RESEARCH NOTE

SOCRATES: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for COVID-19

Lander Willem[1]*, Thang Van Hoang[2], Sebastian Funk[3], Pietro Coletti[2], Philippe Beutels[1,4] and Niel Hens[1,2]

Abstract

Objective: Establishing a social contact data sharing initiative and an interactive tool to assess mitigation strategies for COVID-19.

Results: We organized data sharing of published social contact surveys via online repositories and formatting guidelines. We analyzed this social contact data in terms of weighted social contact matrices, next generation matrices, relative incidence and R_0. We incorporated location-specific physical distancing measures (e.g. school closure or at work) and capture their effect on transmission dynamics. All methods have been implemented in an online application based on R Shiny and applied to COVID-19 with age-specific susceptibility and infectiousness. Using our online tool with the available social contact data, we illustrate that physical distancing could have a considerable impact on reducing transmission for COVID-19. The effect itself depends on assumptions made about disease-specific characteristics and the choice of intervention(s).

Keywords: Social contact data, User interface, Transmission dynamics, Infectious diseases, Epidemics, Social distancing, Behavioral changes, Data sharing initiative, Open-source, COVID-19

Introduction

Given the pandemic of SARS-CoV-2, which causes COVID-19 disease, it is of great importance to consider intervention strategies to slow down SARS-CoV-2 spread, and thus decrease surge capacity problems arising to health care provision and essential supplies [1, 2]. Physical distancing on a large scale, first at the epicenter of the outbreak in Wuhan, and later in other locations was shown to slow down SARS-CoV-2 spread (e.g. in Shanghai) [3].

Social contact surveys have proven to be an invaluable source of information about how people mix in the population [4–6] and explained close contact infectious disease data well [7–9]. For example, adapted social mixing during the A(H1N1)v2009 pandemic was fundamental to reproduce the observed incidence patterns [10]. In terms of prevention strategies, social contact data from the POLYMOD project [5] have been used to quantify the impact of school closure on the spread of airborne infections [11]. This was done by comparing the basic reproduction number R_0, or the average number of secondary infections caused by a single infectious individual in a completely susceptible population, derived from mixing patterns observed on weekends or during a holiday period with those derived from mixing patterns observed on weekdays.

In this research note, we highlight a social contact data sharing initiative and present an online tool to facilitate data access and analysis. Physical distancing measures can be mimicked with this tool by excluding the contribution of mixing patterns at specific locations to investigate the impact on disease transmission and guide

*Correspondence: lander.willem@uantwerp.be
[1] Centre for Health Economic Research and Modelling Infectious Diseases, University of Antwerp, Antwerp, Belgium
Full list of author information is available at the end of the article

policy makers. As a case study in light of COVID-19, we exploit our application to quantify the potential impact of school closure and physical distancing at work due to non-pharmaceutical interventions, a shift from common workplaces to teleworking at home or (temporary) unemployment.

Main text

Methods

Following a systematic literature review [4], corresponding authors were contacted to share their data subject to ethical approvals and GDPR compliance. All data have been refactored according to guidelines we developed during a Social Contact Data Hackaton in 2017 as part of the TransMID project. Each survey is split into multiple files to capture participant, contact, survey day, household and time-use data. For each data type, there is one "common" file and one "extra" file in which more specific variables related to the survey are included. Each data set contains a dictionary to interpret the columns (see http://www.socialcontactdata.org for more information).

To extrapolate survey data to the country level and obtain social contact rates on a weekly basis, we incorporate participant weights accounting for age and the number of observations during week (5/7) and weekend (2/7) days. We use the United Nation's World Population Prospects [12] as reference and constrain weights to a maximum of 3 to limit the influence of single participants. The social contact matrix m_{ij} can be estimated by:

$$m_{ij} = \frac{\sum_{t=1}^{T_i} w_{it}^d y_{ijt}}{\sum_{t=1}^{T_i} w_{it}^d}, \tag{1}$$

where w_{it}^d denotes the weight for participant t of age i who was surveyed on day type $d \in \{$ weekday , weekend $\}$, y_{ijt} denotes the reported number of contacts made by participant t of age i with someone of age j and T_i denotes all participants of age i. By nature, contacts are reciprocal and thus $m_{ij}N_i$ should be equal to $m_{ji}N_j$. To resolve differences in reporting, reciprocity can be imposed by:

$$m_{ij}^{\text{reciprocal}} = \frac{m_{ij}N_i + m_{ji}N_j}{2N_i}, \tag{2}$$

with N_i and N_j the population size in age class i and j, respectively [13]. This reciprocal behavior might not be valid for specific contact types, e.g. contacts at work for retail workers are most likely not contacts at work for their customers.

Transmission dynamics can be represented by the next generation matrix G with elements g_{ij} that indicate the average number of secondary infections in age class i through the introduction of a single infectious individual of age class j into a fully susceptible population [14]. The next generation matrix is defined by:

$$G = DMq, \tag{3}$$

with D the mean duration of infectiousness, M the contact matrix and q a proportionality factor [9, 11]. The proportionality factor q combines several disease-specific characteristics that are related to susceptibility and infectiousness. Equation 3 can be reformulated as:

$$g_{ij} = D * m_{ij} * s_i * k_j * \hat{q}, \tag{4}$$

where s_i denotes the susceptibility of age group i, k_j the infectiousness of age group j and $\hat{q}$ other disease-specific factors. The leading right eigenvector of G is proportional to the expected incidence by age and R_0 can be calculated as the dominant eigenvalue of G [5].

To evaluate intervention strategies, we focus on the relative impact of adjusted social contact patterns on R_0 in line with the so-called *social contact hypothesis* [7] by cancelling disease specific features:

$$\frac{R_{0a}}{R_{0b}} = \frac{\max(\text{eigen}(DM_aq))}{\max(\text{eigen}(DM_bq))} = \frac{\max(\text{eigen}(M_a * S * K))}{\max(\text{eigen}(M_b * S * K))}, \tag{5}$$

where indices a and b refer to the different conditions, and S and K account for age-specific susceptibility and infectiousness, respectively [11]. Physical distancing can be evaluated by the elimination or reduction of location-specific subsets of the social contact data. Contacts reported at multiple locations are assigned to a single location in the following hierarchical order: home, work, school, transport, leisure and other locations. We simulate school closure by excluding all contacts reported at school. We evaluate physical distancing at work by applying a proportional reduction of the social contacts reported at work ($p_{\text{workplace}}^{\text{distancing}}$). To combine the effect of school closure and distancing at work, the social contact matrix M is calculated as:

$$M = M_{\text{home}} + (M_{\text{work}} * (1 - p_{\text{workplace}}^{\text{distancing}}))$$
$$+ (M_{\text{school}} * 0) + M_{\text{transport}} + M_{\text{leisure}} + M_{\text{other}} \tag{6}$$

We developed an interactive application to access and analyze social contact data based on R packages *shiny* [15] and *socialmixr* [16]. The user interface enables the selection of country-specific data, age categories, type of day, contact duration, intensity and gender. Using selection boxes, the user can opt to disable the assumption of reciprocity and participant weights. The user can also enable distancing strategies such as school closure or

Willem *et al. BMC Res Notes*   *(2020) 13:293*

Page 3 of 8

physical distancing at work, or include age-specific transmission parameters.

The user interface contains a plot of the social contact matrix and the principal results of the social contact analysis: *M*, relative incidences, the reference demography, participant statistics info on the data sets. Relative R_0 and *M* ratios are printed if reactive strategies are selected.

As COVID-19 case study, we estimate the effect of school closure and physical distancing at work on disease transmission dynamics. In order to do this, we use 3 age classes: 0–18 years, 19–60 years and over 60 years of age. For each country, we calculate contact rates after excluding data from holiday periods. We capture transmission dynamics with 0%, 20%, 40% and 60% distancing at work, with and without school closure. As proof of concept, we include the scenario where children are less vulnerable compared to elderly $[s_i = k_j = (0.5, 1, 1.5)]$, instead of uniform susceptibility and infectiousness.

Results
The http://www.socialcontactdata.org initiative, status 25th May 2020, includes data for Belgium, Finland, Germany, Italy, Luxembourg, Netherlands, Poland and the UK from POLYMOD [5], as well as data from other studies on social mixing in France [17], China [18], Hong Kong [19], Peru [20], UK [21], Russia [22], Zimbabwe [23], Vietnam [24], South Africa and Zambia [25]. All data are available on Zenodo [26–35] and can be retrieved within R using the *socialmixr* package.

The SOcial Contact RATES (Socrates) data tool [36, 37] enables quick and convenient generation of social contact matrices, relevant for the spread of infectious diseases. Figure 1 presents a screenshot of the user interface. The potential of using social contact patterns to simulate infectious disease transmission are endless, and we hope with this initiative to support data-driven modeling endeavors. The survey data from France and Zimbabwe contain multiple days per participant, hence we included only the first day for each participant to minimize the effect of reporting fatigue.

We demonstrate the effect of physical distancing at work and school closure on R_0 in Fig. 2. If we assume uniform susceptibility and infectiousness, we predict for most countries a 10% decrease in R_0 with workplace distancing of 60%. For Poland and Hong Kong, the reduction is slightly higher. The analysis for Peru shows little impact of workplace distancing since only few contacts were reported "at work", whereas a substantial proportion of contacts was reported at the market or street. Cultural

differences in how "at work" is understood should be considered when interpreting results. The data for Zimbabwe contains also relatively few reported contacts at work, which translates into a limited impact of workplace distancing in our analysis. The estimated R_0 reduction due to school closure is more country-specific, e.g. 10% reduction for Belgium and Vietnam, but 20% for Italy, Luxembourg and France. If we assume that elderly are more vulnerable compared to children, as might be the case for COVID-19 [38], the impact of school closure decreases dramatically. The positive effect of physical distancing at work on R_0 remains the same or increases.

The predicted relative incidences, as presented in Fig. 3, highlight the impact of school closure compared to an increase in physical distancing at work by age. The relative incidence in people 18–60 years of age decreases with an increasing proportion of workplace distancing, which is of interest if this age group is more vulnerable compared to children. The relative incidence in the age group above 60 years of age increases in all situations compared to no intervention. This does not imply that the absolute number of cases in this age group would rise.

Limitations
Most survey designs were based on the POLYMOD survey though each survey had additional features and objectives which provide useful additional information. At the moment, we do not capture the full potential of each data set yet. Our case study elaborates on adapted school and work contacts and does not capture compensation behavior due to not being at school or work. This might be valid for a pandemic situation but not for regular (school) holidays. Social distancing due to (pandemic) scares are also not included yet.

The current application contains a local version of each data set, with some additional data reformatting. Our aim is to enable a direct link to Zenodo repositories. Note that some social contact surveys are available on Zenodo but not (yet) included in Socrates. E.g., the data from China [18] contains grouped contacts, which require different methodology. We omitted data from the UK [21], Zambia and South Africa [23] from our case study because only infants or adults were recruited.

Note that we will continue to develop this open-source tool [37] and thus the input/output/plots/scenarios might change in future editions.
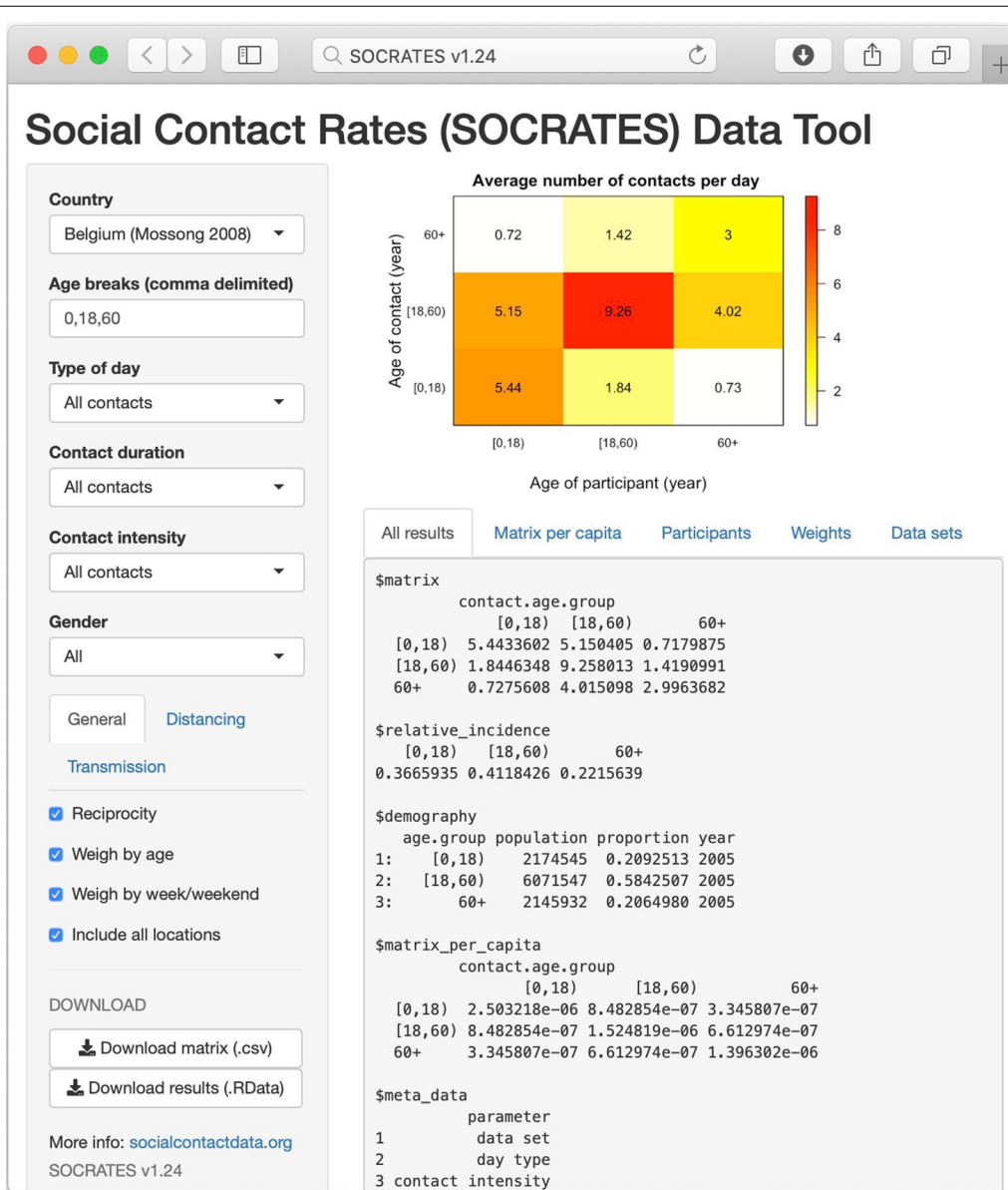
Fig. 1 Screenshot of the online SOCRATES application [36]. The user interface enables the selection of country data in combination with temporal and contact features. The social contact matrix is shown on the right-hand side in addition to principal results and statistics. When users incorporate physical distancing at work or school closure, using the "Distancing" panel on the left-hand side, the R_0 ratio is added to the output (not shown)
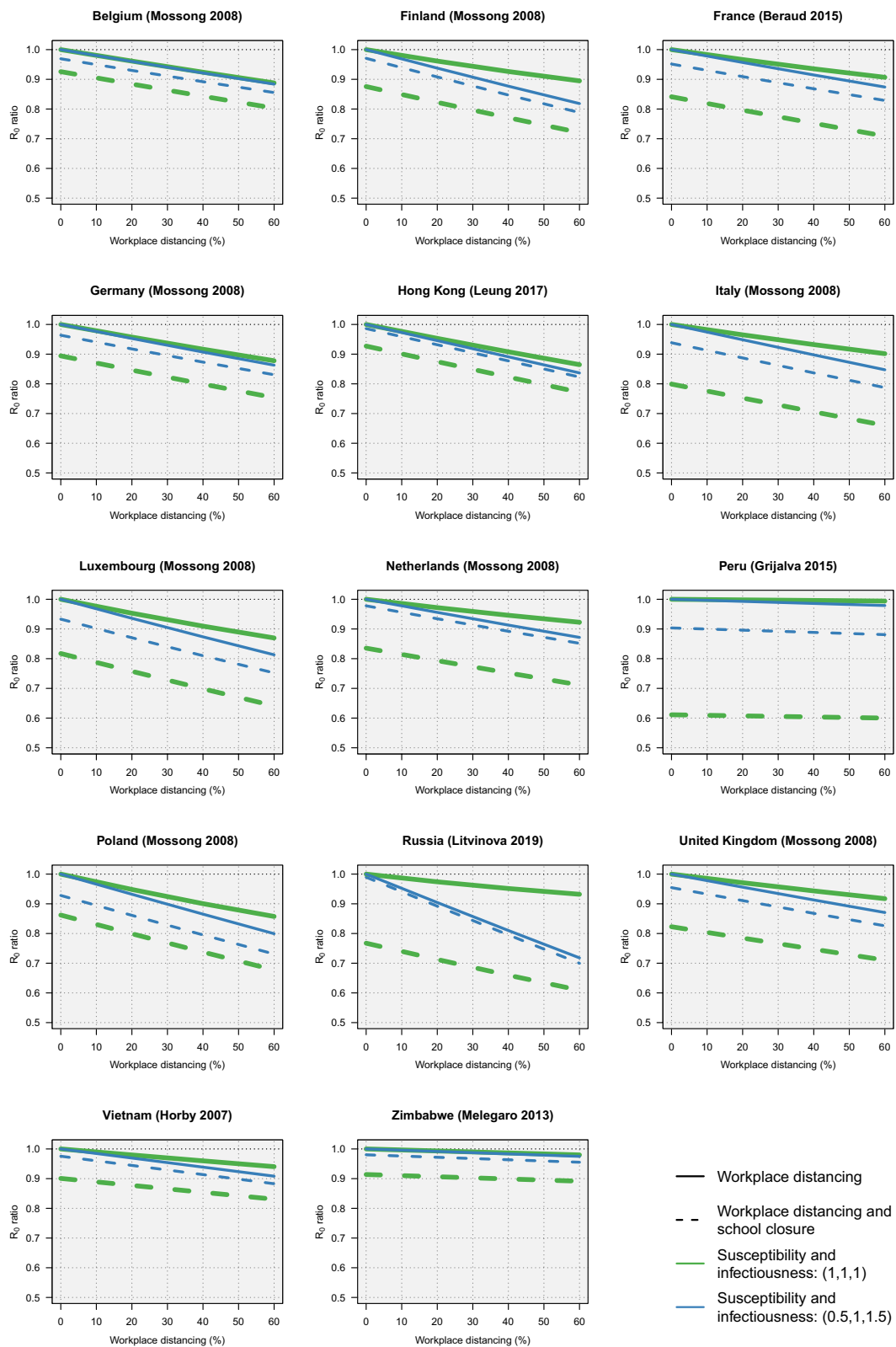
Fig. 2 Predicted R_0 ratio by country due to physical distancing at work and/or school closure. The impact on R_0 is shown with uniform susceptible and infectiousness parameters (1,1,1) and when children are less vulnerable compared to elderly (0.5,1,1.5). The age classes are 0–18 years, 19–60 years and over 60 years of age. Distancing is incorporated by a reduction of location-specific social contacts

Willem *et al. BMC Res Notes*   (2020) 13:293

Page 6 of 8
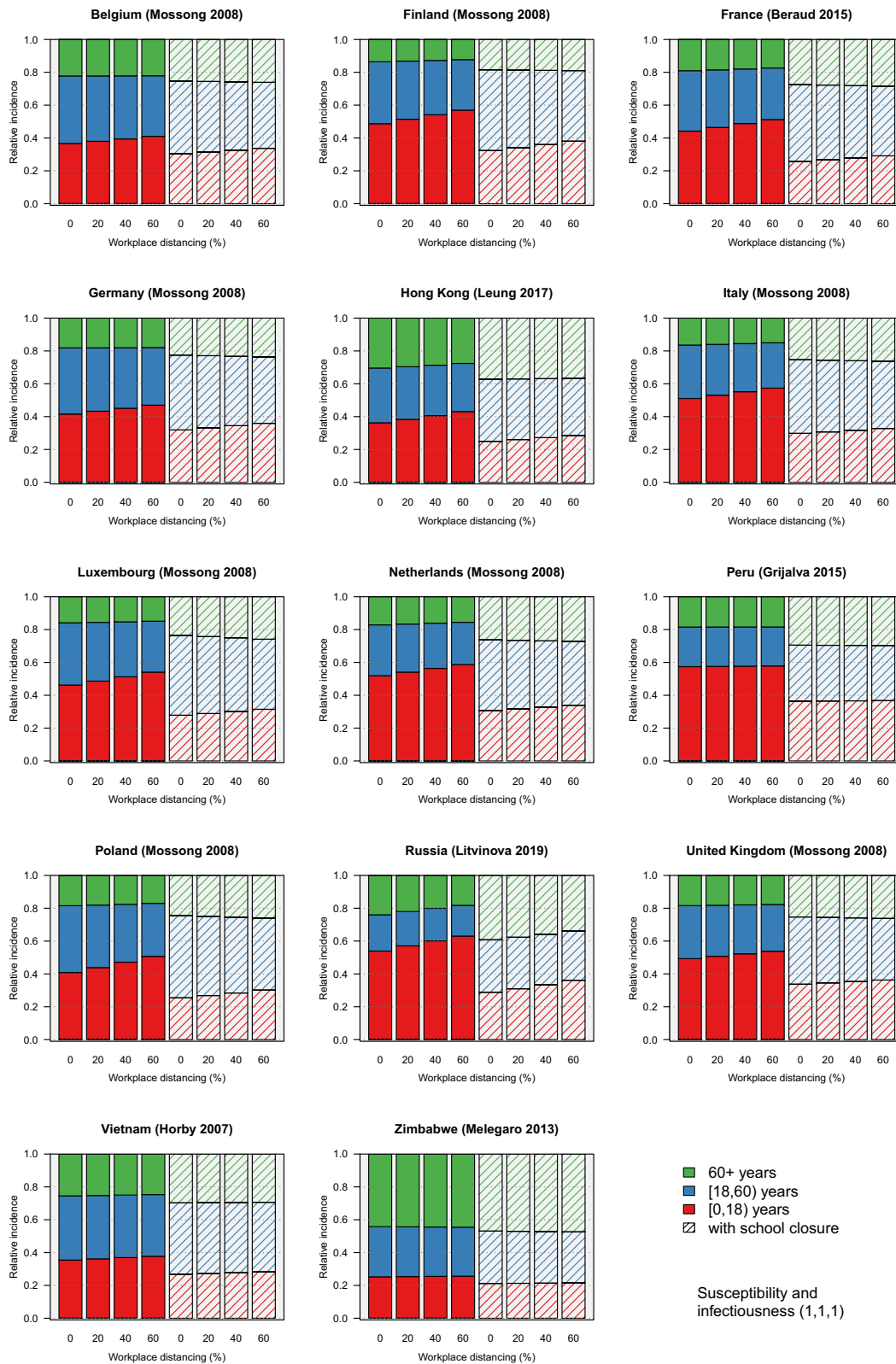
Fig. 3 Predicted age-specific relative incidence by country with physical distancing at work and/or school closure. The analysis presented here does not account for age-specific vulnerability. Distancing is incorporated by a reduction of location-specific social contacts

Willem *et al. BMC Res Notes*   (2020) 13:293

Page 7 of 8

Authors' contributions
NH conceived the study. TVH and PC collected and reformatted social contact data. LW and NH wrote a first draft of the paper. LW, TVH, SF and NH developed the online tool. PB substantively revised the manuscript. All authors contributed to the final version of the paper. All authors read and approved the final manuscript.

Availability of data and materials
All data sets are available on Zenodo [26–34]. We also share all R-code on Zenodo [37].

Ethics approval and consent to participate
The social contact data sharing initiative is part of the ERC consolidator grant "TransMID" which received ethical approval from the Hasselt University Medical Ethical Committee (CME2016/618).

Consent for publication
Not applicable.

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

Author details
[1] Centre for Health Economic Research and Modelling Infectious Diseases, University of Antwerp, Antwerp, Belgium. [2] Interuniversity Institute of Biostatistics and Statistical Bioinformatics, Data Science Institute, Hasselt University, Hasselt, Belgium. [3] Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK. [4] School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia.

References
1. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle P-Y, D'Ortenzio E, Yazdanpanah Y, Eholie SP, Altmann M, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. Lancet. 2020;395:871–7.
2. Verelst F, Kuylen E, Beutels P. Indications for healthcare surge capacity in European countries facing an exponential increase in coronavirus disease (COVID-19) cases, March 2020. Euro Surveill. 2020;25(13):2000323. https://doi.org/10.2807/1560-7917.ES.2020.25.13.2000323.
3. Lu H, Ai J, Shen Y, Li Y, Li T, Zhou X, Zhang H, Zhang Q, Ling Y, Wang S, et al. A descriptive study of the impact of diseases control and prevention on the epidemics dynamics and clinical features of SARS-CoV-2 outbreak in Shanghai, lessons learned for metropolis epidemics prevention. medRxiv 2020. https://doi.org/10.1101/2020.02.19.20025031.
4. Hoang TV, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, Beutels P, Hens N. A systematic review of social contact surveys to inform transmission models of close-contact infections. Epidemiology. 2019;30(5):723–36.
5. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3):74.
6. Willem L, Van Kerckhove K, Chao DL, Hens N, Beutels P. A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. PLoS ONE. 2012;7(11):e48695.
7. 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. 2006;164(10):936–44.
8. Ogunjimi B, Hens N, Goeyvaerts N, Aerts M, Van Damme P, Beutels P. Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. Math Biosci. 2009;218(2):80–7.
9. Goeyvaerts N, Hens N, Ogunjimi B, Aerts M, Shkedy Z, Van Damme P, Beutels P. Estimating infectious disease parameters from data on social contacts and serological status. J R Stat Soc Ser C Appl Stat. 2010;59(2):255–77.
10. Eames K, Tilston N, White P, Adams E, Edmunds W. The impact of illness and the impact of school closure on social contact patterns. Health Technol Assess. 2010;14(34):267–312.
11. Hens N, Ayele GM, Goeyvaerts N, Aerts M, Mossong J, Edmunds JW, Beutels P. Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. BMC Infect Dis. 2009;9(1):187.
12. Population Division, Department of Economic and Social Affairs, United Nation. Wpp2015: World population prospects 2015. R package version 1.1-0. 2016. https://CRAN.R-project.org/package=wpp2015.
13. Held L, Hens N, O'Neill PD, Wallinga J. Handbook of infectious disease data analysis. Boca Raton: Chapman and Hall/CRC; 2019.
14. Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J Math Biol. 1990;28(4):365–82.
15. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. Shiny: web application framework for R. R package version 1.2.0. 2018. https://CRAN.R-project.org/package=shiny.
16. Funk, S. Socialmixr: social mixing matrices for infectious disease modelling. R package version 0.1.6. 2020. https://CRAN.R-project.org/package=socialmixr.
17. Béraud G, Kazmercziak S, Beutels P, Levy-Bruhl D, Lenne X, Mielcarek N, Yazdanpanah Y, Boëlle P-Y, Hens N, Dervaux B. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. PLoS ONE. 2015;10(7):e0133203.
18. Zhang J, Klepac P, Read JM, Rosello A, Wang X, Lai S, Li M, Song Y, Wei Q, Jiang H, et al. patterns of human social contact and contact with animals in Shanghai, China. Sci Rep. 2019;9(1):1–11.
19. Leung K, Jit M, Lau EH, Wu JT. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. Sci Rep. 2017;7(1):1–12.
20. Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, Hens N, et al. A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. PLoS ONE. 2015;10(3):e0118457.
21. van Hoek AJ, Andrews N, Campbell H, Amirthalingam G, Edmunds WJ, Miller E. The social life of infants in the context of infectious disease transmission; social contacts and mixing patterns of the very young. PLoS ONE. 2013;8(10):e76180.
22. Litvinova M, Liu Q-H, Kulikov ES, Ajelli M. Reactive school closure weakens the network of social interactions and reduces the spread of influenza. Proc Natl Acad Sci USA. 2019;116(27):13174–81.
23. Melegaro A, Del Fava E, Poletti P, Merler S, Nyamukapa C, Williams J, Gregson S, Manfredi P. Social contact structures and time use patterns in the Manicaland Province of Zimbabwe. PLoS ONE. 2017;12(1):e0170459.
24. Horby P, Thai PQ, Hens N, Yen NTT, Mai LQ, Thoang DD, Linh NM, Huong NT, Alexander N, Edmunds WJ, et al. Social contact patterns in Vietnam and implications for the control of infectious diseases. PLoS ONE. 2011;6(2):e16965.
25. Dodd PJ, Looker C, Plumb ID, Bond V, Schaap A, Shanaube K, Muyoyeta M, Vynnycky E, Godfrey-Faussett P, Corbett EL, et al. Age-and sex-specific

Willem *et al. BMC Res Notes*   (2020) 13:293

Page 8 of 8

social contact patterns and incidence of *Mycobacterium tuberculosis* infection. Am J Epidemiol. 2016;183(2):156–66.

26. Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, Hens N, et al. Peruvian social contact data. Zenodo. https://doi.org/10.5281/zenodo.1215891.

27. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, et al. POLYMOD social contact data. Zenodo. https://doi.org/10.5281/zenodo.1215899.

28. Béraud G, Kazmerczak S, Beutels P, Levy-Bruhl D, Lenne X, Mielcarek N, Yazdanpanah Y, Boëlle P-Y, Hens N, Dervaux B. France social contact data. Zenodo. https://doi.org/10.5281/zenodo.1158452.

29. Ajelli M, Litvinova M. Russian contact matrices by age. Zenodo. https://doi.org/10.5281/zenodo.3232929.

30. Horby P, Thai PQ, Hens N, Yen NTT, Mai LQ, Thoang DD, Linh NM, Huong NT, Alexander N, Edmunds WJ, et al. Social Contact data for Vietnam. Zenodo. https://doi.org/10.5281/zenodo.1289474.

31. Leung K, Jit M, Lau EH, Wu JT. Social contact data for Hong Kong. Zenodo. https://doi.org/10.5281/zenodo.1165562.

32. Melegaro A, Del Fava E, Poletti P, Merler S, Nyamukapa C, Williams J, Gregson S, Manfredi P. Zimbabwe social contact data. Zenodo. https://doi.org/10.5281/zenodo.1251944.

33. Dodd PJ, Looker C, Plumb ID, Bond V, Schaap A, Shanaube K, Muyoyeta M, Vynnycky E, Godfrey-Faussett P, Corbett EL, et al. Social contact data for Zambia and South Africa (CODA Dataset). Zenodo. https://doi.org/10.5281/zenodo.2548693.

34. Zhang J, Klepac P, Read JM, Rosello A, Wang X, Lai S, Li M, Song Y, Wei Q, Jiang H, et al. Social contact data for China Mainland. Zenodo. https://doi.org/10.5281/zenodo.3516113.

35. van Hoek AJ, Andrews N, Campbell H, Amirthalingam G, Edmunds WJ, Miller E. Social contact data for UK. Zenodo. https://doi.org/10.5281/zenodo.1409507.

36. Social Contact Rates (SOCRATES) Data Tool: as part of the socialcontactdata.org initiative. TransMID. http://www.socialcontactdata.org.

37. Willem L, Hoang VT, Funk S, Coletti P, Beutels P, Hens N. Social contact rates (SOCRATES) data tool (v1.5). Zenodo. https://doi.org/10.5281/zenodo.3706788.

38. Guan W-J, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, Liu L, Shan H, Lei C-L, Hui DS, et al. Clinical characteristics of Coronavirus disease 2019 in China. N Engl J Med. 2020;382:1708–20.

Publisher's Note