2), 187–205.

Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, M. Monod, Imperial College COVID-19 Response Team, A. C. Ghani, C. A. Donnelly, S. Riley, M. A. C. Vollmer, N. M. Ferguson, L. C. Okell, and S. Bhatt (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature 584*, 257261.

Gostic, K. M., L. McGough, E. B. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, R. Kahn, R. Niehus, J. A. Hay, P. M. De Salazar, J. Hellewell, S. Meakin, J. D. Munday, N. I. Bosse, K. Sherrat, R. N. Thompson, L. F. White, J. S. Huisman, J. Scire, S. Bonhoeffer, T. Stadler, J. Wallinga, S. Funk, M. Lipsitch, and S. Cobey (2020, 12). Practical considerations for measuring the effective reproductive number, R_t. *PLOS Comput. Biol. 16*(12), 1–21.

Grimée, M., M. Bekker-Nielsen Dunbar, F. Hofmann, and L. Held (2021). Modelling the effect of a border closure between Switzerland and Italy on the spatiotemporal spread of COVID-19 in Switzerland. *medRxiv*.

Hale, T., T. Boby, N. Angrist, E. Cameron-Blake, L. Hallas, B. Kira, S. Majumdar, A. Petherick, T. Phillips, J. Tatlow, and S. Webster (2020). Variation in government responses to COVID-19 working paper. https://covidtracker.bsg.ox.ac.uk.

Held, L., M. Höhle, and M. Hofmann (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Stat. Model 5*(3), 187–199.

Held, L., S. Meyer, and J. Bracher (2017). Probabilistic forecasting in infectious disease epidemiology: the 13th Armitage lecture. *Stat. Med 36*(22), 3443–3460.

Held, L. and M. Paul (2012). Modeling seasonality in space-time infectious disease surveillance data. *Biom. J. 54*(6), 824–843.

Höhle, M. (2016). Infectious disease modelling. In A. Lawson, S. Banerjee, R. Haining, and M. Ugarte (Eds.), *Handbook of Spatial Epidemiology*, pp. 477–500. Chapman & Hall/CRC Handbooks of Modern Statistical Methods.

Jackson, C., P. Mangtani, J. Hawker, B. Olowokure, and E. Vynnycky (2014). The effects of school closures on influenza outbreaks and pandemics: Systematic review of simulation studies. *PLOS ONE 9*(5), 1–10.

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz (2020). Superspreading and the effect of individual variation on disease emergence. *Nature 438*, 355–359.

Luca, G. D., K. van Kerckhove, P. Coletti, C. Poletto, N. Bossuyt, N. Hens, and V. Colizza (2018). The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium. *BMC Infect. Dis. 18*(29).

Meyer, S. and L. Held (2017). Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics 18*, 338–351.

Mistry, D., M. Litvinova, A. Pastore y Piontti, M. Chinazzi, L. Fumanelli, M. F. C. Gomes, S. A. Haque, Q.-H. Liu, K. Mu, X. Xiong, M. E. Halloran, I. M. Longini Jr., S. Merler, M. Ajelli, and A. Vespignani (2020). Inferring high-resolution human mixing patterns for disease modeling. *arXiv.* https://arxiv.org/abs/2003.01214.

Nishiura, H., D. Klinkenberg, M. Roberts, and J. A. P. Heesterbeek (2009). Early epidemiological assessment of the virulence of emerging infectious diseases: A case study of an influenza pandemic. *PLOS ONE 4*(8), 1–8.

Noufaily, A. (2020). Underreporting and reporting delays. In L. Held, N. Hens, P. O'Neill, and J. Wallinga (Eds.), *Handbook of Infectious Disease Data Analysis*, Handbooks of Modern Statistical Methods, pp. 437–454. CRC Press.

Russell, T. W., N. Golding, J. Hellewell, S. Abbott, L. Wright, C. A. B. Pearson, K. van Zandvoort, C. I. Jarvis, H. Gibbs, Y. Liu, R. M. Eggo, W. J. Edmunds, A. J. Kucharski, and CMMID COVID-19 working group (2020). Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med 18*(1), 332.

Ssentongo, P., C. Fronterre, A. Geronimo, S. J. Greybush, P. K. Mbabazi, J. Muvawala, S. B. Nahalamba, P. O. Omadi, B. T. Opar, S. A. Sinnar, Y. Wang, A. J. Whalen, L. Held, C. Jewell, A. J. B. Muwanguzi, H. Greatrex, M. M. Norton, P. Diggle, and S. J. Schiff (2021). Tracking and predicting the African COVID-19 pandemic. *Proc. Natl. Acad. Sci. (in press).*

Vlachos, J., E. Hertegård, and H. B. Svaleryd (2021). The effects of school closures on SARS-CoV-2 among parents and teachers. *Proc. Natl. Acad. Sci 118*(9).

Wakefield, J., T. C. Dong, and V. N. Minin (2020). Spatio-temporal analysis of surveillance data. In L. Held, N. Hens, P. O'Neill, and J. Wallinga (Eds.), *Handbook of Infectious Disease Data Analysis*, Handbooks of Modern Statistical Methods, pp. 455–476. Chapman & Hall/CRC.