Why We Need More Data before the Next Pandemic

Nigel Gilbert[*] [a]   Edmund Chattoe-Brown [b]   Christopher Watts [c]

Duncan Robertson [d]

[a]   Centre for Research in Social Simulation (CRESS), University of Surrey (United Kingdom)
[b]   School of Media, Communication and Sociology, University of Leicester (United Kingdom)
[c]   Independent researcher, Cambridgeshire (United Kingdom)
[d]   School of Business and Economics, Loughborough University (United Kingdom)

Abstract

Attempts to control the current pandemic through public health interventions have been driven by predictions based on modelling, thus bringing epidemiological models to the forefront of policy and public interest. It is almost inevitable that there will be further pandemics and controlling, suppressing and ameliorating their effects will undoubtedly involve the use of models. However, the accuracy and usefulness of models are highly dependent on the data that are used to calibrate and validate them. In this article, we consider the data needed by the two main types of epidemiological modelling (compartmental and agent-based) and the adequacy of the currently available data sources. We conclude that at present the data for epidemiological modelling of pandemics is seriously deficient and we make suggestions about how it would need to be improved. Finally, we argue that it is important to initiate efforts to collect appropriate data for modelling now, rather than waiting for the next pandemic.

Keywords: Covid-19; time use data; epidemiological model; social network; social contacts.

Acknowledgements

---

*   ✉ n.gilbert@surrey.ac.uk

1 Introduction

The Covid-19 pandemic has highlighted the limitations of existing data on social behaviour for modelling infection transmission in a population. Without the right kinds of data, policymakers have had to rely on models that assume unchanging behaviour in a population or a subset of the population. While these approaches may be satisfactory for the initial growth phase of an epidemic, limitations in understanding evolving behavioural heterogeneity are likely to result in forecasts that are materially wrong. This is part of a wider problem in sociology (and the other social sciences) which is that theories, conceptual models and frameworks shape the kinds of data that are collected and deemed to be important.

To make better models and for these to be calibrated and validated against the actual spread of disease, we will require better data. In this context, data is better in at least two senses. Firstly, we may just need more detailed data to reflect the changes in behaviour resulting from public health interventions (hereafter PHIs). Secondly, we may need different *kinds* of data, for example about how people decide if they can or will comply. The contagion of an infectious disease is fundamentally mediated by individuals interacting socially, and therefore appropriate and reliable social data is needed to generate models that accurately reflect the aggregate dynamics of individuals. While demographic and contact data about individuals have been collected in the past, the Covid-19 epidemic has highlighted the need to collect behavioural and attitudinal data that are both more fine-grained and more dynamic in distinctive ways (Biggs & Littlejohn, 2021). Such data can be used to produce more accurate models and to suggest answers to important questions such as why there have been a higher proportion of fatalities among ethnic minorities and in deprived areas (Khunti et al., 2020), how many sites of infection are connected by key workers (Daly, 2020), what happens to contact rates if schools, restaurants, shops or sporting events are closed (Santamaria, 2020), and why China, South Korea, Taiwan had better outcomes than the UK and the US in the first wave of the pandemic (e.g. Lee et al., 2020).

We first consider what data is required by the two main epidemiological modelling techniques: compartmental and agent-based modelling, then review current data sources for individual-level human interactions, and finally call for more appropriate and better quality data while considering some of the practical and ethical issues that may arise from its collection. While we are not the first to raise these modelling concerns — see for example, Squazzoni et al., 2020; Manzo & Van de Rijt, 2020 — our emphasis is on data requirements.

However, our aim is not only to contribute to modelling and epidemiology. By critically evaluating the theories implicit in these models and developing arguments for new forms of data, we also highlight the value of this data for meeting other challenges in sociological theory. For example, the integration of social networks, time use and spatial behaviour is also required in understanding large scale shifts of public sentiment. Epidemic models are thus a useful microcosm for the wider challenge of developing sociological theory by integrating it with appropriate novel kinds of data.

2 Challenges for SEIR Compartmental Models

The most common and, so far, the most influential style of epidemiological modelling uses a system of differential equations to represent the flow of a population through a set of compartments, each holding a stock of individuals in a distinct disease state (for example, Susceptible, Exposed, Infectious, and Removed states). Hence these models are often called SEIR models.

Fixed transition rates control the flow of individuals through these compartments (Anderson & May, 1992). A typical example is the SEIR model of Davies et al. (2020b) which was used to support advice to the UK Government on possible interventions to control Covid-19.

A concept useful for considering the data needs of SEIR models is that of the Data Generation Process (hereafter DGP) proposed by Hendry & Richard (1983, pp. 111–112). The DGP is the actual set of social interactions and events by which a real system evolves, as distinct from measurements of that system and theories about how it operates. In the context of a disease, for example, the DGP involves individuals being in various disease states (e.g., susceptible; infectious and showing symptoms; infectious and asymptomatic) and the various mechanisms of transition between those states (Davies et al., 2020b, Figure 1, p. e376). Based on this conceptual approach, we can rapidly identify data issues relevant to Covid-19 policy. Is there a measurement (or set of measurements) that can put individuals into disease categories with adequate reliability? Is that set of measures being applied to an adequate (and unbiased) sample of the populations in each compartment so that its values can potentially be generalised to the epidemic as a whole? How often is that measurement being carried out to assess changes in the system (including the effects of policies being implemented)?

From the perspective of effective modelling, it is clearly problematic if only those who have symptoms are tested or only those who are sick enough to be hospitalised are diagnosed to assess prevalence (Burgess et al., 2020). This issue has a knock-on effect for transitions between states in two different ways. Firstly, it is rarely feasible to do what is required to measure transition rates directly (i.e., take an adequate sample known reliably to be in a particular state and then track how many of them become, for example, seriously ill). Furthermore, there is a danger that one would have to carry out this process not once but repeatedly to get a sense of the extent to which it might depend on other factors like age.[1] Secondly, in practice, transition rates are often estimated from the overall progress of the disease itself (Davies et al., 2020a) and, clearly, if the progress of the disease is mismeasured (or the model fitted is mis-specified), then the transition probabilities will be mismeasured too. Unfortunately, a lack of measures or measure unreliability are not the only forces operating to produce incorrect estimates of transition probabilities between states. The point of PHIs during a pandemic is specifically to *change* at least some of these transition probabilities over time (but to an unknown extent *ex ante*). The extent to which interventions do change transition rates (for example between Susceptible and Exposed via mask wearing) is one important area where data is currently lacking.

It is this risk of various kinds of mismeasurement and mis-specification that requires us to concentrate much more carefully on the detail of the DGP so we can see how it leads to specific data demands. Some transitions, as far as we know, are predominantly determined by the characteristics of the disease. If an individual is infected, the disease will progress (although whether symptoms develop or the disease becomes serious are also associated with attributes such as age). Nevertheless, the outcomes for those with Covid-19 serious enough for hospitalisation are now significantly better than at the beginning of the pandemic (see, for example, Jorge et al., 2021). Thus, over time, this particular transition probability (from serious illness to death) has been falling. It is presumably both the evolution of the virus and the actions of the medical profession that have changed that transition probability. But for other transitions (most notably from susceptible to infected), it is the actions of the individual, society and/or

---

1. This also creates a practical problem for the compartmental approach. The more detailed the model becomes, the smaller the amount of data in each compartment and the more possible transitions there may be between compartments. Both aspects mean that transition rates cannot be estimated so robustly. Other modelling approaches like Agent-Based Modelling do not represent detail in this "atomising" way.

the government that have significant impacts on these transition probabilities.

In the early stages of the pandemic, individuals may have responded with little more than increased handwashing or being more careful how they coughed. But as soon as governments started telling people how to behave, those behaviours significantly increased from their baseline, although there was not complete compliance. This makes it clear that regular and comparable measurements of compliance are crucial (see for example, Hills & Eraso, 2021, but note the use of a convenience sample.) There is also a question about whether "compliance fatigue" — either based on psychology or material circumstances like running out of savings — needs to be factored into the relative effectiveness of intervention policies (see, for example, Hoeben et al., 2021).

Thus modelling approaches based on assuming fixed transition parameters, although possibly useful for the physiological aspects of a disease, are likely to be less than adequate for social behaviours subject to agency and policy (Manzo, 2020; Manzo & Van de Rijt, 2020). Moreover, attempting to *estimate* those transition parameters may be problematic. This is further complicated by the likely existence of feedback: for example, when estimated transition probabilities are used in models that then inform policies imposing interventions that change those same transition probabilities.

It follows that models cannot depend on the presumption that parameters remain constant but must be fed by frequently updated and representative data to recalibrate them. After the government said that people should not do x, what did the reported compliance rate actually change to? Are compliance rates dropping generally? Further, even among those who comply, what actual effect has compliance had on relevant contacts? (For some people with very limited lifestyles such as isolated elderly people, these may be almost unchanged. For others, the shift may be dramatic.) This requires that the data collected on compliance and contacts should be generic (e.g., how many times did you go shopping last week?) and not tied to particular policies or modelling approaches.

We should also recognise that measurements such as the results of tests for new diseases (or new forms of existing diseases) will not be available immediately and the best that may be possible is for models to be calibrated using previous data from the most relevant pandemics (which itself is an argument for ongoing consistent surveillance data systems — see Chua et al., 2021). The modelling strategy adopted will thus have to be good at assessing the various forms of uncertainty that may become clearer as the situation develops. The idea of modelling contingent on new data is not common in existing research (but see Birrell et al., 2021 for an example of "nowcasting").

Another implication of this analysis is the explicit role that agency plays in the effectiveness of policies. As soon as governments have to decide what to mandate, and the public have to decide whether to comply, it no longer seems credible to model an epidemic as people simply being "pushed" through disease states — as physiology alone does. We need to know how people reason about different forms of compliance and whether there are differences in the kinds of things they reason about (SoleimanvandiAzar et al., 2021). For example, reasoning based on misinformation may be countered (partially at least) by communication campaigns while basic needs (such as an adequate household income) may require changes to welfare policy. No amount of stressing the individual and social harms of Covid will allow a household to function without an income. More generally, the effect of drives such as sex and intoxication (whether drugs or alcohol) on undermining compliance seem to have received limited attention. This issue taps into the more general question about whether statistical approaches and probabilities can effectively represent decision processes and whether qualitative research on compliance

and reasoning about it will therefore have an important contribution to make (Denford et al., 2020; Santelli et al., 2020).

A further aspect of the data issue is the extent to which specific theories or modelling approaches shape the kind of data that is typically collected or used. For example, it is clear that epidemics have a spatial dimension and for this reason geographers and transport researchers collect data on "space" and its manifestations (such as travel). But it is also clear that epidemics have a network dimension — as with inviting friends to your house or asking kin in another household to provide childcare for schooling at home (Azad & Devi, 2020). The challenge is to develop models that can handle both dimensions and their interaction in appropriate ways. For example, there are many jobs which involve travel only to a static workplace, but there are also jobs that involve moving from site to site without much interaction with others (e.g., maintenance workers and delivery drivers) and yet others where the incumbent remains in one place but meets with a large number of "clients" (e.g., shop assistants and hospital workers).

To understand possible routes of transmission we need to have a much better overall sense of the composition of society and the resulting nuanced structure of its contact dynamics than we currently do. This detail cannot simply be represented as either a coherent network or a map since it involves timing, decisions and human goals at the very least. Owing to social complexity, it should not be assumed that small population segments, such as reconstituted families and boarding school pupils, do not have a significant impact on the dynamics of contact. Effective data of this kind will have to focus much less on averages and much more on distinctive modes of behaviour and categories of people. To take an example from Davies et al.'s (2020b) model, which assumes that the epidemic in each county evolves independently: how many kinds of people regularly reside in two or more counties and why? Some are obvious, like reconstituted families. Others, like the armed forces and construction workers on large projects, are much less so. Reading the current epidemiological modelling literature, even categories like households feel generic by sociological standards.

Another interesting example of this issue is provided by the current state of time diary research. Time diaries are generally argued to be a much more effective way to record how long people spend on activities such as work, sleep and childcare than using conventional surveys because they are supposed to be completed without relying on potentially inaccurate retrospective recall (Paolisso & Hames, 2010; Gmel & Daeppen, 2007). However, because they are considered onerous, they are rarely completed for more than a few days, which creates problems for identifying distinct behaviour on weekends, during school holidays and so on, and it is only relatively recently that a few of them have started to include data about where people are physically located when carrying out these activities and (broadly) who they carry them out with (e.g., Elgethun, 2007; Mullan & Chatzitheochari, 2019). It is exactly over these longer periods, invisible in current data, that modes of life like "term time" and "vacation" (which are popularly associated with particular modes of epidemic spread — see, for example, Brooks-Pollock et al., 2021) can be characterised.

For epidemiological purposes, it may not be enough to know how many people are "at work" between 0900 and 1000 on a Tuesday morning, because in addition one needs to know how many of them are likely to be physically co-located.[2] Similarly, it is not enough to know

---

2. The number, spatial distribution and size of sites such as schools, workplaces and cinemas is also an aspect of social life that does not seem to fit with existing social science. What implications does it have if everyone goes to the nearest small shop for food rather than a large supermarket with a much larger catchment area? What effect does the extent to which school catchment areas overlap have on infection? To answer such questions, we must first have a conceptual framework that allows us to see that they need asking.

that people are "with friends" as a broad category; one needs to know which friends and how often. In other areas, social science has been protected from being highly controversial by a tendency to look at aggregates and averages that allow individuals to remain anonymous. But in the limit, epidemiological models may require us to know exactly who was where and when, which will raise considerable issues of ethics and trust.

To sum up, there are several areas where new data may be needed or existing data may have to be used differently in order to maximise the benefits obtainable from compartmental models of epidemics and models of social behaviour in general:

- The possible role of historical data or new forms of routine data collection (such as that for dynamic contact structures) in providing estimates for models of emerging diseases will have to be systematically assessed. It is inevitable that such models will not have all the data they need in the early stages. What is the best performance that can be achieved in full awareness of that fact?

- Conceptual development has to proceed hand in hand with the identification of data requirements to pinpoint the aspects of social behaviour that compartmental models, network and geographical approaches tend to downplay or reinvent to fit their frameworks. For example, no existing framework pays enough attention to compulsory association and negative social relations.

- Longitudinal data is needed on both compliance with policy and the actual effects of compliance (whether intended or not). How many people comply with edicts banning visits to other households, and how much does this change their actual patterns of contact? Modelling will need to integrate this data going forward rather than simply making timeless assumptions. Arguably, dynamic change is not an optional refinement to epidemic models but a fundamental aspect of the logic for PHIs.

- Neglected concepts give rise to novel data requirements. For example, how do people decide whether to comply and what needs or wants is this kind of decision based on? Depending on this, can compliance fail in its effects or even be counter-productive (as when the French curfew, rather than suppressing contacts, may simply have bunched them at the start and end of the day, Dimeglio et al., 2021).

- Data on where people are and when needs to be integrated with data about environmental characteristics such as levels of pollution, humidity, and strength of air movements that may affect the likelihood of transmission of infection (Bontempi et al., 2020).

- There seems to be a specific gap in our understanding of dynamic contact structures at the boundary between several fields of research. Networks are clearly important in human contact but do not explain all contacts (while shopping for example). Similarly, while spatial proximity may be the immediate cause of infection, how that proximity comes to pass depends on human agency and structural issues (such as the need to work). Time diaries identify some patterns in contacts but do so in terms of very rough categories like "at work" or "socialising" that may not be sufficient for epidemic control. There does not seem to be any existing conceptual framework that engages with the structured diversity of life worlds in terms of their implications for the nature and timing of contacts (for example, the proportion of jobs that are "site based" rather than "multi-site").

One way to sum up these issues sociologically is to emphasise that data is needed for *process* (rather than variable or narrative) based accounts of social life (Abbott, 2016).

3   The Challenges for Agent-Based Models

While compartmental epidemiological models use differential equations to represent the dynamics of an epidemic in terms of aggregate, population-level quantities, agent-based models (ABM) explicitly represent individuals' attributes and interactions (Gilbert, 2019). Patterns emerge from these micro-level processes to explain the macro-level behaviour of real-world systems. Agent-based models can include heterogeneity in individuals' attributes and behaviour, stochastic variability in processes, and social networks constraining who interacts with whom.

Some of these aspects have been adopted by mathematical modellers. For example, the compartmental model of Davies et al. (2020b) breaks a population down into sub- compartments by age and geographical area, simulates stochastic variation in flows between compartments, and computes infectious contact rates from data that is age-based, area- specific, and context-specific ("home", "work", "school", or "other"). But an agent-based model can go further, by generating the transmission network of who infected whom, and generating sequences of the times each individual is in each disease state (Lorig et al., 2021). The personal life histories and distinctive social contexts of simulated agents can be used to compute the agent's behaviour and bodily reaction to becoming infected, and this in turn can affect the behaviour of others. These interdependencies can lead to much greater variation in aggregate behaviour than an equation-based model might suggest.

Phenomena that an equation-based modeller must neglect but an agent-based modeller might address include over-dispersion and inequality in infections. Over-dispersion is the statistical phenomenon whereby the majority of infections are caused by a small minority of infectious individuals (Endo et al., 2020). In the case of Covid-19, 80% of infections may have been caused by as few as 10% of infected people (Lewis, 2021). This phenomenon clearly corresponds to the idea that different modes of life (the nature of one's work for example) may lead to very different opportunities to infect and be infected. The degree of over-dispersion produced by an infectious disease affects the utility of particular responses, such as contact tracing (Endo et al., 2021). To study over-dispersion, or the effectiveness of measures like contact tracing, calls for a modelling approach that generates explicit information about who infects whom (rather than *rates* of infection). Likewise, an unequal distribution of cases, such as the higher case and fatality rates among ethnic minorities and lower socio-economic subpopulations, calls for a model that represents people's racial, ethnic, and economic diversity and its causal impacts on their modes of life (for example use of care homes versus family care). Model outcomes can also show surprising variation, including bursty dynamics, even if similar sets of inputs (seed infections, demographics, contact rates etc.) are used, making prediction from a sample set of simulation runs more difficult.

However, the benefits of agent-based simulations come with drawbacks, in terms of model intelligibility, computational cost, and compatibility with other scientific techniques (Chattoe-Brown, 2021). With more complexity represented, including heterogeneity, stochasticity, and networks of interdependencies, it can be much harder to understand why a model has produced the result it has. Even if we accept the complexity, to represent the same-sized population that an equation-based model can tackle requires more computer memory and time. Moreover, some of the techniques developed by epidemiologists for mathematical models remain inapplicable for ABMs. For instance, it will be harder to give mathematical explanations for system-level behaviour if an ABM computes agents' behaviour using algorithms rather than equations. Furthermore, a key method of calibrating and inferring from models, Bayesian data fitting, may be unavailable since for all but the simplest ABMs we cannot define a likelihood function.

Despite these drawbacks, progress has been made, in part accelerated as a result of the pandemic crisis and the increased interest in epidemiological modelling. Addressing the need for computing power, some ABMs have been run on high-performance computing grids (Chang et al., 2020; Mahmood et al., 2020), and scientists' collaboration with professional programmers from the computer gaming industry has resulted in dramatic speed improvements (Improbable, 2020). And, just as engagement with professional software developers led to improved confidence in the epidemiological model at Imperial College (Ferguson et al., 2020), so similar engagements may boost confidence in the code behind ABMs, and lead to higher professional standards of documentation and visual output from the models, addressing the intelligibility issue to some extent. Finally, while the use of likelihood function-free ABMs in Bayesian data fitting is in its infancy, Approximate Bayesian Computing with ABMs is a growing field (e.g., Carrella et al., 2020; Carrella, 2021).

In February 2020, when epidemiologists most needed off-the-shelf modelling tools they could customise rapidly to the new infectious disease, there were few agent-based modellers working in epidemiology. There were no ABMs to rival the compartmental models in terms of being ready to deploy, or of having a reputation for trustworthiness and applicability, or in being integrated with the work of other scientists and public health officials. The best model in a crisis is always one you have available, and suitable ABMs were not available then. One of the positive developments from the Covid-19 pandemic should be that this is no longer the case by the time of the next crisis.

Nevertheless, even if ABMs were to become more easily available for pandemic modelling, there would still be significant challenges to obtain the data they require. For calibration and validation, agent-based models ideally need longitudinal data at the individual level about people's progress through disease states. That is, they need to know, for a sample of the population, when each person entered and left each disease state, together with demographic and contextual data about that person at each transition. They also need to know about the interactions of each member of the sample with others. This is much finer-grained data than that required by compartmental models, which only deal with aggregates of "representative" agents.

4   The Need for Better Data

As the arguments above have shown, there is a need for better data to enable epidemics to be managed more effectively, where "better" means more timely, more detailed and more relevant. We need data on individuals' activities (underpinned by their goals and decisions) and their proximity to and interactions with others, and ideally this data needs to be broken down by age, social class, ethnic group, health status and location, because each of the latter variables are likely to affect the probability of transmission. In this section we review the current availability of such data, show that it falls far short of what is required and suggest what preparations need to be made to obtain the relevant data in advance of future pandemics.

During the Covid-19 pandemic, almost all epidemiological models of the spread of infection in the UK relied on one of two sources for data on interaction: the European Commission funded POLYMOD dataset and the BBC Pandemic dataset.

4.1   The POLYMOD Dataset

The POLYMOD dataset (Mossong et al., 2008) consists of data about 97,904 contacts between 7,290 people from eight European countries (Belgium, Germany, Finland, Great Britain, Italy,

Luxembourg, The Netherlands, and Poland) The data is derived from diaries that documented participants' physical and nonphysical contacts for a single day. Participants detailed the location and duration of each contact. The diaries also contained basic demographic information about the participant and the contact.

The surveys were conducted between May 2005 and September 2006 using a quota sample that was broadly representative of the population in each of the 8 countries in terms of geographical spread, age, and sex. Children and adolescents were oversampled. Only one person in each household was asked to participate in the study. Paper diaries were either sent by mail or given face to face to participants. Participants were coached by telephone or in person on how to fill in the diary. Diaries for young children were completed by a parent or guardian on their behalf.

The diaries recorded basic sociodemographic information about the participant, including employment status, level of completed education, household composition, age, and sex. Participants were assigned a random day of the week to record every person they had contact with between 5 a.m. and 5 a.m. the following morning. They were instructed to record contacted individuals only once in the diary. A contact was defined as either skin-to-skin contact such as a kiss or handshake (a physical contact), or a two-way conversation with three or more words in the physical presence of another person but no skin-to-skin contact (a nonphysical contact). Participants were also asked to provide estimates of the age and sex of each contact person. For each contact, participants were asked to record location (one of home, work, school, leisure, transport, or other), the total duration of the time spent together (less than 5 min, 5-15 min, 15 min to 1 h, 1-4 h, or 4 h or more) and the frequency of usual contacts with this individual (daily or almost daily, about once or twice a week, about once or twice a month, less than once a month, or for the first time).

The anonymised data, with details about each contact, the demographic data about the participant and the participant's household composition are freely available (Mossong, 2020). The dataset can easily be transformed to matrices of age-based contact rates by using the *R* package socialmixr (Funk, 2018). This package will generate re-scaled rates for particular populations, given the age-based demographic structure of that population. This enables the POLYMOD data to be applied to countries and geographical units other than those represented in the original data sample (Prem et al., 2017).

But despite having been designed specifically for use in epidemiological modelling (Mossong et al., 2008), the POLYMOD contact matrices have a number of features that cast doubt on their suitability for that task in the context of the Covid-19 pandemic. First, the POLYMOD study asked participants to record skin-to-skin contacts ("physical") and contacts for short conversations ("non-physical"). As Mossong et al. (2008) admit, both of these relations are distinct from spatial proximity relations, i.e., contacts who came within a given distance of the participant. While their notion of 'contact' may be appropriate for some epidemics (e.g., HIV transmission requires physical contact), it is not for others (e.g., Covid-19 can be transmitted when two people are in close proximity, without need for either physical or non-physical contact as defined by Mossong et al.).

Secondly, as well as proximity, POLYMOD data omit information important for understanding the potential for transmission of Covid-19 and other airborne diseases. In their review of the early literature on Covid-19 transmission dynamics, Cevik et al. (2020) identify not just that proximity to infectious people is a factor, but also that transmission is more likely to occur indoors rather than outdoors, in poorly ventilated and small rooms, and where people are breathing more heavily, such as while talking loudly in noisy environments and singing

in choirs. The POLYMOD matrices only record where social contact occurred using a coarse "generic" classification. We might assume that "home" contacts are nearly all indoors, except perhaps in warmer, drier months of the year, and "work" and "school" contacts are mostly indoors, but the "Other" category could include locations as varied as pubs, gyms, buses, and parks, with greatly varying transmission risks. Also unrecorded are what activities were being performed in the locations, and how many other people may have been in proximity but not engaging in conversations or skin-to-skin contacts with participants. A POLYMOD participant could spend an evening drinking alone in a poorly ventilated pub, crowded with people talking loudly, all of whom are complete strangers to the study participant. In such a situation, there is a high risk of airborne transmission, but no "contact" would be recorded. The mismatch between POLYMOD contact relations and likely transmission routes is all the more frustrating when one considers that the contact matrices are used to evaluate non-pharmaceutical interventions for which proximity is relevant, including social distancing, mask wearing, isolation of suspected cases and of vulnerable people, and closures of schools and non-essential businesses (Davies et al., 2020b).

Thirdly, we note that the POLYMOD matrices do not report the duration of contacts. Eight hours spent in the same bedroom with a partner counts as one contact, but so does a three-word conversation with a bus driver. Transmission risk increases with time spent in close contact with an infectious person (Cevik et al., 2020). There is also some suggestion that longer contacts provide a higher dose of viral particulates and raise the chance of severe illness. But the epidemiological models using POLYMOD contacts treat each contact equally when they update at regular time steps (e.g., Davies et al., 2020b).

Fourthly, the POLYMOD matrices omit with whom contacts were made, and in particular whether today's set of contacts overlap with yesterday's. In the home, nearly all contacts will be with close family, in schools and many workplaces they will mostly be with the same set of people, but in some occupations one may encounter a wide range of new people every day. But repetition of contact is important for correlations between transmission risks. If one's family, schoolmates, or work colleagues were non-infectious yesterday, there is a good chance they are still so today. If one of them were infectious, they will probably remain a threat to others for several days until they isolate. Studies of epidemics in network models bear out the importance of contact partners' identities (Newman, 2002). Epidemics traverse random networks faster than regular, spatial ones. The universal mixing assumed by differential equation-based models is analogous to a dynamically changing random network. Real-world social networks will change much less often from day to day, and will exhibit spatial and social regularities with the potential to slow epidemic spread.

Fifthly, the POLYMOD matrices state only average numbers of contacts per person, not deviations from mean rates. Everyone in an age group is given the same role in transmission. There can be no "superspreaders" when using a POLYMOD matrix. But school teachers and those in customer-facing roles are likely to have a much larger number of contacts than others. The POLYMOD data gives no indication of participants' occupations or lifestyles, so we cannot identify particular types of participant who might have unusual contact patterns.

Using scale-free networks, which have long-tailed distributions in the number of neighbours per node, Newman (2002) has studied cases where the higher variance in number of contact partners led to faster epidemics than in other types of network with the same mean number of partners. Given the POLYMOD source data, a list of contacts, it would be possible to calculate the variance in the number of contact partners for each pair of age groups, although this calculation is not offered by the socialmixr R package of Funk (2018). But it is

not clear how it could be incorporated into a differential equation-based model, other than by subdividing the population according to whether they had relatively higher or lower contact rates with particular age bands, although it could be added to an agent-based model without much difficulty.

Finally, we note that the POLYMOD matrices offer little hint as to how contact rates may change during the course of a pandemic. There is no information about the social mechanisms that generated the contacts in the data set, other than the typology of where the contacts took place (home, work, school, etc.). If those mechanisms change — for example, as a result of a person choosing to stay home while ill, or because their workplace has been closed — we have no guidance as to how their contact rate would change. Closing the schools can be represented as setting "school"-context contacts to zero or to some small percentage to allow for the continuing attendance of the children of key workers (as was permitted in the UK), but do the school children then spend all their free time at home, or do they go to the park? Should "home" and "other" contacts increase as opportunity to visit them increases? This is an advantage of the modelling approach proposed by Dignum (2021) in which goals (such as sociability) may vary in their outcomes (e.g., people meeting in the park rather than at home).

Just as people are heterogeneous in contact rates during normal times, their responses to interventions may be heterogeneous. Key workers may still have to attend their workplaces with their children needing to be supervised in schools, even while their schoolmates are being home-schooled by parents off-work. As well as varying in duties and the capability to change contact rates, people may vary in their motivations, with some reluctant to modify their favourite behaviours. So a lot of complex human behaviour has to be represented by simple alterations to model parameters. Davies et al. (2020b) represented non-pharmaceutical interventions (NPIs) purely as reductions in contact rates, but little rationale was given for the specific sizes of reductions. The impression is of modelling experiments, sophisticated in many other respects, resting on rather arbitrary assumptions when it comes to the representation of NPIs. But if the purpose of the model is to estimate the effectiveness of those NPIs, this is a serious weakness.

To sum up, the POLYMOD matrices are based on the wrong relations, and encourage the misrepresentation of social contacts in their duration, repetition of partners, variance in rates, and degree to which interventions and behavioural responses to the disease will alter them. They also lack context, both in the activities where the contacts occur and the agency of the individuals who are in contact.

4.2   The BBC Pandemic Datasets

In 2018, the UK TV channel, BBC4, broadcast a documentary about a simulation of a pandemic spreading through the UK. The programme was based on the collection of two datasets: one derived from data reported from a nationally available app, "BBC Pandemic", that people were encouraged to download onto their personal smartphones, and one from data from an app that people living in a small English town, Haslemere, were encouraged to use. The national app recorded volunteers' movements and self-reported contacts with other people over the course of one day. The resulting data, obtained from 28,947 app users, provides user profiles, including age, location logs (hourly to one kilometre resolution) and self-reported basic descriptions of those the volunteers encountered (Klepac et al., 2018). The second dataset consists of 1,616 daily contact events and 1,257 unique social links among adults in Haslemere (Kissler et al., 2020). The 469 volunteers in the town (constituting 4.2% of the town's inhabitants) were asked to download and carry a smartphone app and their proximity to other

volunteers was recorded at 5 minute intervals over a period of three days.

While in many ways the BBC Pandemic datasets are pathbreaking, they have limitations. The complete national dataset is not available for research because of ethical constraints resulting from the potential identifiability of the participants, but a redacted dataset was released in early 2021 (Conlan et al., 2021). Timestamped proximity data is available from the Haslemere dataset, but without any demographic information about the individuals, again because of ethical constraints. The datasets only include those aged 16 or over who owned mobile phones and were interested in participating, potentially introducing volunteer bias into the sample.

The national dataset is based on a larger sample than the POLYMOD dataset and has the advantage of including spatial information (although only with 1km granularity, which means that it is not possible to identify the type of location where the contact was made). However, it suffers from most of the same disadvantages as POLYMOD when used for epidemiological modelling. The Haslemere dataset is of course representative neither in the location of its sample, nor in terms of its population (which is much more middle class than the UK as a whole), and the sample size is small.

5   Conclusion: Recommendations for Better Data

This paper has highlighted why we need more appropriate data before the next pandemic, and more generally for the development of sociological theory, but the question remains *what* data and *how* should this be obtained?

Compartmental models can only show aggregate dynamics. Agent-based modelling provides a method to produce more detailed, fine resolution models of individual interaction potentially leading to more accurate and more useful models of epidemic dynamics, but these models require both more detailed, individual- rather than collective-level data and contextual information about physical locations and individual goals and decisions.

To summarise, the calibration and validation of good epidemic models needs data about:

- The everyday lives of a representative sample of individuals, with their demographic characteristics (age, sex, ethnicity, disability status etc.), the composition of their households, and their economic status (employment status, income band or occupational class). The location, the type of place they are in (including physical characteristics such as ventilation), the number of people who are in close proximity to them, and their contacts, both physical (e.g., touching) and social (e.g., talking) need to be observed for representative days (e.g., both weekdays and weekends, and in different seasons).

- The perceptions of the sample and the meaning they ascribe to government edicts and community and peer group pressures and norms. These perceptions also involve, for example, the feasibility of compliance and the valuation of general adherence relative to individual circumstance.

- Changes in behaviour as a result of social, physical and medical changes due to the pandemic and the constraints under which people are acting (e.g., the closure of schools).

These data should be obtained for a cohort observed over as long a period as possible (e.g., over the whole duration of the epidemic). Although we are considering the opportunities resulting from the Covid pandemic, we should also ensure that the data gathering strategy is

robust enough that if another pandemic were to arise with different transmission characteristics, or another kind of emergency occurred, for example widespread civil unrest, we could easily use the data for public benefit.

One obvious way of gathering this data would be through using smartphone apps. Previous attempts such as the BBC Pandemic data gathering in Haslemere collected proximity information only when other people who have the app installed come close, which may be a low proportion of the population. In contrast, the NHS COVID-19 app, which has a relatively high take-up in the population,[3] only collects very limited information to encourage its use (Park, 2021). One way forward would be to extend existing methods such as time diaries and network data acquisition (Gershuny et al., 2021; Sullivan et al., 2020) using smartphone sensors by recording Bluetooth or video data, while using privacy preserving technologies (see Cadman & Freeman, 2020 for an account of some of the technical and organisational difficulties of doing so). The pandemic has changed the public's perception of the trade-off between privacy and public benefit (Farjam et al., 2021), and this may make recruiting respondents to research which has demonstrable benefits for controlling epidemics easier. However, we are also mindful that there is a "political" dimension to institutional data collection processes (see, for example, Bruno et al., 2014; Alteri et al., 2021) and simply presenting a "reasoned" argument may not be sufficient to build a coalition in support of the necessary data.

Data collection of this kind is expensive, both in terms of paying for people to plan, organise, host, and curate data sets, and for the actual collection. In terms of financial cost, the POLYMOD data set cost €1.8million (CORDIS, 2009). The richer data that we believe is required to model a future pandemic would probably cost at least an order of magnitude more: perhaps € 20 million annually. However, such an expenditure is insignificant in relation to the benefit it could provide during a pandemic. Furthermore, novel forms of data — about goals and decisions for example — would be valuable to many other areas of sociology such as understanding changes in political sentiment, the underpinnings of social capital and social support, and the various ways in which societies can undergo differentiation on a large scale. Despite the cost, the economic and social savings and the impact on health would make the investment well worthwhile. And since setting up such data collection takes time, we need to start now.

References

Abbott, A. (2016). *Processual Sociology*. Chicago: University of Chicago Press.

Alteri, L., Parks, L., Raffini, L., & Vitale, T. (2021). Covid-19 and the Structural Crisis of Liberal Democracies. Determinants and Consequences of the Governance of Pandemic. *Partecipazione e Conflitto*, *14*(1), 1–37. https://doi.org/10.1285/i20356609v14i1p01

Anderson, R.M. & May, R.M. (1992). *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press.

Azad, S. & Devi, S. (2020). Tracking the Spread of Covid-19 in India via Social Networks in the Early Phase of the Pandemic. *Journal of Travel Medicine*, *27*(8), taaa130. https://doi.

---

3. The app had been downloaded about 21 million times by December 2020, (Statista, https://www.statista.com/statistics/1190062/covid-19-app-downloads-uk/) which can be compared with the adult population of England and Wales in 2020 of about 60 million.

org/10.1093/jtm/taaa130

Biggs, A.T. & Littlejohn, L.F. (2021). Revisiting the Initial Covid-19 Pandemic Projections. *The Lancet Microbe*, *2*(3), e91–e92. https://doi.org/10.1016/S2666-5247(21)00029-X

Birrell, P., Blake, J., van Leeuwen, E., Gent, N., & De Angelis, D. (2021). Real-Time Now-casting and Forecasting of Covid-19 Dynamics in England: The First Wave. *Philosophical Transactions of the Royal Society B*, *376*(1829), 20200279. https://doi.org/10.1098/rstb.2020.0279

Bontempi, E., Vergalli, S., & Squazzoni, F. (2020). Understanding Covid-19 Diffusion Requires an Interdisciplinary, Multi-dimensional Approach. *Environmental Research*, *188*, 109814. https://doi.org/10.1016/j.envres.2020.109814

Brooks-Pollock, E., Christensen, H., Trickey, A., Hemani, G., Nixon, E., Thomas, A.C., & Danon, L. (2021). High Covid-19 Transmission Potential Associated with Re-opening Universities Can Be Mitigated with Layered Interventions. *Nature Communications*, *12*, 1507. https://doi.org/10.1038/s41467-021-25169-3

Bruno, I., Didier, E., & Vitale, T. (2014). Statactivism: Forms of Action between Disclosure and Affirmation. *Partecipazione e Conflitto*, *7*(2), 198–220. https://doi.org/10.1285/i20356609v7i2p198

Burgess, S., Ponsford, M.J., & Gill, D. (2020). Editorial: Are We Underestimating Seroprevalence of SARS-CoV-2? *British Medical Journal*, *370*, m3364. https://doi.org/10.1136/bmj.m3364

Cadman, P. & Freeman, S. (2020). A New Proximity Risk Calculation for the NHS Test & Trace Covid App. *Zuehlke*. https://www.zuehlke.com/en/insights/a-new-proximity-risk-calculation-for-the-nhs-test-trace-covid-app

Carrella, E. (2021). No Free Lunch When Estimating Simulation Parameters. *Journal of Artificial Societies and Social Simulation*, *24*(2), 7. https://doi.org/10.18564/jasss.4572

Carrella, E., Bailey, R., & Madsen, J.K. (2020). Calibrating Agent-Based Models with Linear Regressions. *Journal of Artificial Societies and Social Simulation*, *23*(1), 7. https://doi.org/10.18564/jasss.4150

Cevik, M., Marcus, J.L., Buckee, C., & Smith, T.C. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Transmission Dynamics Should Inform Policy. *Clinical Infectious Diseases*, *73*(Supplement 2), S170–S176. https://doi.org/10.1093/cid/ciaa1442

Chang, S.L., Harding, N., Zachreson, C., Cliff, O.M., & Prokopenko, M. (2020). Modelling Transmission and Control of the Covid-19 Pandemic in Australia. *Nature Communications*, *11*, 5710. https://doi.org/10.1038/s41467-020-19393-6

Chattoe-Brown, E. (2021). Why Questions Like "Do Networks Matter?" Matter to Methodology: How Agent-Based Modelling Makes It Possible to Answer Them. *International Journal of Social Research Methodology*, *24*(4), 429–442. https://doi.org/10.1080/13645579.2020.1801602

Chua, A.Q., Al Knawy, B., Grant, B., Legido-Quigley, H., Lee, W.-C., Leung, G.M., & Maurer-Stroh, S. (2021). How the Lessons of Previous Epidemics Helped Successful Countries Fight Covid-19. *British Medical Journal*, *372*, n486. https://doi.org/10.1136/bmj.n486

Conlan, A.J.K., Klepac, P., Kucharski, A.J., Kissler, S.M., Tang, M.L., Fry, H., & Gog, J.R. (2021). Human Mobility Data from the BBC Pandemic Project. *medRxiv*. https://doi.org/10.1101/2021.02.19.21252079

CORDIS (2009). Improving Public Health Policy in Europe through Modelling and Economic Evaluation of Interventions for the Control of Infectious Diseases. *CORDIS EU Research Results*. https://cordis.europa.eu/project/id/502084

Daly, M. (2020). Covid-19 and Care Homes in England: What Happened and Why? *Social Policy and Administration*, *54*(7), 985–998. https://doi.org/10.1111/spol.12645

Davies, N.G., Klepac, P., Liu, Y., Prem, K., Jit, M., CMMID Covid-19 Working Group, & Eggo, R.M. (2020a). Age-dependent Effects in the Transmission and Control of Covid-19 Epidemics. *Nature Medicine*, *26*, 1205–1211. https://doi.org/10.1038/s41591-020-0962-9

Davies, N.G., Kucharski, A.J., Eggo, R.M., Gimma, A., & Edmunds, W.J. on behalf of the Centre for the Mathematical Modelling of Infectious Diseases Covid-19 Working Group (2020b). Effects of Non-pharmaceutical Interventions on Covid-19 Cases, Deaths, and Demand for Hospital Services in the UK: A Modelling Study. *Lancet Public Health*, *5*(7), E375–E385. https://doi.org/10.1016/S2468-2667(20)30133-X

Denford, S., Morton, K.S., Lambert, H., Zhang, J., Smith, L.E., Rubin, J.G., & Yardley, L. (2020). Understanding Patterns of Adherence to Covid-19 Mitigation Measures: A Qualitative Interview Study. *medRxiv*. https://doi.org/10.1101/2020.12.11.20247528

Dignum, F. (Ed.). (2021). *Social Simulation for a Crisis: Results and Lessons from Simulating the Covid-19 Crisis*. Cham: Springer.

Dimeglio, C., Miedougé, M., Loubes, J.M., Mansuy, J.M., & Izopet, J. (2021). Side Effect of a 6 pm Curfew for Preventing the Spread of SARS-CoV-2: A Modeling Study from Toulouse, France. *Journal of Infection*, *82*(5), 186–230. https://doi.org/10.1016/j.jinf.2021.01.021

Elgethun, K., Yost, M.G., Fitzpatrick, C.T., Nyerges, T.L., & Fenske, R.A. (2007). Comparison of Global Positioning System (GPS) Tracking and Parent-Report Diaries to Characterize Children's Time-Location Patterns. *Journal of Exposure Science and Environmental Epidemiology*, *17*(2), 196–206. https://doi.org/10.1038/sj.jes.7500496

Endo, A., Abbott, S., Kucharski, A., & Funk, S. (2020). Estimating the Overdispersion in Covid-19 Transmission Using Outbreak Sizes outside China [version 3; peer review: 2 approved]. *Wellcome Open Research*, *5*, 67. https://doi.org/10.12688/wellcomeopenres.15842.3

Endo, A., Leclerc, Q., Knight, G., Medley, G., Atkins, K., Funk, S., & Kucharski, A. (2021). Implication of Backward Contact Tracing in the Presence of Overdispersed Transmission in Covid-19 Outbreaks [version 2; peer review: 2 approved]. *Wellcome Open Research*, *5*, 239. https://doi.org/10.12688/wellcomeopenres.16344.2

Farjam, M., Bianchi, F., Squazzoni, F., & Bravo, G. (2021). Dangerous Liaisons: An Online Experiment on the Role of Scientific Experts and Politicians in Ensuring Public Support for Anti-Covid Measures. *Royal Society Open Science*, *8*(3), 201310. https://doi.org/10.1098/rsos.201310

Ferguson, N.M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., & Ghani, A.C. (2020). *Report 9 – Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce Covid-19 Mortality and Healthcare Demand.* https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/

Funk, S. (2018). *Socialmixr: Social Mixing Matrices for Infectious Disease Modelling* [data collection]. The Comprehensive R Archive Network, R package version 0.0.1. https://CRAN.R-project.org/package=socialmixr

Gershuny, J., Sullivan, O., Sevilla, A., Vega-Rapun, M., Foliano, F., Lamote de Grignon, J., & Walthery, P. (2021). A New Perspective from Time Use Research on the Effects of Social Restrictions on Covid-19 Behavioral Infection Risk. *PLOS ONE*, *16*(2), e0245551. https://doi.org/10.1371/journal.pone.0245551

Gilbert, N. (2019). *Agent-Based Models* (2nd ed.). London: Sage.

Gmel, G. & Daeppen, J.B. (2007). Recall Bias for Seven-Day Recall Measurement of Alcohol Consumption among Emergency Department Patients: Implications for Case-Crossover Designs. *Journal of Studies on Alcohol and Drugs*, *68*(2), 303–310. https://doi.org/10.15288/jsad.2007.68.303

Hendry, D.F. & Richard, J.F. (1983). The Econometric Analysis of Economic Time Series. *International Statistical Review*, *51*(2), 111–148. https://doi.org/10.2307/1402738

Hills, S. & Eraso, Y. (2021). Factors Associated with Non-Adherence to Social Distancing Rules during the Covid-19 Pandemic: A Logistic Regression Analysis. *BMC Public Health*, *21*, 352. https://doi.org/10.1186/s12889-021-10379-7

Hoeben, E.M., Bernasco, W., Liebst, L.S., van Baak, C., & Lindegaard R.M. (2021). Social Distancing Compliance: A Video Observational Analysis. *PLOS ONE*, *16*(3), e0248221. https://doi.org/10.1371/journal.Pone.0248221

Improbable (2020). *Synthetic Environment Technology Accelerates Pandemic Modelling.* https://www.improbable.io/blog/improbable-synthetic-environment-technology-accelerates-uk-pandemic-modelling

Jorge, A., D'Silva, K., Cohen, A., Wallace, Z.S., McCormick, N., Zhang, Y., & Choi, H.K. (2021). Temporal Trends in Severe Covid-19 Outcomes in Patients with Rheumatic Disease: A Cohort Study. *Lancet Rheumatology*, *3*(2), e131–e137. https://doi.org/10.1016/S2665-9913(20)30422-7

Khunti, K., Singh, A.K., Pareek, M., & Hanif, W. (2020). Is Ethnicity Linked to Incidence or Outcomes of Covid-19? *British Medical Journal*, *369*, m1548. https://doi.org/10.1136/bmj.m1548

Kissler, S.M., Klepac, P., Tang, M., Conlan, A.J.K., & Gog, J.R. (2020). Sparking "The BBC Four Pandemic": Leveraging Citizen Science and Mobile Phones to Model the Spread of Disease. *BioRxiv*, 479154. https://doi.org/10.1101/479154

Klepac, P., Kissler, S., & Gog, J. (2018). Contagion! The BBC Four Pandemic – the Model behind the Documentary. *Epidemics*, *24*, 49–59. https://doi.org/10.1016/j.epidem.2018.03.003

Lee, D., Heo, K., & Seo, Y. (2020). Covid-19 in South Korea: Lessons for Developing Countries. *World Development*, *135*, 105057. https://doi.org/10.1016/j.worlddev.2020.105057

Lewis, D. (2021). Superspreading Drives the Covid Pandemic – and Could Help to Tame It. *Nature*, *590*(7847), 544–546. https://doi.org/10.1038/d41586-021-00460-x

Lorig, F., Johansson, E., & Davidsson, P. (2021). Agent-Based Social Simulation of the Covid-19 Pandemic: A Systematic Review. *Journal of Artificial Societies and Social Simulation*, *24*(3), 5. https://doi.org/10.18564/jasss.4601

Mahmood, I., Arabnejad, H., Suleimenova, D., Sassoon, I., Marshan, A. Serrano-Rico, A. & Groen, D. (2020). FACS: A Geospatial Agent-Based Simulator for Analysing Covid-19 Spread and Public Health Measures on Local Regions. *Journal of Simulation*. https://doi.org/10.1080/17477778.2020.1800422

Manzo, G. (2020). Complex Social Networks Are Missing in the Dominant Covid-19 Epidemic Models. *Sociologica*, *14*(1), 31–49. https://doi.org/10.6092/issn.1971-8853/10839

Manzo, G. & Van de Rijt, A. (2020). Halting SARS-CoV-2 by Targeting High-Contact Individuals. *Journal of Artificial Societies and Social Simulation*, *23*(4), 10. https://doi.org/10.18564/jasss.4435

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., & Edmunds, W.J. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine*, *5*, e74. https://doi.org/10.1371/journal.pmed.0050074

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., & Edmunds, W.J. (2020). *POLYMOD Social Contact Data (Version 2)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1043437

Mullan, K. & Chatzitheochari, S. (2019). Changing Times Together? A Time-Diary Analysis of Family Time in the Digital Age in the United Kingdom. *Journal of Marriage and Family*, *81*(4), 795–811. https://doi.org/10.1111/jomf.12564

Newman, M.E.J. (2002). Spread of Epidemic Disease on Networks. *Physical Review E*, *66*(1), 016128. https://doi.org/10.1103/PhysRevE.66.016128

Paolisso, M. & Hames, R. (2010). Time Diary versus Instantaneous Sampling: A Comparison of Two Behavioral Research Methods. *Field Methods*, *22*(4), 357–377. https://doi.org/10.1177/1525822X10379200

Park, J. (2021). Governing a Pandemic with Data on the Contactless Path to AI: Personal Data, Public Health, and the Digital Divide in South Korea, Europe and the United States in Tracking of Covid-19. *Partecipazione e Conflitto*, *14*(1), 79–112. https://doi.org/10.1285/i20356609v14i1p79

Prem, K., Cook, A.R., & Jit, M. (2017). Projecting Social Contact Matrices in 152 Countries Using Contact Surveys and Demographic Data. *PLOS Computational Biology*, *13*(9), e1005697. https://doi.org/10.1371/journal.pcbi.1005697

Santamaria, C., Sermi, F., Spyratos, S., Iacus, S.M., Annunziato, A., Tarchi, D., & Vespe, M. (2020). Measuring the Impact of Covid-19 Confinement Measures on Human Mobility Using Mobile Positioning Data. A European Regional Analysis. *Safety Science*, *132*, 104925. https://doi.org/10.1016/j.ssci.2020.104925

Santelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., & Vineis, P. (2020). Five Ways to Ensure that Models Serve Society: A Manifesto. *Nature Human Behaviour*, *582*, 482–484. https://doi.org/10.1038/d41586-020-01812-9

SoleimanvandiAzar, N., Irandoost, S.F., Ahmadi, S., Xosravi, T., Ranjbar, H., Mansourian, M., & Lebni, J.Y. (2021). Explaining the Reasons for Not Maintaining the Health Guidelines to Prevent Covid-19 in High-Risk Jobs: A Qualitative Study in Iran. *BMC Public Health*, *21*(1), 1–15. https://doi.org/10.1186/s12889-021-10889-4

Squazzoni, F., Polhill, J.G., Edmonds, B., Ahrweiler, P., Antosz, P., Scholz, G. & Gilbert, N. (2020). Computational Models that Matter during a Global Pandemic Outbreak: A Call to Action. *Journal of Artificial Societies and Social Simulation*, *23*(2), 10. https://doi.org/10.18564/jasss.4298

Sullivan, O., Gershuny, J., Sevilla, A., Walthery, P., & Vega-Rapun, M. (2020). Time Use Diary Design for Our Times – an Overview, Presenting a Click-and-Drag Diary Instrument (CaDDI) for Online Application. *Journal of Time Use Research*, *15*(1), 1–17. https://jtur.iatur.org/home/article/c73705a3-2c6f-46d4-9616-0f197e40455c

Nigel Gilbert – Centre for Research in Social Simulation (CRESS), University of Surrey (United Kingdom)

🆔 https://orcid.org/0000-0002-5937-2410

✉ n.gilbert@surrey.ac.uk; ⤤ https://cress.soc.surrey.ac.uk/web/people/ngilbert

Nigel Gilbert holds a Distinguished Chair at the University of Surrey (United Kingdom) and is a computational social scientist. He was one of the first to use agent-based models in the social sciences and has since published widely on the methodology underlying computer modelling and on the application of simulation for applied and policy related problems.

Edmund Chattoe-Brown – School of Media, Communication and Sociology, University of Leicester (United Kingdom)

🆔 https://orcid.org/0000-0001-8232-6896

⤤ https://www2.le.ac.uk/departments/sociology/people/echattoebrown

Edmund Chattoe-Brown's career has focused on the value of research methods (particularly Agent-Based Modelling) in generating warranted knowledge of society. His aim has been to make such models both more usable in social science generally and particularly more empirical. The results have been published in 17 different peer reviewed journals across the sciences to date.

Christopher Watts – Independent researcher, Cambridgeshire (United Kingdom)

🆔 https://orcid.org/0000-0002-0861-9815

⤤ https://github.com/innovative-simulator

Christopher J. Watts specialises in agent-based computer simulation. He is currently an independent researcher based in Cambridgeshire (United Kingdom). He previously worked for the Universities of Munich (in Geography), Surrey (Sociology) and Warwick (Business School). His background includes degrees in Operational Research, and Philosophy.

Duncan Robertson – School of Business and Economics, Loughborough University (United Kingdom)

🆔 https://orcid.org/0000-0002-7801-5451

⤤ https://www.lboro.ac.uk/departments/sbe/staff/duncan-robertson/

Duncan Robertson joined the Management Science and Operations Group at Loughborough in 2014. He is a Fellow and Trustee of St. Catherine's College in the University of Oxford. He originally trained as a physicist at Imperial College London.