Data needs for integrated economic-epidemiological models of pandemic mitigation policies

David J. Haw [a,*], Christian Morgenstern [a], Giovanni Forchini [b,a], Rob Johnson [a], Patrick Doohan [a], Peter C. Smith [c,d], Katharina D. Hauck [a]

[a] MRC Centre for Global Infectious Disease Analysis & WHO Collaborating Centre for Infectious Disease Modelling, Abdul Latif Jameel Institute for Disease and Emergency Analytics, Imperial College London, United Kingdom
[b] USBE, Umeå Universitet, SE-901 87 Umeå, Sweden
[c] Department of Economics and Public Policy, Imperial College Business School, United Kingdom
[d] Centre for Health Economics, University of York, United Kingdom

ARTICLE INFO

ABSTRACT

The COVID-19 pandemic and the mitigation policies implemented in response to it have resulted in economic losses worldwide. Attempts to understand the relationship between economics and epidemiology has led to a new generation of integrated mathematical models. The data needs for these models transcend those of the individual fields, especially where human interaction patterns are closely linked with economic activity. In this article, we reflect upon modelling efforts to date, discussing the data needs that they have identified, both for understanding the consequences of the pandemic and policy responses to it through analysis of historic data and for the further development of this new and exciting interdisciplinary field.

1. Introduction

Economic disruption during the COVID-19 pandemic has led to a wealth of mathematical models that integrate the fields of economics and epidemiology, e.g. Haw et al. (2022), Eichenbaum et al. (2022), Acemoglu et al. (2020), Çakmaklı et al. (2020), Pichler et al. (2020). A common theme across such studies is the association of an economic cost to the efforts made in mitigating the COVID-19 pandemic. Models analyse the impact of policy interventions (or lack thereof) and behavioural effects on the dynamics of infection transmission and the economy of a particular location.

In this article, we discuss the data needs of integrated models, focusing on studies that model the impact of lockdowns and/or individuals' behaviour: mandated closures of non-essential business activity that impact on both disease transmission and economic activity or output. In these models, force-of-infection is usually affected by economic closures, i.e. mandated closures that reduce transmission and also manifest in economic variables. The mechanistic relationship between economic closures and transmission hazard yields mathematical models that project the impact of an economic decision on the spread of a disease, or vice versa. Fig. 1 illustrates the causal relationship between integrated model variables and their manifestation in current data sources reported longitudinally throughout the COVID-19 pandemic. The economic variables "policy" and "behaviour" denote any degree

of mandated and voluntary mitigation respectively. We define an integrated model as one in which at least one of the bold arrows is accounted for, i.e. an explicit causal relationship between an economic variable and an epidemiological one. The exact manifestation of this relationship is dependent on the question being asked, though questions regarding the interplay between disease dynamics and economic activity are increasingly prevalent. Table 1 contains a breakdown of the data sources referenced in the figure, together with other existing sources discussed in this manuscript. With many new, large data sources emerging in recent times, and many corresponding studies still in the process of publication, any attempt to classify data for integrated models would be quickly outdated. Our review of current data sources is thus not exhaustive, but we aim to acknowledge the broad data types that have recently emerged and identify common limitations faced when addressing topical scientific questions. We also extend the notion of "data needs" beyond data that does not exist, including inconsistency of data across countries or institutions, imprecision of survey questions, and public inaccessibility of data.

Motivated almost entirely by the COVID-19 pandemic, reflection on data sources since December 2019 allows us to identify data needs for modelling the macro-economic impact of a pandemic on the economy, and the micro-economic impact on the behaviour and decisions of individual consumers and businesses. Our discussion assumes a readership


Table 1
Current key economic data sources for integrated modelling. Data needs, as defined in our introduction, are highlighted in the right-hand column, "limitations".

Data type/Source	Notes	Limitations
GVA		
Country-specific delayed reporting, e.g. Office for National Statistics (2022) (UK), Australian Bureau of Statistics (2022) (Australia), Badan Pusat Statistik (2022) (Indonesia)	Sector-specific contributions to GDP	Reported retrospectively, by monthly interval Classification of businesses into sectors is inconsistent over time and between countries
Transactions		
Yodlee data analytics (Yodlee, 2022)	Clean tagged and anonymised consumer transaction data of public and private merchants	Available only at high monetary cost
Transctions from major banks, used to inform (ONS, 2022)	Weekly indices of value and volume of credit/debit card transactions by group of Merchant Category Codes (MCCs)	Not freely available; Spatially disaggregated by location of HQ; MCCs do not align with sectors of GVA/IO table configurations
Mobility		
Google COVID-19 Community Mobility Reports (Google, 2022)	Changes over time in visits to: retail and recreation, supermarket and pharmacy, parks, public transport, workplaces, residential; Available by country/region	Baseline data is Jan/Feb 2020, giving no true pre-pandemic reference for seasonality etc.; No distinction between Workplaces
Stringency		
COVID-19 behaviour tracker (Hale et al., 2021; Ritchie et al., 2020)	Stringency index comprised of: school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport stay-at-home requirements, public information campaigns, restrictions on internal movements international travel controls	Index and constituents highly coarse grained at national level
Sector-stratified contact rates		
Two known sources: (Béraud et al., 2015; Danon et al., 2013)	Workforce surveys reporting daily contact rates	Some sample sizes are small and not representative; Limited disaggregation of sectors
Education sector breakdown		
ONS COVID-19 Schools Infection Survey (SIS, Office for National Statistics, 2020c)	COVID-19 Swab test and survey of schoolchildren; Includes household composition	Not freely available
Attendance in education (GOVschools, 2022)	School and early years settings	No linkage to household structures
Survey data		
COVID-19 behaviour tracker (Imperial College/YouGov)	Global insights on people's behaviours in response to COVID-19	Survey discontinued.
REACT-1 (Riley et al., 2021)	Longitudinal swab and survey data; Large number of variables reported at individual level within and within households	Limited identification of employment sector; Not freely available;
ONS COVID-19 Infection Survey (SIS, Office for National Statistics, 2020a)	Longitudinal swab and survey data stratified by age, region and vaccine status	Not freely available; Full survey contents not publically available
ONS Labour Force Survey (LFS, Office for National Statistics, 2020b)	Employment circumstances of the UK population	Quarterly temporal aggregation (UK)
ONS Survey on Living Conditions (SLC, Office for National Statistics, 2020d)	Information on household resources, housing, labour, education, pensions and health	Only two time points: May 2021 and September 2021; Two binary output variables: Ability to afford unexpected expense/holiday, Ability to pay usual expenses

familiar with the basic underlying models. Many of our observations are rooted in the integrated model DAEDALUS (Haw et al., 2022), though broader literature searches of relevant models (Brodeur et al., 2021) have shown these observations to be somewhat universal. Many of the existing datasets identified are unique to the United Kingdom, though our focus is on the nature of the variables and the scope of the study, rather than the population from which such data was obtained. The sections in this article highlight data needs when modelling the relationship between economic activity and the transmission of an infectious disease in a population. We begin by breaking down the

interface between economics and epidemiology, identifying the data heterogeneities required for alignment of models. We then provide a brief introduction to economic models for a readership of epidemiologists, before discussing a crucial metric of economic output, namely Gross Value Added (GVA), and the way it shapes economic modelling. One particularly difficult heterogeneity to fully encapsulate is that of geographical space, which we discuss in detail. We then explain the relationship between economic activity and physical contact outside the workplace, and the broader issue of changes in individual behaviour that give rise to changes in consumer demand. Our discussion of models
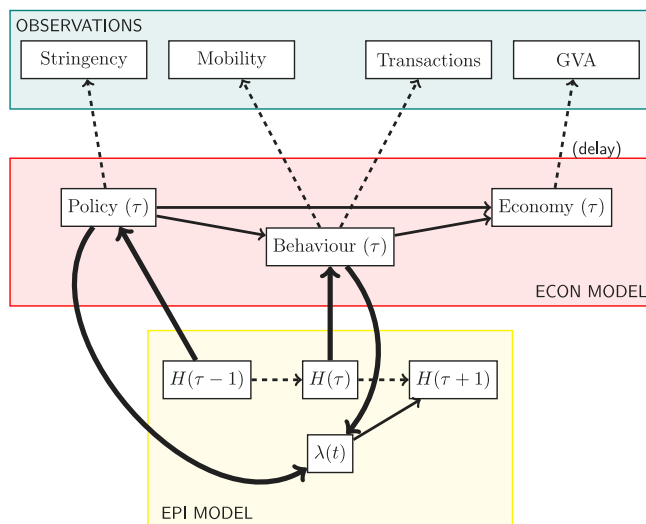
Fig. 1. Schematic of integrated epi/econ models: epidemiological variables are grouped at the bottom (yellow), economic variables displayed in the centre of the figure (red), and current economic data at the top (green). For simplicity we exclude core epidemiological data (e.g. hospitalisations and deaths) from this diagram. The letter τ indexes time intervals, typically of one month in duration. GVA in period τ is reported in period $\tau + 1$. In the epidemiological model, $H(\tau)$ denotes total hospital occupancy in period τ, and $\lambda(t)$ the (continuously varying) Force Of Infection (FOI). Solid arrows denote dependencies between variables. Bold arrows show the causal relationships between economic and epidemiological phenomena. Dashed arrows show the manifestation of economic variables in current data sources, and indicate the progression of time periods in the epidemiological part of the model.

and data sources will become increasingly granular, aiming to illustrate the varying levels of model complexity that researchers may wish to consider. As always, complexity is dependent on the question we are asking, but may also be limited by the data sources that are available to inform a model. A broad research question to keep in mind throughout our discussion is the following: "how can we characterise an optimal mitigation strategy that accounts for economic impact, and how could we find such a strategy in a future outbreak?" We conclude with a table of notable known data sources that illustrate the data needs discussed, briefly summarising their contents and limitations.

2. The interface of economics and epidemiology

Economics meets epidemiology where people meet people. Much of the heterogeneity in epidemiological models is born of the differences in contact rates (people per day) with respect to demographic phenomena, most notably age (Mossong et al., 2008; Prem et al., 2017) and space (Haw et al., 2019). Such heterogeneities are key in replicating observed data in mathematical models (Gog et al., 2014). The relationship between economics and epidemiology that we focus on can be summarised as follows: partial economic closures reduce the active workforce interacting with each other and with customers, causing a reduction in contacts between people, reducing transmission risk. This reduced workforce also yields a reduction in production of goods and services. Likewise, the incentive to reduce contact with others causes a reduction in demand as fewer people choose to spend money outside the home, and labour force is reduced due to direct impact of infection in workers. Crucially, changes in contact patterns occur not just between workers of the same sector, but also between workers and customers (of other sectors or in the community), or between members of the community (patrons of bars/restaurants etc.). The heterogeneity necessary for a dynamic macroeconomic integrated model is therefore a mapping between economic activity and human contact at the same degree of disaggregation. This mapping can work

in either direction: mandated closures yield reductions in contact, epidemiological outcomes trigger mandated closures, or perceived threat due to an outbreak yields changes in behaviour and therefore in economic activity (c.f. the interactions between policy, behaviour, FOI $\lambda(t)$ and hospital occupancy, as an indicator of threat, in Fig. 1). Sectors of particular note are: education (in which pupils/students are the customer), hospitality/entertainment and travel/hotels, as these sectors, when open, correlate strongly with contact and therefore transmission (Haw et al., 2022). As we see a wealth of data emerge including simple metrics for behavioural change (mobility Google, 2022, mask wearing Riley et al., 2021 etc.), we see a clear modelling need whereby human behaviour is an explicit component of Force-Of-Infection.

Estimates of sector-stratified contact rates in the literature are rare (Béraud et al., 2015; Danon et al., 2013), with little sector disaggregation and sparse responses in many sectors. These studies have, however, been used as the basis of many economically stratified models in the wake of COVID-19 (Janiak et al., 2021; Hill et al., 2021). In order to study the relationship between economic variables (e.g. GVA, discussed below) and contact rates, we require disaggregate data for both phenomena simultaneously, and at different points during the recent pandemic, i.e. under different regulations and different patterns of incidence/prevalence, and from a part of the world relevant to the country being modelled.

Intrinsic to sector- or workplace-stratified contact data is the proportion of workers working remotely, and corresponding changes in productivity. We must also differentiate between capacity for remote work should this be mandated, and baseline remote working rates, for which pre-COVID data will be long outdated. The latter represents a new workplace configuration that serves as the initial condition for a future outbreak, and may still not have been established in much of the world. Remote working data is therefore an ongoing need when modelling human workplace contact.

Human contact is a result of physical presence in the workplace, whereas productivity stems from active workforce. A key data need here is changes in number of workers and proportion of time spent working remotely. The dominant data need identified here is a volume of contact surveys throughout the recent pandemic. A wealth of survey data identifies only some formal sectors, such as healthcare, education and hospitality, with a focus more on the nature of the work than its interaction with rest of the economy (Riley et al., 2021). Such data in fact incentivises a modelling need, i.e. disaggregation by work type.

Assortative mixing by age is crucial to encapsulate the dynamics of airborne pathogen transmission (Keeling and Rohani, 2008). In modelling terms, when adding further heterogeneities such as economic sector, the populations of these additional compartments may be non-uniform in age. There is therefore the potential for a change in sector-specific infection rates. For example, if customer-facing roles in hospitality are dominated by younger workers who are also densely connected outside of the workplace, then we will observe a more rapid initial spread of a pathogen within this sector. In order to observe such effects in a model, we require age stratification of our populations in all other compartments.

Evidence from the COVID-19 pandemic has shown heterogeneity in infection risk and severity with respect to socio-economic status (Lo et al., 2021; Mena et al., 2021). Integrated modelling is a key tool in identifying how this relates to the workplace setting. In order to fully characterise the impact of stratified economic closures, we must explore the differential impact on different socio-economic groups. This may manifest in terms of workforce composition, household size, social contact patterns, ability to work remotely, ability to afford childcare, and access to medical care. There are two clearly identifiable missing links in current data: a mapping from labour force in formal economic sectors to information that is crucial for epidemiological models, most notably infection risk and contact patterns by occupation, and information on age, socio-economic status and employment type with respect to customer exchange. Reported contact patterns with respect to such

disaggregation throughout different stages of economic closure would be gold standard.

In the education sector, it is the pupils and students that are the consumers, and it is between these consumers that much disease transmission occurs when the sector is active. For example, in the UK, efforts in online teaching and bubbling of classes or year groups have reduced such contacts and have been somewhat dynamic in response to outbreaks within a given institution. Furthermore, school closures create a demand for childcare that may impact the workforce in other sectors. Estimated loss in labour force and GDP due to school closures has been inferred from a combination of school closure data and household survey data (Raitzer et al., 2020), though sources used predate the COVID-19 pandemic and so estimates do not account for patterns that arise due to additional closures. Refined estimates are therefore reliant on longitudinal variants of these datasets throughout the pandemic.

The calibration of integrated models to a disease outbreak requires epidemiological data — typically hospitalisations or deaths, owing to the more consistent reporting over time. Such consistency is a broader epidemiological problem and therefore is left to other articles in this issue. Moreover, though the disaggregation of the population into economic sector is an important heterogeneity for integrated epi-econ models, this heterogeneity is not necessarily directly manifest in epidemiological data resulting from changes in contact patterns: in the case of COVID-19, hospitalisations are observed in the older and more vulnerable population, for which there is little motivation for economic stratification.

We identify the following data needs for the description of heterogeneity in the interface of economics and epidemiology:

- Proportion of employees and proportion of time spent working remotely, alongside any measured changes in productivity (though the latter may be inferred from sectoral outputs)
- Longitudinal contact survey data regarding number and nature of contacts, stratified by employment status, economic sector and type of workplace, age and socio-economic status
- Childcare protocols during periods of school closure (household survey level; school openings for children of essential workers; socio-economic status of schools)
- Daily reported numbers of school attendance and absence by reason for absence (available for the UK Office for National Statistics, 2020c)
- Consistent longitudinal reporting of an epidemiological variables such as hospitalisations

3. Economic models

Models of the economy can be broadly classified into two domains: macro-economic models (describing the economy as a whole, or highly aggregated by economic sector — akin to compartmental models in epidemiology) and micro-economic models (describing individual economic agents such as consumers, workers or firms — akin to agent-based models (ABMs) or individual contact network models in epidemiology). Unlike the study of communicable disease, where SIR- or SIS-like dynamics characterised by force-of-infection (FOI) form the basis of all models, there is little consensus on the essential elements of an epi-econ model. Most studies focus on a single, closed economy, though some model transmission and/or economic exchange between heterogeneous regions of a country, or a country and its main trading partners (Deb et al., 2021; Verschuur et al., 2021).

Macro-economic models focus on aggregate economic outcomes, such as the magnitude and nature of perturbations to the economy e.g. Acemoglu et al. (2020). These macro-economic models of the economy include input–output models (IO models), general equilibrium models (GEMs), growth models and aggregate demand — aggregate

supply models (AD-AS models). Data needs for such models are one-time surveys or time series of macro-economic outcomes: gross domestic product (GDP), gross value added (GVA) by economic sector, labour force, labour productivity, capital, consumption, savings, investments, and input–output tables (IO tables, which describe supply chains and the interdependencies between economic sectors). Though there is a clear mapping between model variables and data variables in macro-economics, data sources such as IO tables are typically available only retrospectively, and with substantial temporal aggregation (periods of one year).

Micro-economic models study the behaviour of economic agents and how they change as a result of exogenous shocks e.g. Eichenbaum et al. (2022). Such models attempt to model the factors underlying changes in goods and services produced, or the changes in the supply of labour or mode of working (office working vs. remote working), changes in consumption, or changes in productivity. Micro studies may allow for heterogeneity across agents, but more commonly are formulated as Representative Agent Models (RAMs). They assume that there is a limited number of classes of agents that all behave identically (in the same class). This typically involves the modelling of preferences to describe the typical behaviour of an individual in each class, assuming homogeneity in behaviour within the class.

Some integrated models incorporate an endogenous behaviour change of individuals via a utility function that usually incorporates perceived risk or cost of infection, aside from other benefits and costs, which impacts consumption and/or labour supply (Eichenbaum et al., 2022). This association is informed by individuals' preferences and internal trade-offs between the benefits of consumption and working, and the costs of infection. Integrated epi-econ models with endogenous behaviour often incorporate risk of infection via a simple SIR-like outbreak model. Such models, however, remain theoretical due to the abstract nature of individual preferences. Data inputs include consumption and hours worked (Eichenbaum et al., 2022), though no consensus on the arguments and the functional form of utility functions is apparent in the literature. Indicators for such behavioural variables have been quantified on an international scale (Imperial College London/YouGov, 2020), though data needs for the calibration of utility functions may also depend on the choice of function in a given model.

4. GVA and epi-econ modelling

When modelling a heterogeneous economy, sector-stratified production data is advantageous over aggregate measures such as national GDP. GVA by economic sector is readily available for many countries and represents the difference between the value of output and intermediate inputs (goods and services purchased from other sectors) such that total GVA across all sectors yields GDP. A GVA of £1bn could therefore correspond to an output of £1.5bn with intermediate inputs of £500 m, or an output of £2bn with intermediate inputs of £1bn.

Economic closures affect both supply (workers are sent home, with an associated drop in productivity and economic output), and demand (there is less opportunity for consumption). This affects particularly services, or goods where market exchange cannot be online. Moreover, with knowledge of an impending lockdown, companies in a specific sector may reduce output expecting a reduced demand for consumption in order to avoid over-production. It is also feasible that the remote production of goods and services became more efficient as the pandemic progressed. For example, the initial lockdown in the UK (March 2020) showed notably lower GVA across many sectors compared to the November 2020 lockdown, despite similar mandated closures. Changes in investment are often considered negligible in early months of a pandemic, but they become increasingly important when studying long-term projections after a dramatic perturbation to the economy.

Current reporting of GVA by economic sector presents two notable immediate limitations: a delay in reporting means that the current state of the economy often cannot be used for calibration or projections, and the temporal resolution of such data is at best aggregated by month, showing only coarse dynamic trends on the timescale of a pandemic. Moreover, thresholds of time intervals align to calendar months and so do not typically map to changes in mandated economic closures. The monthly Office for National Statistics (ONS) surveys of economic activity in the UK on which GVA is based (Office for National Statistics, 2022) are regularly incomplete upon first publication and are usually updated as more data become available. This results in the need to re-calibrate models to the entire pandemic on a regular basis in order to accurately account for economic variables. The nature of GVA as an aggregate variable renders high temporal resolution problematic, though weekly or fortnightly data is not infeasible. More important is the need for nowcasts, i.e. projections that provide estimates for current economic activity, to fill the void created by reporting delay, which provide a methodological solution to this data gap. This is already available for national GDP data across the OECD countries (Woloszko, 2020) and an extension to other economic time series would be very welcome.

GVA data for each country, like much economic data, is disaggregated by economic sector. The sector stratifications are, however, not uniform between countries. A baseline 117 sector configuration, when presented, is universal, though data is typically presented in aggregate form that is suited to the economy of a given country. This is problematic when subdivisions of the reported sectors pose notably different risk of disease transmission, such as the subdivision of the retail sector into "online" and "in-person". When modelling at supranational level, whether multiple interacting open economies or simply comparison between closed economies of different countries, a simple data need involves a standardised view of a single economy, mediated by a global organisation. Even a coarse but unified description would be of immediate appeal. A significant amount of excellent work has been already done by the OECD (OECD, 2022), IMF (IMF, 2022) and World Bank (World Bank, 2022) in this space. A broader overview is provided by Kose et al. (2022).

Additional macroeconomic variables used in some integrated models include labour force (number of workers), labour supply (number of potential workers, including those furloughed and unemployed) and workforce productivity (ratio of GVA to labour force). Intuitively, reduced economic activity in a sector corresponds to a reduced labour force due to mandated closures or furloughs, together with increased remote working, and hence fewer contacts in the workplace. It may also be the case that total labour force is reduced due to actual or anticipated reductions in demand, and an associated reduction in the demand for labour. Also, the relationship between productivity and labour force might be nonlinear at the business level (Serrasqueiro and Nunes, 2008), and may differ by sector. For example, a loss of labour force may be to some extent counteracted by increased productivity in remaining workers.

The contact patterns experienced between workers is dependent on the worker population. Furthermore, the workforce requirements in a given sector in terms of skill set or training as well as overall number may change due to mandated closures. For example, the hospitality sector may stay active for home delivery from restaurants, thus requiring kitchen staff and delivery staff, but not service staff. Workforce composition at different stages of closure may be informed by labour earnings as a fraction of GVA (Office for National Statistics, 2021), though this data yields no explicit information regarding contact patterns. Crucial to understanding this relationship is the linkage of labour earnings with contact survey data discussed above.

In many low- and middle-income countries (LMICs), a considerable portion of economic activity is attributed to the "informal sector" (or the "grey economy") and the unregulated labour market, i.e. trade that is not taxed or monitored by the government, and day labourers that are effectively self-employed. As a result, such activity does not appear in official data, and is difficult to quantify empirically (Octavia, 2020; Cwiakala-Malys et al., 2020). Since it is difficult to subject the unregulated informal sector to mandated closure, reduction in transmission cannot be achieved via sending workers home. Rather, reductions in activity of the informal sector may result from reductions in demand, either as a result of mandated closures in formal sectors or behaviour change due perceived threat of infection (Adom et al., 2020). We also face the possibility of increased informal activity as activity in formal sectors is limited, though evidence shows this not to be the case (Komin et al., 2021). The case for financial support of the informal sector in LMICs is made clear in the literature (Nnabuife et al., 2020; Komin et al., 2021), though much of the data is qualitative. A quantitative approach requires not just the collection of new data, but a methodology for identifying informal supply and demand.

School closures have long been known as an effective infection control measure (Chao et al., 2010). Indeed, contact rates in children are often stratified with respect to school terms (Prem et al., 2017), and from a modelling perspective these contact rates can be seen to drive an epidemic trajectory (Haw et al., 2021). However, the sector labelled "education" often aggregates output across primary, secondary and further education, spanning an age range of 4-21+, and includes both students and staff. The variability in contact patterns across this sector, and indeed compliance to NPIs such as distancing and mask wearing, does not require explanation. Moreover, the economic value of education is usually under-estimated in national accounts because monetary valuations of educational output that relies on input prices or fees paid do not reflect the true value of education. While efforts are made in some national accounts to correct for the undervaluation, the benefits of better education that manifest over the lifetime of individuals in terms of higher incomes and better educational opportunity are difficult to estimate and fully account for (Johnson et al., 2020). Focusing on the GVA of an aggregate education sector therefore tends to underestimate the long-term cost of school closures, and an adjustment to the GVA contribution from this sector is crucial in quantifying the true economic impact of school closures. Measures of such components are available over the COVID-19 pandemic in the UK (Office for National Statistics, 2020c).

From this discussion we identify the following data needs as developments beyond current GVA reporting:

- Separation of GVA data into constituent parts: output and intermediate inputs, labour earnings, labour productivity, capital investments etc.
- Employment data (hours worked/full-time-equivalent (FTE) labour force/furlough) aligned with GVA
- GVA nowcasts, with improved temporal resolution where possible
- Optimal disaggregation of the economy that is consistent between countries
- Quantification of activity in the informal sector (greater relevance in low- and middle-income countries)
- Estimates of the true economic contribution of the education sector to the long-term welfare of an economy

5. Spatial structures

Spatial disaggregation in epidemic models is often key to replicating epidemic timing (Mills and Riley, 2014a). Moreover, localised mitigation strategies require such heterogeneity in the underlying models. The relevance of spatial heterogeneity is driven by population density and human movement patterns, which are easily quantified owing to mobility data, e.g. from mobile phone use (Grantz et al., 2020). When developing spatially disaggregate transmission models, travel data is key to accurately modelling human movement. If we wish to add a spatial component to integrated models, where transmission risk and

economic activity are causally related, then we require some spatial disaggregation of our economic variables.

Economic activity data is typically not geographically disaggregated within a given country, since national accounts are maintained at national level, and single production chains regularly involve multiple locations. Exceptions include countries with a federal structure such as the USA, where much economic data is aggregate at state level, though spatial units are typically large and do not reflect the distribution and movement of the country's population. An integrated model must therefore rely on (relatively) spatially aggregate data regarding economic activity. Including spatial heterogeneity in an integrated mathematical model requires both spatial and economic heterogeneity in the force-of-infection and hence requires an economically disaggregated description of human contact for each location. The application of such a model would be a scenario in which economic closures are geographically targeted. The resulting economic disruption would, however, be dependent on the geographical distribution of different components of supply chains within a sector. The level of heterogeneity in economic data required to support such a framework means that this particular avenue of research would certainly be complicated and in many cases would be infeasible. Mobility data over the COVID-19 pandemic may give an indication of physical presence under different configurations of economic closure, but the concept of locally evaluated GVA remains problematic at high resolution, hence so is the problem of associating costs to localised business closures.

Another challenge in integrated spatial modelling lies in the observation that people do not always work where they live: contributing to the economy of a city and to the community of a rural village is not uncommon in many western countries. Many models aim to study the role of household/workplace/community transmission, without further stratification of the nature of the workplace (Riley and Ferguson, 2006; Haw et al., 2020). Crucially such studies typically employ agent-based modelling owing to the heterogeneity that emerges, which introduces a degree of complexity and computational demand that may prove infeasible to include in immediate outbreak response. When taking a compartmental approach, it is naïve to assume uniformity of commuting behaviour across sectors, though some simple stratifications of the sector-specific workforces may suffice to encapsulate this phenomenon, for example the urban/rural distinction for home and for the workplace, alongside the aforementioned number of workers able to work remotely, and proportion of time doing so. Though it may be possible to infer commuting patterns from economic closures and travel data combined, there are additional movement patterns that correlate strongly with economic closures. Explicit commuting data would bring clarity to this research question.

A second manifestation of space in integrated modelling focuses not on distance but on type of location in which people spend time. Google mobility data (Google, 2022) reports percentage change in visits to locations stratified as follows: "retail/recreation", "grocery/pharmacy", "parks", "transit stations", "workplace" and "residential". Much of the interest in the relationship between economics and epidemiology lies in the categories "retail/recreation" and "workplace", for which no further disaggregation is available. Furthermore, transaction data collected by banks and credit card companies often contain limited geographical information as a physical location for the transaction is broadly reported but it is attributed to the location of the headquarters of a company rather than the location of the transaction itself (Yodlee, 2022).

We have highlighted the difficulty of spatially disaggregate economic modelling. We propose the following possible avenues for data sourcing in response to some specific research questions:

- Availability of satellite image data/Geolocation data to assess contact patterns and economic activity by location (e.g. retail parks) (Lott, 2021)

- Understanding the transmission risk associated with commuting: decomposition of workforce and workplace data by urban/rural; survey data including commute distances, times and travel means (to inform agent-based models); remote working data (as previously discussed).
- Modelling to inform targeted closures of specific premises: data should focus on the nature of a location associated with economic activity, rather than any manifestation of geographical distance, and on the presence of such locations in different geographical areas.
- Mitigating animal disease outbreaks: this is an obvious exception to the cases discussed above, whereby we would aim construct a geographically explicit model of the farming supply chain.

6. The physical nature of a transaction

Central to an integrated epi-econ model is a description of the relationship between transactions (part of economic activity) and transmission potential (epidemiology). The latter is dependent on the nature of contact between individuals: physical proximity, duration, presence of a screen/mask, physical contact, mutual touching of equipment, online payment, delivery vs. collection of goods etc. Data regarding individual transactions (contributing to retail sales data (ONS, 2022)) alludes to this by reporting physical presence or absence of a payment card. The limitations of this variable are as follows: a distinction is made between a contactless payment using a card and contactless mobile payment, but no distinction is made between contactless mobile payment and online payment. Rich datasets with transaction data, data on economic linkages, supply chains and financial transaction of physical assets exist (Yodlee, 2022) but tend to be associated with high subscription fees.

Though human mobility data and transaction data is abundant, they are reported independently of one another and hence any relationship can only be inferred. If we wish to quantify risk aversion for application to mechanistic modelling, then we require data displaying the physical nature of transactions. We acknowledge the possibility of linking movement data with transaction data, though each is highly aggregate in different dimensions. Heterogeneity in the physical nature of a transaction results in heterogeneity in contact rates and risk of transmission given contact. Furthermore, if certain sector closures correlate more strongly with transactions involving physical presence, then economic activity in such sectors may serve as a predictor for change in epidemiologically relevant contact patterns, which is the essence of integrated modelling.

We illustrate this by contrasting the COVID-19 pandemic with a Norovirus outbreak: the former is typically transmitted by droplets and is heavily dose dependent, whereas the latter is typically transmitted via contaminated surfaces. The role of Perspex screens and contactless payment therefore differ greatly between the two cases. For reasons of data confidentiality, it is infeasible to expect individual transaction data, though we propose aggregation by physical location such that the epidemiological distinction between different transaction environments is retained.

We propose the following candidate data needs, where some historic data may already exist privately:

- total spend in the retail/hospitality/entertainment/transport sector, disaggregated by physical nature of transaction.
- Proxy variables for the latter may include the following: proportion of transactions made indoors; online payment and collection/delivery; NPIs in place on site.
- Crucial to understanding the physical nature of a transaction is a longitudinal component, displaying the change in physical nature of a transaction at different stages of an outbreak (ideally including the "no outbreak" scenario).
- Availability of commercial datasets to academic institutions free for research purposes

7. Understanding behavioural response

Many governments have responded throughout the recent pandemic with mandated limitations of business/economic sectors considered non-essential. Such mandates impose a limit on economic activity. The behavioural response of consumers may further reduce demand for goods and services, and the anticipation of businesses the demand for labour. For example, if restaurants are subject to a mandate of table service only, there will be reduced demand for dining (upper limit on covers), with a knock-on reduction in demand for intermediate goods and services (sourcing fewer ingredients), and employment (reduced staffing). However, perceived risk of infection by consumers may result in even greater reduction in demand. The combined effect of the mandate and behavioural effects will manifest via reduced GVA for that sector. Moreover, the behavioural component may respond to broader range of phenomena than the mandate alone: the perceived threat may depend on knowledge of prevalence or hospital occupancy, rhetoric promoted in the media, or "behavioural fatigue" (Harvey, 2020; Mills and Riley, 2014b). Such phenomena are difficult to quantify directly due to their subjective nature, but they are manifest in metrics such as mobility, reported contacts and financial transactions. If we wish to derive mathematical description of behaviour from survey data obtained throughout the course of the COVID-19 pandemic then analysis of a large volume of these less subjective data is a good place to start, and can inform more carefully articulated questions for future surveys.

GVA and employment alone show only the overall output, but do not help to understand which factors impact on the supply and demand-side drivers that contribute to this output. We therefore propose the following: a combination of contact survey and business activity data in hospitality venues. Crucially, observations under different sets of mandated restrictions (and necessarily different concurrent levels of prevalence/hospital occupancy) would allow for the quantification of change in contact patterns throughout the pandemic. A large volume of such data would allow modellers to explore a functional relationship between state variables and behavioural response. Such data does exist for the UK (Riley et al., 2021; Imperial College London/YouGov, 2020), though much retrospective analysis is required to identify the power of these surveys, and to identify specific modifications required in the light of research questions and modelling studies that have emerged since the conception of early COVID-19 questionnaires.

8. How to proceed

Much of the available economic data regarding changes in activity throughout the COVID-19 pandemic is aggregate: in space (by country), in time (discretely by month at best) and in the output of economic goods and service under consideration. As a result, often only the aggregate outcome of economic phenomena is manifest. Disaggregation of transactions, timing and output is often needed to understand the relationship between FOI, economic activity and perceived threat. Disaggregation of the education sector into primary/secondary/further, and localised data of outbreaks and closures in schools, would allow us to refine our mechanistic description of mitigation strategies with respect to this sector.

Where large datasets do exist regarding transactions, mobility and contact surveys, there is an opportunity for inferring parameter values that describe the interface between economics and epidemiology. Moving forward, however, explicit coupling of transaction data with the physical nature of a transaction would both confirm (or deny) and further refine estimates from more indirect methods.

Social contact surveys are commonplace in epidemiology. However, they typically do not relay direct information regarding employment sector. "Customer facing" and "healthcare/patient facing" are standard practice (Riley et al., 2021), though there remains substantial heterogeneity in the nature of a contact within these categories.

It is important to acknowledge the difference between a descriptive study of the relationship between economics and epidemiology throughout the COVID-19 pandemic, and the development of predictive models that require assumptions describing this relationship. With better calibrated fixed parameters, our theoretical integrated econ-epi models would gain predictive power, though in the early stages of this interdisciplinary field we are far from consensus on how such models should function. It is therefore somewhat early to describe the full data needs in the health economics of infectious diseases.

Among the data needs we have identified are many that will never exist for past outbreaks, meaning that retrospective studies will necessarily require a greater armoury of inference methods in order to extract information about the interface of economics and epidemiology. Moreover, the baseline scenario of "no pandemic" has changed as a result of the last two years, so that a more detailed description of the pre-COVID world is beyond our reach.

We have focused on the data needs for models in which economic mandates impact the dynamics of transmission. The converse, in which economic activity is dependent on epidemiological variables, poses the additional difficulty of quantifying behavioural response. Models must also distinguish between intervention measures that reduce some economic activity (e.g. enforced business closures, limiting capacity in public spaces), and those that do not (e.g. mask wearing, Perspex screens). The latter point is relevant in broader epidemiological studies and so we do not address it directly here.

Prior to the emergence of COVID-19, it was widely expected that the next respiratory pandemic to hit the globe would be influenza A. Proven wrong by a pathogen with double the severity, we are reminded that the nature of future outbreaks is a known unknown, and that not all disease outbreaks require the same data. The discussion of data needs for epidemic models alone is therefore ongoing, and data for integrated models will depend on the way in which any newly emerging pathogen impacts the economy. We acknowledge that our discussion is centred around airborne pathogen transmission, where physical proximity can drive transmission. We propose that a more general indicator of transmission potential must differentiate between different types of contact reported, e.g. handshake vs. shared enclosed space.

It is perhaps ambitious to envisage a mathematical model that is fully dynamic in its description of both the economy and infectious disease transmission, calibrated in real time to metrics for both phenomena. But it is perhaps also unnecessary. As modellers explore the theoretical space of integrated epi-econ models, we hope to identify precise requirements of disaggregation in data that sufficiently informs our description of the interface between economics and epidemiology, whilst remaining feasible to report in the long term.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

References

Acemoglu, D., Chernozhukov, V., Werning, I., Whinston, M.D., 2020. A multi-risk SIR model with optimally targeted lockdown.

Adom, Dickson, Adu-Mensah, Jephtar, Appiah, Sekyere, 2020. Hand-to-mouth work culture and the COVID-19 lockdown restrictions: experiences of selected informal sector workers in Kumasi, Ghana. 1, pp. 45–63, URL: https://royalliteglobal.com/advanced-humanities/article/view/237.

Australian Bureau of Statistics, 2022. Australian system of national accounts. URL: https://www.abs.gov.au/statistics/economy/national-accounts/australian-system-national-accounts/latest-release.

Badan Pusat Statistik, 2022. PDB seri 2010 (milyar rupiah), 2022. URL: https://www.bps.go.id/indicator/11/65/1/-2010-version-gdp-at-2010-version.html.

Béraud, Guillaume, Kazmerczak, Sabine, Beutels, Philippe, Levy-Bruhl, Daniel, Lenne, Xavier, Mielcarek, Nathalie, Yazdanpanah, Yazdan, Bolle, Pierre-Yves, Hens, Niel, Dervaux, Benoit, 2015. The french connection: the first large population-based contact survey in france relevant for the spread of infectious diseases. PLoS One 10 (7), 1–22. http://dx.doi.org/10.1371/journal.pone.0133203.

Brodeur, Abel, Gray, David, Islam, Anik, Bhuiyan, Suraiya, 2021. A literature review of the economics of COVID-19. J. Econ. Surv. 35 (4), 1007–1044. http://dx.doi.org/10.1111/joes.12423, URL: http://dx.doi.org/10.1111/joes.12423.

Çakmaklı, C., Demiralp, S., Kalemli-Özcan, S., Yesiltas, S., Yildirim, M.A., 2020. COVID-19 and emerging markets: A SIR model, demand shocks and capital flows.

Chao, Dennis L., Halloran, M. Elizabeth, Longini, Jr., Ira M., 2010. School opening dates predict pandemic influenza A(H1N1) outbreaks in the United States. J. Infect. Dis. (ISSN: 0022-1899) 202 (6), 877–880. http://dx.doi.org/10.1086/655810, eprint: https://academic.oup.com/jid/article-pdf/202/6/877/18062872/202-6-877.pdf.

Cwiakala-Malys, Anna, Piotrowska, Iwona, Durbajlo-Mrowiec, Małgorzata, Lagowski, Paweł, 2020. Measurement of grey economy and system ways of limiting it – polish example. Eur. Res. Stud. J. XXIII (Special 1), 997–1023, URL: https://EconPapers.repec.org/RePEc:ers:journl:v:xxiii:y:2020:i:special1:p:997-1023.

Danon, Leon, Read, Jonathan M., House, Thomas A., Vernon, Matthew C., Keeling, Matt J., 2013. Social encounter networks: characterizing Great Britain. Proc. R. Soc. B: Biol. Sci. 280.

Deb, Pragyan, Furceri, Davide, Ostry, Jonathan D., Tawk, Nour, 2021. The economic effects of COVID-19 containment measures. Open Econ. Rev. 33 (1), 1–32. http://dx.doi.org/10.1007/s11079-021-09638-2.

Eichenbaum, Martin S., Rebelo, Sergio, Trabandt, Mathias, 2022. The macroeconomics of testing and quarantining. J. Econom. Dynam. Control (ISSN: 0165-1889) 104337. http://dx.doi.org/10.1016/j.jedc.2022.104337, URL: https://www.sciencedirect.com/science/article/pii/S0165188922000422.

Gog, Julia R., Ballesteros, Sébastien, Viboud, Cecile G., Simonsen, Lone, Bjørnstad, Ottar N., Shaman, Jeffrey L., Chao, Dennis L., Khan, Farida, Grenfell, Bryan T., 2014. Spatial transmission of 2009 pandemic influenza in the US. PLoS Comput. Biol. 10, http://dx.doi.org/10.1371/journal.pcbi.1003635.

Google, 2022. COVID-19 community mobility reports. https://www.google.com/covid19/mobility/.

GOVschools, 2022. Attendance in education and early years settings during the coronavirus (COVID-19) pandemic. URL: https://explore-education-statistics.service.gov.uk/find-statistics/attendance-in-education-and-early-years-settings-during-the-coronavirus-covid-19-outbreak.

Grantz, Kyra H., Meredith, Hannah R., Cummings, Derek A.T., Metcalf, C. Jessica E., Grenfell, Bryan T., Giles, John R., Mehta, Shruti, Solomon, Sunil, Labrique, Alain, Kishore, Nishant, Buckee, Caroline O., Wesolowski, Amy, 2020. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. Nature Commun. 11 (4961).

Hale, Thomas, Angrist, Noam, Goldszmidt, Rafael, Kira, Beatriz, Petherick, Anna, Phillips, Toby, Webster, Samuel, Cameron-Blake, Emily, Hallas, Laura, Majumdar, Saptarshi, Tatlow, Helen, 2021. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat. Hum. Behav. https://www.nature.com/articles/s41562-021-01079-8.

Harvey, Nigel, 2020. Behavioral fatigue: real phenomenon, naïve construct, or policy contrivance? Front. Psychol. (ISSN: 1664-1078) 11, http://dx.doi.org/10.3389/fpsyg.2020.589892, URL: https://www.frontiersin.org/article/10.3389/fpsyg.2020.589892.

Haw, David J., Biggerstaff, Matthew, Prasad, Pragati, Walker, Joseph, Grenfell, Bryan, Arinaminpathy, Nimalan, 2021. Using real-time data to guide decision-making during an influenza pandemic: a modelling analysis. http://dx.doi.org/10.1101/2021.06.09.21258618, medRxiv, eprint:https://www.medrxiv.org/content/early/2021/06/12/2021.06.09.21258618.

Haw, David J., Cummings, Derek A.T., Lessler, Justin, Salje, Henrik, Read, Jonathan M., Riley, Steven, 2019. Differential mobility and local variation in infection attack rate. PLoS Comput. Biol. 15 (1), 1–16. http://dx.doi.org/10.1371/journal.pcbi.1006600.

Haw, David J., et al., 2020. Strong spatial embedding of social networks generates nonstandard epidemic dynamics independent of degree distribution and clustering. Proc. Natl. Acad. Sci. 117 (38), 23636–23642. http://dx.doi.org/10.1073/pnas.1910181117, eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1910181117.

Haw, D.J., et al., 2022. Optimizing social and economic activity while containing SARS-CoV-2 transmission using DAEDALUS. Nat. Comput. Sci. 2, http://dx.doi.org/10.1038/s43588-022-00233-0.

Hill, Edward M., Atkins, Benjamin D., Keeling, Matt J., Dyson, Louise, Tildesley, Michael J., 2021. A network modelling approach to assess non-pharmaceutical disease controls in a worker population: An application to SARS-CoV-2. PLoS Comput. Biol. 17 (6), 1–24. http://dx.doi.org/10.1371/journal.pcbi.1009058.

IMF, 2022. IMF data portal. URL: https://data.imf.org/?sk=388dfa60-1d26-4ade-b505-a05a558d9a42.

Imperial College London/YouGov, 2020. COVID-19 behaviour tracker. URL: https://www.imperial.ac.uk/global-health-innovation/what-we-do/our-response-to-covid-19/covid-19-behaviour-tracker/.

Janiak, Alexandre, Machado, Caio, Turén, Javier, 2021. Covid-19 contagion, economic activity and business reopening protocols. J. Econ. Behav. Organ. (ISSN: 0167-2681) 182, 264–284. http://dx.doi.org/10.1016/j.jebo.2020.12.016, URL: https://www.sciencedirect.com/science/article/pii/S0167268120304753.

Johnson, R., et al., 2020. Valuing Lives, Education and the Economy in an Epidemic: Societal Benefit of SARS-CoV-2 Booster Vaccinations in Indonesia. Tech. Rep., Imperial College London, URL: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-51-indonesia/.

Keeling, M.J., Rohani, P., 2008. Modeling Infectious Diseases in Humans and Animals. Princeton University Press, ISBN: 9780691116174, URL: https://books.google.co.uk/books?id=G8enmS23c6YC.

Komin, Wichaya, Thepparp, Rungnapa, Subsing, Borvorn, Engstrom, David, 2021. Covid-19 and its impact on informal sector workers: a case study of Thailand. Asia Pac. J. Soc. Work Dev. 31 (1–2), 80–88. http://dx.doi.org/10.1080/02185385.2020.1832564, eprint.

Kose, M. Ayhan, Kurlat, Sergio, Ohnsorge, Franziska, Sugawara, Naotaka, 2022. A cross-country database of fiscal space. J. Int. Money Finance (ISSN: 0261-5606) 128, 102682. http://dx.doi.org/10.1016/j.jimonfin.2022.102682, URL: https://www.sciencedirect.com/science/article/pii/S0261560622000857.

Lo, Chun-Han, et al., 2021. Race, ethnicity, community-level socioeconomic factors, and risk of COVID-19 in the United States and the United Kingdom. 38.

Lott, Charlie, 2021. Assessment of retail parking lots using remote sensing. URL: https://storymaps.arcgis.com/stories/0c5acd83794b4a2fad13255eb1c63b47.

Mena, Gonzalo E., et al., 2021. Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. Science 372 (6545), http://dx.doi.org/10.1126/science.abg5298, eabg5298, eprint: https://www.science.org/doi/pdf/10.1126/science.abg5298.

Mills, Harriet L., Riley, Steven, 2014a. The spatial resolution of epidemic peaks. PLoS Comput. Biol. 10 (4), 1–9. http://dx.doi.org/10.1371/journal.pcbi.1003561.

Mills, Harriet L., Riley, Steven, 2014b. The spatial resolution of epidemic peaks. PLoS Comput. Biol. 10 (4), 1–9. http://dx.doi.org/10.1371/journal.pcbi.1003561.

Mossong, Joël, Hens, Niel, Jit, Mark, Beutels, Philippe, Auranen, Kari, Mikolajczyk, Rafael, Massari, Marco, Salmaso, Stefania, Tomba, Gianpaolo Scalia, Wallinga, Jacco, Heijne, Janneke, Sadkowska-Todys, Malgorzata, Rosinska, Magdalena, Edmunds, W. John, 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 5 (3), 1. http://dx.doi.org/10.1371/journal.pmed.0050074.

Nnabuife, Ezimma K., Okoli, Ifeanyi E., Clara, Anugwu Chika, 2020. Informal sector and Nigerian economic prospects: the covid–19 experience. Eur. J. Bus. Manag. Res. 5 (4).

Octavia, Joanna, 2020. Towards a national database of the informal sector: pandemic response and future recommendations for Indonesia. https://warwick.ac.uk/fac/soc/ier/news/?newsItem=8a1785d772941b9901729ec4277c58e4.

OECD, 2022. Catalogue of OECD databases. Data across 38 member countries. URL: https://data.oecd.org/searchresults/?r=+f/type/datasets.

Office for National Statistics, 2020a. COVID-19 infection survey (CIS). URL: https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/covid19infectionsurveycis.

Office for National Statistics, 2020b. Labour force survey (LFS). URL: https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/labourforcesurvey.

Office for National Statistics, 2020c. Schools infection survey (SIS). URL: https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/covid19schoolsinfectionsurveyis.

Office for National Statistics, 2020d. Survey on living conditionsy (SLC). URL: https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/surveyonlivingconditions.

Office for National Statistics, 2021. Labour costs and labour income, UK. URL: https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/covid19infectionsurveycis.

Office for National Statistics, 2022. Gross value added (GVA). URL: https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/monthlygrossdomesticproductbygrossvalueadded.

ONS, Retail sales, Great Britain: September 2022, URL: https://www.ons.gov.uk/businessindustryandtrade/retailindustry/bulletins/retailsales/september2022.

Pichler, A., Pangallo, M., del Rio-Chanona, R.M., Lafond, F., Farmer, J.D., 2020. Production networks and epidemic spreading: How to restart the UK economy?

Prem, Kiesha, Cook, Alex R., Jit, Mark, 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput. Biol. 13 (9), 1–21. http://dx.doi.org/10.1371/journal.pcbi.1005697.

Raitzer, David A., Lavado, Rouselle F., Rabajante, Jomar, Javier, Xylee, Garces, Ludigil, Amoranto, Glenita, 2020. Cost–Benefit Analysis of Face-to-Face Closure of Schools to Control COVID-19 in the Philippines. In: ADB Briefs. (162), ISBN: 9789292572327, http://dx.doi.org/10.1111/apa.15371.7, URL: https://www.adb.org/sites/default/files/publication/662321/adb-brief-162-cost-benefit-analysis-closure-schools-philippines.pdf.

Riley, S, Atchison, C, Ashby, D, Donnelly, CA, Barclay, WS, Cooke, GS, Ward, H, Darzi, A, Elliott, P, null, null, 2021. Real-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: Study protocol [version 2; peer review: 2 approved]. Wellcome Open Res. 5 (200), http://dx.doi.org/10.12688/wellcomeopenres.16228.2.

Riley, Steven, Ferguson, Neil M., 2006. Smallpox transmission and control: Spatial dynamics in Great Britain. Proc. Natl. Acad. Sci. 103 (33), 12637–12642. http://dx.doi.org/10.1073/pnas.0510873103, eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.0510873103.

Ritchie, Hannah, Mathieu, Edouard, Rodés-Guirao, Lucas, Appel, Cameron, Giattino, Charlie, Ortiz-Ospina, Esteban, Hasell, Joe, Macdonald, Bobbie, Beltekian, Diana, Roser, Max, 2020. Coronavirus pandemic (COVID-19). https://ourworldindata.org/coronavirus.

Serrasqueiro, Zélia Silva, Nunes, Paulo Maçãs, 2008. Performance and size: empirical evidence from Portuguese SMEs. Small Bus. Econ. 31, 195–217. http://dx.doi.org/10.1007/s11187-007-9092-8, URL: https://link.springer.com/article/10.1007/s11187-007-9092-8.

Verschuur, Jasper, Koks, Elco E., Hall, Jim W., 2021. Global economic impacts of COVID-19 lockdown measures stand out in high-frequency shipping data. PLoS One 16 (4), e0248818. http://dx.doi.org/10.1371/journal.pone.0248818.

Woloszko, N., 2020. Tracking Activity in Real Time with Google Trends. Tech. Rep., http://dx.doi.org/10.1787/18151973.

World Bank, 2022. World Bank Open Data - Free and open access to global development data. URL: https://data.worldbank.org/.

Yodlee, Inc., 2022. ENVESTNET yodlee investor data analytics. URL: https://www.yodlee.com/data-analytics/investor-data.