COMMENT

Check for updates

Addressing the socioeconomic divide in computational modeling for infectious diseases

Michele Tizzoni [1✉], Elaine O. Nsoesie [2,3], Laetitia Gauvin[1],
Márton Karsai [4,5], Nicola Perra [6] & Shweta Bansal[7]

The COVID-19 pandemic has highlighted how structural social inequities fundamentally shape disease dynamics, yet these concepts are often at the margins of the computational modeling community. Building on recent research studies in the area of digital and computational epidemiology, we provide a set of practical and methodological recommendations to address socioeconomic vulnerabilities in epidemic models.

Socioeconomic factors in infectious disease modeling and surveillance: the need for a comprehensive approach

The investigation of social determinants of health and disease stands at the core of social epidemiology, a discipline whose tradition dates back to the mid-20th century[1]. Social conditions and in particular social disparities related to income, wealth, race, ethnicity, gender, and education, to cite only a few, are known to affect the health status of individuals and they reflect in unequal health outcomes when it comes to disease burden[2].

In recent decades, significant effort has been devoted to investigating the relationship between differences in socioeconomic conditions and the prevalence of non-communicable diseases[3]; however, the socioeconomic divide represents a key factor in the spread of infectious diseases as well. The effect of socioeconomic status (SES) on the spread of respiratory infections has been recognized in several past and recent epidemics, for instance in the case of the 1918 and 2009 flu pandemics lower SES was found to be associated with the highest disease burden[4–6]. Similarly, socioeconomic disparities, such as unequal access to care and sanitation, have been shown to be important in the West African Ebola outbreak[7] and in the spread of vector-borne diseases, such as malaria[8] and dengue[9,10].

The COVID-19 pandemic has revealed and further exacerbated such differences across several dimensions. During the pandemic, health outcomes have been significantly different by social strata, with inequalities in the distribution of infections, hospitalizations, and deaths closely matching income, occupational and racial disparities[11–13]. In the early phase of the pandemic, such inequalities were strongly linked to the affordability of non-pharmaceutical interventions (NPIs), as the feasibility of adopting prolonged social distancing measures has been a privilege of a few[14,15]. Following the rapid development of efficacious vaccines, inequities in health outcomes

[1] ISI Foundation, Turin, Italy. [2] Department of Global Health, School of Public Health, Boston University, Boston, MA, USA. [3] Center for Antiracist Research, Boston University, Boston, MA, USA. [4] Department of Network and Data Science, Central European University, 1100 Vienna, Austria. [5] Alfréd Rényi Institute of Mathematics, 1053 Budapest, Hungary. [6] School of Mathematical Sciences, Queen Mary University of London, London, UK. [7] Department of Biology, Georgetown University, Washington, DC, USA. ✉email: michele.tizzoni@isi.it

have been further driven by disparities in vaccine distribution, and accessibility, especially across the Global North-South Divide[16].

While particular disparities in health outcomes were not obvious at the start of the pandemic, and many others are yet to be discovered, the populations that have borne the greatest morbidity and mortality burden are the same populations that tend to have the highest burden of disease and limited access to optimal healthcare. These disparities are largely driven by structural factors and therefore require collaboration with social scientists, historians, and economists to understand the impact of past and current factors on the health outcomes of these populations, and design studies that focus on addressing underlying factors rather than unjustly blaming the individuals affected.

Despite the clear understanding of the importance of socioeconomic inequalities in disease transmission dynamics, the epidemic modeling community has often neglected these aspects in traditional mathematical approaches. One reason is the lack of an empirically driven mechanistic description of the interaction between inequalities and disease outcomes. Most mathematical models account for variations in health risk by age, and by occupational status, often limited to the dichotomy of student/ worker. Contact heterogeneity, which is known to be key in defining the risk of infection, is usually assumed to be encoded into the demographic structure of a population[17]; therefore, implying that countries or regions with similar demographics will experience similar epidemic trajectories. Despite their extensive and successful application in many real-world settings to infer key epidemiological parameters, evaluate different epidemic scenarios and inform public health interventions, many computational models remain agnostic about socioeconomic disparities, and they provide, by definition, only partial views of transmission mechanisms at play. However, as models are becoming more and more a standard tool for decision makers to inform public health policies, such as the adoption of NPIs for entire populations, this might in turn lead to widening social and health inequities. For example, models that assume that the risk of exposure and infection is the same for every individual in the population could lead to the implementation of interventions that are the same across a population irrespective of the social factors that create unequal exposure. Depending on the assumptions, such interventions will likely favor one group more than another.

In recent years, researchers have advocated for the extended use of computational and digital tools to tackle the emergence of novel infectious diseases and to rapidly face new outbreaks[18,19]. More recently, the importance of including social aspects in infectious disease modeling has been highlighted by numerous authors[20–22]. In this comment, we argue that the field of digital and computational epidemiology may remedy some of the challenges of socioeconomic inequalities in outbreak science. We provide some relevant examples of studies that demonstrate the opportunities of digital approaches to these issues and we conclude by making a set of practical recommendations to advance the field toward a more comprehensive approach.

Computational and digital epidemiology approaches to address the socioeconomic divide in the COVID-19 pandemic

Despite the challenges posed by socioeconomic inequalities to disease modeling and surveillance, the COVID-19 pandemic has also shown the promise of computational and digital epidemiology to address these gaps. Indeed, key insights into the effects of social inequalities during the COVID-19 crisis have come from the analysis of novel digital traces and their integration into epidemic models. Through the analysis of mobility patterns derived from de-identified mobile phone data, several studies

have revealed that individuals belonging to higher SES could better afford the adoption of health-protecting behaviors, such as reducing their mobility and social distancing[15,23,24]. As a consequence, disadvantaged groups that were not able to limit their social interactions experienced the highest rates of infections. Socioeconomic constraints to mobility reductions were associated with income levels, especially in the USA[15,25], but also more broadly with the structure of the labor market, as found in France[26], Italy[27], and Colombia[28]. In general, workers in informal sectors, or in essential services, such as agricultural workers, were incentivized or required to continue working away from home despite the restrictions, leading to higher infection risks.

By exposing the hard social constraints that limited the adherence to physical distancing in many countries, researchers have underscored the need for more equitable policies in response to COVID-19 and provided practical guidance to achieve them with the aid of computational models. For instance, by mapping the movements of about 100 million people to half-million points of interest in the US, and creating a network model of SARS-CoV-2 based on these data, Chang and collaborators identified a range of optimal reopening strategies to minimize the burden of infections among the most deprived populations[29]. A modeling study of the spread of SARS-CoV-2 in Santiago de Chile highlighted how the deep socioeconomic inequalities of the Santiago population, and the associated disparities in mobility reductions, significantly delayed the end of the first COVID-19 wave[30]. Counterfactual scenario simulations showed that a more equitable social distancing would have prevented more than 80% of deaths reported in Santiago, in the same period. Similarly, by combining detailed mobility data describing contact rates between households in the metropolitan area of Philadelphia, with a computational model of epidemic spread, Nande et al. demonstrated that evictions, which would inevitably follow mass unemployment due to the COVID-19 closures, lead to a significant increase in COVID-19 cases[31]. Such an increase would mainly affect denser and poorer neighborhoods of the city, widening the disparities in health risks associated with the pandemic. As a consequence, eviction moratoria could be an effective public health measure to avoid rapid surges in COVID-19 cases.

All these modeling efforts usually incorporated socioeconomic factors effectively, through the integration of behavioral data, such as mobility traces, to calibrate classic age- and spatially structured epidemic models. Fewer studies, instead, defined an epidemic model where socioeconomic disparities are encoded into its mathematical formulation. For instance, Ma et al.[32] developed a compartmental model with assortative mixing derived from the census distribution of ethnic groups in the city of New York to explain the high burden of disease in minority populations.

While both approaches either indirectly (the former) or directly (the latter) take into account socioeconomic disparities, only the latter offers the opportunity to actively investigate the mechanisms of infection inequality, and identify strategies to prevent them.

Building on these examples, in what follows, we make a set of specific recommendations to advance the field by bringing the concept of socioeconomic equity to all the three main aspects that stand at the core of the data-model loop in computational epidemiology: surveillance data, behavioral data, and epidemic models (see Fig. 1).

Equity in disease surveillance. Disparities in exposure, susceptibility, transmission, and treatment lead to certain populations bearing a higher disease burden than others. These disparities are driven by unequal access to resources that promote health at the
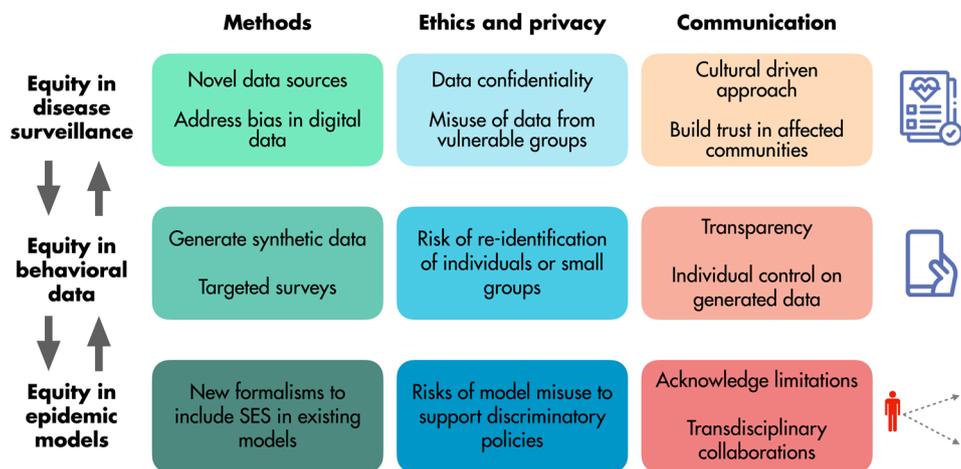
Fig. 1 Practical guidelines to introduce an equity lens into the computational modeling of infectious diseases and related challenges. The figure illustrates and summarizes key recommendations and associated challenges, grouped by area of relevance in the data-model framework of computational epidemiology: surveillance data, behavioral data, and epidemic models. SES socioeconomic status.

individual and community level (i.e., social determinants of health), discrimination against lower SES, displaced populations (e.g., refugees), and structural racism against people of color[33–35]. To address these disparities, we need reliable data on infection burden and access to public health resources for different socio-economic groups. Such data are critical to the prevention of infection and to characterize the disparate impacts of interventions across social and economic strata that can create or exacerbate health inequities. We, therefore, need to design disease surveillance systems that are focused on promoting equity.

Incorporating an equity lens into surveillance systems would differ depending on the data sources, and local context. Digital data collected from high-volume healthcare data, participatory surveillance systems, or from mining digital traces from social media, mobile phone usage, and Internet usage are opportunistic in nature meaning they are not usually generated for disease surveillance purposes. Such novel data streams can improve timeliness, spatial and temporal resolution, and provide access to unreachable populations for effective infectious disease surveillance, but such data must be used with caution and with explicit equity lens[36–39]. Digital healthcare records for passive syndromic surveillance, for example, are dependent on healthcare accessibility, healthcare-seeking behavior, and other reporting issues[40]. Studies in the US for influenza-like illness have demonstrated that disease burden can be underestimated in low SES populations from healthcare-based surveillance[41,42], which not only produces misleading estimates of disease burden but also underestimates the extent of health disparities. Until healthcare systems reach equity, statistical approaches must be developed to account for the measurement biases and quantify uncertainty in such data for infectious disease modeling applications[42].

Demographers and social scientists have been developing methods to account for bias due to nonrepresentative samples. For instance, in the early work by Zagheni and Weber[43], the authors accounted for variations in Internet penetration rates to correct their migration estimates based on e-mail data. Similar approaches could be extended to the field of digital epidemiology. In some cases, it is not always possible to address the bias inherent in digital data. Therefore, communicating these biases in a way that is easy for the audience (i.e., the public and policymakers) to understand is important.

Addressing limitations in disease surveillance systems requires a community and cultural context-driven approach to data collection. For example, knowing that Black or African American

people have a higher prevalence of the risk factors associated with severe COVID-19 disease should have prompted action and policy to collect consistent racial data at the beginning of the pandemic in the United States to identify and effectively respond to these disparities. Unfortunately, that was not the case. As of September 2020, only New York State was reporting race data for 100% of COVID-19 cases, while this percentage was zero for many other states[44]. These disparities are often driven by discrimination and bias that is well documented but might not always be discussed in the context of disease surveillance. Collaborating with public health scientists who study health disparities, historians and social scientists with expertise in the factors that have long influenced health outcomes, and community leaders is extremely important. However, the context and individuals represented will differ by country and region.

Furthermore, racial and ethnic discrimination has impacted the trust of healthcare providers within certain communities. It is therefore our responsibility as public health officers and researchers to earn the trust of these communities to enable the submission and collection of data for public health surveillance. A first step to developing equitable disease surveillance systems is designing policies that focus on addressing health disparities. A focus on health disparities will then require data that captures populations affected, disaggregated by social identity groups and social status. While aggregated data can demonstrate associations between disease burden and SES, disaggregated data is critical to accurately measuring health disparities. Second, data collection should also focus on capturing the interaction between different social processes that interact to create disadvantages. Individuals from low SES might have other social disadvantages depending on the context. In the US, for example, SES tends to intersect with certain racial or ethnic identities. While in other countries, low SES might be higher for vulnerable populations such as refugees. We acknowledge that SES interacts with different infectious diseases in varied ways; therefore, it is important to determine what SES factors are relevant for a particular infectious disease. For example, improving some SES factors might lead to improved access to quality housing, which might reduce the incidence of some respiratory conditions. However, improved housing is not associated with the reduced spread of all infectious diseases. Third, modelers and analysts should conduct studies to inform the detail and structure required in disease surveillance data for effective computational disease modeling studies focused on health disparities. And importantly, fourth, policies must be

created to enforce protection and deter misuse of data from minority, underserved and under-resourced populations.

Equity in behavioral data. The COVID-19 pandemic has shown that human behavioral patterns are an essential factor to consider for the better understanding, modeling, and forecasting of an ongoing epidemic. As equity in disease surveillance data is needed to identify disparities in disease exposure and transmission, equity in behavioral data is needed to address such disparities in statistical and computational epidemic models. We identify three main directions for action to integrate equity into behavioral data that are relevant to infectious disease dynamics.

First, socioeconomic differences in behavioral patterns could be captured by leveraging existing socioeconomic data sources, such as routinely collected census survey data.

Socioeconomic data underlying the observed disparities in health outcomes, such as income distributions, or population stratifications by racial and ethnic groups, are generally available in several countries, from official statistical sources, at very high spatial resolutions, and simple interaction models can be developed with them. As an example, the IPUMS International Historical Geographic Information System[45] represents a rich source to describe the distribution of essential or frontline workers—who are significantly more exposed than other groups[46]—across subpopulations. Similarly, the American Community Survey provides population numbers by race or ethnicity, at the census block level, that can be used to define varying rates of exposure or mixing patterns for different identity groups[32]. Despite their potential biases, these data sources provide usually the most accurate socioeconomic indicators available in several countries.

A second approach could be combining novel digital data sources and census data to generate synthetic data. Digital trace data can be collected in an aggregate way to preserve individual privacy and further calibrated to match the socioeconomic structure reported by census, as done by Replica to generate racial disaggregated mobility patterns[47]. The pandemic emergency has encouraged large tech companies to release digital trace data, in particular near real-time movement data: mobility reports provided by Google and Apple, mobility maps shared by Meta through its Data for Good Program, and mobility indicators made available by location intelligence companies, are all valuable inputs for epidemic models[48]. Future work should focus on developing the most statistically appropriate methods to combine such data streams to capture a refined picture of subpopulations. Efforts in this direction have started by building remotely sensed socioeconomic maps in several countries combining multiple data sources like mobility, satellite, night-light emission, or online social media, but still much research is needed[49–52]. Both traditional statistical approaches such as Iterative Proportional Fitting, Monte Carlo sampling, and machine learning techniques such as Self-Organized Maps or Generative Adversarial networks are appropriate candidates to generate synthetic behavioral data from novel data sources[53–55]. The synthetic and inferred data could then be used to represent mobility or contact patterns of subpopulations that will feed epidemic models, in a similar way to what is customarily done for age-dependent contact matrices[56,57].

Finally, targeted data collections should be devised to supplement behavioral data that are passively collected. Although digital data represent a powerful tool to measure behavioral changes during an epidemic, they commonly suffer from observational biases. They can provide insights only about people who have access to digital services, and thus overlook deprived socioeconomic groups. In some cases, active data collection might be the only effective approach to faithfully capture behaviors of underrepresented communities, as called for by the United Nations' Sustainable Development Goals (SDG 17.18). As an example, during the COVID-19 pandemic, UNICEF and the Harvard Humanitarian Initiative developed the Community Rapid Assessment to map protective behaviors across different groups in rural and urban areas through mobile phone surveys[58]. Traditional contact surveys should collect SES information, and new survey campaigns should be focused on low-resource settings[59]. Also, the deployment of proximity sensors[60] or GPS trackers[61] in urban and rural communities represents an alternative to measure contact patterns relevant to disease transmission in hard-to-reach, low-income settings.

Equity in epidemic models. Modeling frameworks that include SES at their core are largely missing and urgently needed[22]. When thinking about possible solutions, it is important to realize how one-fits-all approaches are hardly conceivable. In fact, the details of the implementation are likely to be a function of the data available as well as the type and scale of the model considered.

Arguably, the simplest concrete step in this direction would be extending standard compartmental structures to accommodate different SES as input. Similar to what is customarily done for age, compartments could be stratified by SES. Hence, models for the same disease, but developed for different countries, will reflect not only different age-pyramids but also different socioeconomic structures. Such an approach would allow accounting for differences in healthcare access and for heterogeneities in ability to respond to NPIs across SES. The next step could be capturing the stratification of contacts across both age and SES. Hence, we would move from classic contact matrices $M_{k,j}$ (describing the contact rate of individuals in age bracket k with those in age j) to $M_{k_\alpha j_\beta}$ (describing the interactions of individuals of age k and SES α with individuals of age j and SES β). Such models would allow capturing the correlations in contact patterns and their variations induced by NPIs as a function of socioeconomic indicators.

Some epidemic models are spatially aware and include the coupling between subpopulations (i.e., neighborhoods, regions, countries) through human mobility patterns. Such models could be extended to account for the interplay between SES and movements. To this end, the mobility matrices, describing the travel rates across subpopulations, could be stratified by SES, including the number of individuals of a given SES, traveling between location i and location j. Such extension would provide a more realistic description of human mobility and its variation induced by NPIs. Furthermore, it would allow us to estimate the impact of mitigation measures that target mobility reduction, considering different abilities to comply across SES.

As well-mixed compartmental models often neglect relevant population heterogeneities, other types of models may be needed to describe disease spread with sufficient detail. Agent-based models are the most detailed and complex modeling frameworks. They are based on generating synthetic populations that account for households, workplaces, and schools. Socioeconomic indicators could be used as additional households' features, together with their composition in terms of age and size. These models could be adopted to study the effects of school closures, household mixing, and remote working across different SES and to design interventions that account for inequalities to a higher level of realism.

Each of the proposed directions is of course far from trivial and underlies a clear increase in the complexity of models, some of which are already scratching the boundaries of what is computationally feasible. Furthermore, such extensions will give rise to and describe a wide range of mechanisms, dynamics, and interactions that do not have yet a solid theoretical basis. Just to

offer an example, adherence to NPIs is a complex phenomenon that has been linked to age, gender, education, political beliefs, country of residence, and SES. Hence, in absence of precise data, more expressive models like those we are advocating for would need extra layers of assumptions and parameters.

Furthermore, SES is an aggregated indicator encompassing a wide range of factors that can, directly or indirectly, affect disease spreading and outcomes. Overcrowding in households and workplaces, limited access to healthcare and vaccinations, and limited ability to comply with NPIs by reducing contacts and mobility patterns due to job security or type are just a few examples. Hence, another concrete step toward including inequity in epidemic models is to focus directly on such factors and study the differences they induce across SES as an emergent phenomenon rather than input. Differences in the number of contacts and interactions across groups due to overcrowding could be investigated via compartmental models that allow for differences in the effective transmissibility of a pathogen. The effects of job security and type could be investigated in spatially aware models linking them to variations in mobility patterns. The impact of overcrowding in workplaces could be investigated via agent-based models. Such an approach targets specific causes and mechanisms associated with SES, that might lead to inequalities in disease spreading. It would allow developing a better understanding of their impact on diseases on one side while offering a natural testbed to design specific interventions on the other. As such, it is complementary with respect to what we have described above where SES is considered as an input for the models.

Epidemic models that account for SES as input or that explicitly consider specific drivers of disease transmission associated with SES in their formulation would also allow us to formulate and address new questions. For example, they would enable us to study the impact of health disparities under different (controlled) conditions, disentangle the effects of multiple, and competing, drivers of transmission on health disparities, and design intervention strategies that consider the overall burden as well as health inequality[42,62–67].

Ethics and privacy challenges

We acknowledge that pursuing the above recommendations implies facing some relevant ethical challenges, which should be carefully considered by infectious disease modelers and researchers from multidisciplinary teams before they embark upon research on the topic.

Design of disease surveillance efforts should always aim at protecting the confidentiality of personal information under specific legal safeguards against the risks of disclosure. Individuals should be allowed to opt-out of public health surveillance activities if deemed at risk. The use and share of non-aggregated surveillance data should require the approval of trained research ethics committees[68]. Public health data collection must be conducted in a transparent and accountable manner, minimizing the risk that subjects of public health surveillance may face discrimination.

Similarly, the collection, analysis, and sharing of behavioral indicators from digital traces, such as mobile phone data, should adhere to the highest standards of anonymization, preventing data misuse that could potentially lead to re-identification of individuals or small groups. Differential privacy schemes, based for example on the addition of noise to the original data, should be applied to all data releases[69]. Furthermore, empowering individuals by providing them with more control over the data they generate, for instance through the creation of a Personal Data Store, could represent a solution to the inclusion of marginalized communities in behavioral data collection efforts[70].

Finally, epidemic models of infectious disease spread should be transparent in their assumptions, and the interpretation of their results should always minimize the risk of stigmatization of vulnerable communities. Modeling efforts should never be used to support the adoption or enforcement of discriminatory policies.

In conclusion, we acknowledge that, in some circumstances, benefits may not always outweigh the risks and all modeling studies that focus on minorities or marginalized communities should be preceded by systematic risks and harms assessment, following best practices such as the guidelines provided by UN Global Pulse[71].

Conclusion

The COVID-19 pandemic has demonstrated that socioeconomic inequalities cannot be ignored when it comes to understanding the distribution of disease burden, the behavioral responses to the epidemic, and the entire epidemic dynamics. An equity-focused approach to computational modeling of infectious diseases requires critically assessing how the data, modeling assumptions, and recommended policies impact individuals from different SES. In addition, communicating findings from modeling studies during public health emergencies remains a challenge. Especially, how to quantify and communicate uncertainty to policymakers and the general public. Addressing this challenge requires collaboration with communication experts and community leaders to develop culturally appropriate messages. In addition, the humility to acknowledge limitations in modeling and the compassion to understand the social and political processes that drive individual decision making can go a long way in impacting responsiveness to information communicated during public health emergencies.

In this comment, we have made recommendations on how we can develop approaches to improve the collection and use of surveillance and behavioral data, and how we could incorporate socioeconomic information into epidemic models. While not comprehensive, we hope these recommendations will lead to constructive conversations around the need for digital and computational approaches that are inclusive and focused on reducing rather than exacerbating health disparities during public health emergencies.

References

1. Berkman, L. F., Kawachi, I. & Maria Glymour, M. *Social Epidemiology* (Oxford University Press, 2014).
2. Marmot, M., Friel, S., Bell, R., Houweling, T. A. J. & Taylor, S. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* **372**, 1661–1669 (2008).
3. Mendenhall, E., Kohrt, B. A., Norris, S. A., Ndetei, D. & Prabhakaran, D. Non-communicable disease syndemics: poverty, depression, and diabetes among low-income populations. *Lancet* **389**, 951–963 (2017).
4. Mamelund, S.-E., Shelley-Egan, C. & Rogeberg, O. The association between socioeconomic status and pandemic influenza: systematic review and meta-analysis. *PLoS ONE* **16**, e0244346 (2021).
5. Grantz, K. H. et al. Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918. *Proc. Natl Acad. Sci. USA* **113**, 13839–13844 (2016).
6. Mamelund, S.-E. Social inequality—a forgotten factor in pandemic influenza preparedness. *Tidsskr. Nor. Legeforen.* **137**, 911–913 (2017).
7. Alexander, K. A. et al. What factors might have led to the emergence of ebola in West Africa? *PLoS Negl. Trop. Dis.* **9**, e0003652 (2015).
8. Degarege, A., Fennie, K., Degarege, D., Chennupati, S. & Madhivanan, P. Improving socioeconomic status may reduce the burden of malaria in sub Saharan Africa: a systematic review and meta-analysis. *PLoS ONE* **14**, e0211205 (2019).

9. Kikuti, M. et al. Spatial distribution of dengue in a Brazilian urban slum setting: role of socioeconomic gradient in disease risk. *PLoS Neglected Tropical Dis.* **9**, e0003937 (2015).

10. Rodrigues, N. C. P. et al. Temporal and spatial evolution of dengue incidence in Brazil, 2001-2012. *PLoS ONE* **11**, e0165945 (2016).

11. Bambra, C., Riordan, R., Ford, J. & Matthews, F. The COVID-19 pandemic and health inequalities. *J. Epidemiol. Community Health* **74**, 964–968 (2020).

12. Paul, A., Englert, P. & Varga, M. Socio-economic disparities and COVID-19 in the USA. *J. Phys.: Complex.* **2**, 035017 (2021).

13. The COVID Tracking Project. The COVID Racial Data Tracker. https://covidtracking.com/race/ (2020).

14. Mena, G. E. et al. Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science* **372**, eabg5298 (2021).

15. Jay, J. et al. Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. *Nat. Hum. Behav.* **4**, 1294–1302 (2020).

16. Ye, Y. et al. Equitable access to COVID-19 vaccines makes a life-saving difference to all countries. *Nat. Hum. Behav.* **6**, 207–216 (2022).

17. 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).

18. Rivers, C. et al. Using 'outbreak science' to strengthen the use of models during epidemics. *Nat. Commun.* **10**, 1–3 (2019).

19. Lofgren, E. T. et al. Mathematical models: a key tool for outbreak response. *Proc. Natl Acad. Sci. USA* **111**, 18095–18096 (2014).

20. Bedson, J. et al. A review and agenda for integrated disease models including social and behavioural factors. *Nat. Hum. Behav.* **5**, 834–846 (2021).

21. Buckee, C., Noor, A. & Sattenspiel, L. Thinking clearly about social aspects of infectious disease transmission. *Nature* **595**, 205–213 (2021).

22. Zelner, J. et al. There are no equal opportunity infectors: epidemiological modelers must rethink our approach to inequality in infection risk. *PLoS Comput. Biol.* **18**, e1009795 (2022).

23. Lee, W. D., Qian, M. & Schwanen, T. The association between socioeconomic status and mobility reductions in the early stage of England's COVID-19 epidemic. *Health Place* **69**, 102563 (2021).

24. Garnier, R., Benetka, J. R., Kraemer, J. & Bansal, S. Socioeconomic disparities in social distancing during the COVID-19 pandemic in the United States: observational study. *J. Med. Internet Res.* **23**, e24591 (2021).

25. Weill, J. A., Stigler, M., Deschenes, O. & Springborn, M. R. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proc. Natl Acad. Sci. USA* **117**, 19658–19660 (2020).

26. Valdano, E., Lee, J., Bansal, S., Rubrichi, S. & Colizza, V. Highlighting socio-economic constraints on mobility reductions during COVID-19 restrictions in France can inform effective and equitable pandemic response. *J. Travel Med.* **28**, taab045 (2021).

27. Gauvin, L. et al. Socio-economic determinants of mobility responses during the first wave of COVID-19 in Italy: from provinces to neighbourhoods. *J. R. Soc. Interface* **18**, 20210092 (2021).

28. Heroy, S., Loaiza, I., Pentland, A. & O'Clery, N. COVID-19 policy analysis: labour structure dictates lockdown mobility behaviour. *J. R. Soc. Interface* **18**, 20201035 (2021).

29. Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).

30. Gozzi, N. et al. Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nat. Commun.* **12**, 2429 (2021).

31. Nande, A. et al. The effect of eviction moratoria on the transmission of SARS-CoV-2. *Nat. Commun.* **12**, 2274 (2021).

32. Ma, K. C., Menkir, T. F., Kissler, S., Grad, Y. H. & Lipsitch, M. Modeling the impact of racial and ethnic disparities on COVID-19 epidemic dynamics. *Elife* **10**, e66601 (2021).

33. Tan, S. B., deSouza, P. & Raifman, M. Structural racism and COVID-19 in the USA: a county-level empirical analysis. *J. Racial Ethn. Health Disparities* **9**, 236–246 (2022).

34. Burström, B. & Tao, W. Social determinants of health and inequalities in COVID-19. *Eur. J. Public Health* **30**, 617–618 (2020).

35. Ataguba, O. A. & Ataguba, J. E. Social determinants of health: the role of effective communication in the COVID-19 pandemic in developing countries. *Glob. Health Action* **13**, 1788263 (2020).

36. Nsoesie, E. O., Oladeji, O. & Sengeh, M. D. Digital platforms and non-communicable diseases in sub-Saharan Africa. *Lancet Digit Health* **2**, e158–e159 (2020).

37. Althouse, B. M. et al. Enhancing disease surveillance with novel data streams: challenges and opportunities. *EPJ Data Sci.* **4**, 17 (2015).

38. Henly, S. et al. Disparities in digital reporting of illness: a demographic and socioeconomic assessment. *Prev. Med.* **101**, 18–22 (2017).

39. Bansal, S., Chowell, G., Simonsen, L., Vespignani, A. & Viboud, C. Big data for infectious disease surveillance and modeling. *J. Infect. Dis.* **214**, S375–S379 (2016).

40. Lee, E. C. et al. Deploying digital health data to optimize influenza surveillance at national and local scales. *PLoS Comput. Biol.* **14**, e1006020 (2018).

41. Scarpino, S. V. et al. Socioeconomic bias in influenza surveillance. *PLoS Computational Biol.* **16**, e1007941 (2020).

42. Zipfel, C. M., Colizza, V. & Bansal, S. Health inequities in influenza transmission and surveillance. *PLoS Computational Biol.* **17**, e1008642 (2021).

43. Zagheni, E. & Weber, I. Demographic research with non-representative internet data. *Int. J. Manpow.* **36**, 13–25 (2015).

44. Krieger, N., Testa, C., Hanage, W. P. & Chen, J. T. US racial and ethnic data for COVID-19 cases: still missing in action. *Lancet* **396**, e81 (2020).

45. Manson, S. M., Kugler, T. A., Schroeder, J., Van Riper, D. & Ruggles, S. IPUMS International Historical Geographic Information System: Version 1 (2020).

46. Schnake-Mahl, A. S., Lazo, M., Dureja, K., Ehtesham, N. & Bilal, U. Racial and ethnic inequities in occupational exposure across and between US cities. *SSM - Popul. Health* **16**, 100959 (2021).

47. Disaggregated data is essential to transportation equity. *REPLICA HQ.* https://replicahq.com/disaggregated-data-is-essential-to-transportation-equity/ (2021).

48. Oliver, N. et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci. Adv.* **6**, eabc0764 (2020).

49. Jean, N. et al. Combining satellite imagery and machine learning to predict poverty. *Science* **353**, 790–794 (2016).

50. Dong, Y., Yang, Y., Tang, J., Yang, Y. & Chawla, N. V. Inferring user demographics and social strategies in mobile social networks. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* https://doi.org/10.1145/2623330.2623703 (2014).

51. Chi, G., Fang, H., Chatterjee, S. & Blumenstock, J. E. Microestimates of wealth for all low- and middle-income countries. *Proc. Natl Aca. Sci. USA* **119**, e2113658119 (2022).

52. Abitbol, J. L. & Karsai, M. Interpretable socioeconomic status inference from aerial imagery through urban patterns. *Nat. Mach. Intell.* **2**, 684–692 (2020).

53. Felsenstein, D., Samuels, P. & Grinberger, Y. AASDC: An allocation algorithm for data disaggregation and synthetic database construction, WP 20/16. *The Development of a Dynamic Integrated Model for Disaster Management and Socio-Economic Analysis (DIM2SEA) JAPAN Science and Technology Agency (JST) and Ministry of Science, Technology and Space, Israel (MOST), Jerusalem.* (2016).

54. Huang, Z. & Williamson, P. *A Comparison of Synthetic Reconstruction and Combinatorial Optimisation Approaches to the Creation of Small-area Microdata.* (Department of Geography, University of Liverpool, 2001).

55. Crespo, R., Alvarez, C., Hernandez, I. & García, C. A spatially explicit analysis of chronic diseases in small areas: a case study of diabetes in Santiago, Chile. *Int. J. Health Geogr.* **19**, 24 (2020).

56. Mistry, D. et al. Inferring high-resolution human mixing patterns for disease modeling. *Nat. Commun.* **12**, 323 (2021).

57. Koltai, J., Vásárhelyi, O., Röst, G. & Karsai, M. Reconstructing social mixing patterns via weighted contact matrices from online and representative surveys. *Sci. Rep.* **12**, 4690 (2022).

58. Andersen, C., Huynh, U. K., Toasa, A. O., Wells, C. & Wong, M. Lessons from applying the community rapid assessment method to COVID-19 protective measures in three countries. *CHANCE* **34**, 6–12 (2021).

59. 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).

60. Kiti, M. C. et al. Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Sci.* **5**, 21 (2016).

61. Kauffman, K. et al. Comparing transmission potential networks based on social network surveys, close contacts and environmental overlap in rural Madagascar. *J. R. Soc. Interface* **19**, 20210690 (2022).

62. Speybroeck, N., Van Malderen, C., Harper, S., Müller, B. & Devleesschauwer, B. Simulation models for socioeconomic inequalities in health: a systematic review. *Int. J. Environ. Res. Public Health* **10**, 5750–5780 (2013).

63. Munday, J. D., van Hoek, A. J., John Edmunds, W. & Atkins, K. E. Quantifying the impact of social groups and vaccination on inequalities in infectious diseases using a mathematical model. *BMC Med.* **16**, 162 (2018).

64. Kumar, S., Piper, K., Galloway, D. D., Hadler, J. L. & Grefenstette, J. J. Is population structure sufficient to generate area-level inequalities in influenza rates? An examination using agent-based models. *BMC Public Health* **15**, 947 (2015).

65. Galanis, G. & Hanieh, A. Incorporating social determinants of health into modelling of COVID-19 and other infectious diseases: a baseline socio-economic compartmental model. *Soc. Sci. Med.* **274**, 113794 (2021).

66. Hyder, A. & Leung, B. Social deprivation and burden of influenza: testing hypotheses and gaining insights from a simulation model for the spread of influenza. *Epidemics* **11**, 71–79 (2015).

67. Quinn, S. C. & Kumar, S. Health inequalities and infectious disease epidemics: a challenge for global health security. *Biosecur. Bioterror.* **12**, 263–273 (2014).

68. World Health Organization & Others. Guidance for managing ethical issues in infectious disease outbreaks (2016).

69. de Montjoye, Y.-A., Shmueli, E., Wang, S. S. & Pentland, A. S. openPDS: protecting the privacy of metadata through SafeAnswers. *PLoS ONE* **9**, e98790 (2014).

70. Nanni, M. et al. Give more data, awareness and control to individual citizens, and they will help COVID-19 containment. *Ethics Inf. Technol.* **23**, 1–6 (2021).

71. UN Global Pulse. Risks, harms and benefits assessments, Level 2. https://www.unglobalpulse.org/policy/risk-assessment/ (2020).

Acknowledgements

Author contributions

M.T. conceived and drafted the first version of the manuscript. E.O.N., L.G., M.K., N.P., and S.B. contributed formative ideas, recommendations, and assisted with the drafting and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Michele Tizzoni.

Peer review information *Nature Communications* thanks Renato Casagrandi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.