REVSTAT – Statistical Journal
Volume 18, Number 5, October 2020, 553–564

STATISTICS IN TIMES OF PANDEMICS: THE ROLE OF STATISTICAL AND EPIDEMIOLOGICAL METHODS DURING THE COVID-19 EMERGENCY

(*Invited Paper with Discussion*)

Authors: Baltazar Nunes
– Dep. Epidemiology, Instituto Nacional de Saúde Dr. Ricardo Jorge, and
Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública,
Universidade NOVA de Lisboa,
Portugal
baltazar.nunes@insa.min-saude.pt

Constantino Caetano
– Dep. Epidemiology, Instituto Nacional de Saúde Dr. Ricardo Jorge,
Portugal
constantino.caetano@insa-min-saude.pt

Liliana Antunes
– Dep. Epidemiology, Instituto Nacional de Saúde Dr. Ricardo Jorge,
Portugal
liliana.antunes@insa-min-saude.pt

Carlos Dias
– Dep. Epidemiology, Instituto Nacional de Saúde Dr. Ricardo Jorge, and
Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública,
Universidade NOVA de Lisboa,
Portugal
carlos.dias@insa-min-saude.pt

Received: September 2020 Revised: October 2020 Accepted: October 2020

Abstract:

- Statistical and epidemiological methods play an essential role in producing information for the public health decision process. They allow the collection, analysis, reporting and interpretation of data necessary to inform public health officials in the decision-making process, enabling the diagnosis of the situation, the selection of the most adequate measures and to monitor and evaluate their impact.

Key-Words:

- *epidemiology; SARS-CoV-2; COVID-19; epidemiological surveillance; mathematical modelling.*

AMS Subject Classification:

- 92B15, 92D30.

1. INTRODUCTION

After the identification of the first cases of COVID-19 in December 2019 in Wuhan, China, it took one month (30 January 2020) for the World Health Organization to declare the epidemic a Public Health Emergency of Global concern, and almost two months to be declared a pandemic (11 March 2020) ([11]).

The start of the rampant community transmission of the SARS-CoV-2 virus in Italy and Spain in early February ([11]), which meant that most new cases did not have an epidemiological link, set a decisive date for public health action in Portugal. The first case of COVID-19 was diagnosed in Portugal on the 2nd of March 2020, and it was only a few weeks before community spread of the virus started in Portugal, which brought urgency to the implementation of population-base public health interventions.

Although the infection and disease occurs at the individual level, the epidemic occurs at the population level. This means that the control and mitigation of the infection or disease occurrence cannot be only achieved through individual targeted measures. Population level intervention is crucial to control the spread of the infection and to mitigate the impact of the disease. Among these, one can list vaccination, social distancing, respiratory hygiene, closing of schools, working from home, closing of commerce, restaurants and bars, or even more severe lockdowns, which involve, in addition to the previous measures, stay at home policies.

Public health authorities need information almost on a real time basis to be able to track the infection evolution among the population, the potential effect of interventions and their impact. The information provided can range from descriptive analysis of the distribution of the number of new cases of infection in time and space, population or individual characteristics (age, sex, education, health status), to projections of the infection spread and impact according to different epidemiological scenarios, or control and mitigation strategies.

Throughout an epidemic it is usual for certain geographical locations to be affected differently, probably due to factors that promote the spread of infection, such as population density and connectivity, the inbound and outbound of infected individuals from the region, amongst others. Which means that nowadays, the access to accurate and detailed information is paramount to mount proper public health measures, which can be targeted during a certain period of time, in a certain geographical area or population group.

In the context of a pandemic there are three areas where statistical and epidemiological methods are crucial, namely: epidemiological surveillance, specific epidemiological observational studies aimed at describing or measuring epidemiological parameters and the development of mathematical models, which focus on simulating the disease transmission dynamics and assess impact scenarios under different control and mitigation measures.

2. EPIDEMIOLOGICAL SURVEILLANCE

The epidemiological surveillance is defined as the systematic collection, analysis, interpretation and reporting of data for public health action. During the COVID-19 public health emergency, most of the epidemiological surveillance data comes from the case notification of SARS-CoV-2 infection. Individuals with a laboratory diagnosis of SARS-CoV-2 infection, with or without the presentation of COVID-19 symptoms at diagnosis, are notified and registered in a database. In Portugal, data is collected through the SINAVE system (Sistema Nacional de Vigilância Epidemiológica) ([28]), where medical doctors or laboratories, authorized to perform the SARS-CoV-2 test, report new cases of infection or disease on a daily basis. After the case notification, public health officials perform an epidemiological enquire where data is collected on individual characteristics, disease characteristics (symptoms), probable routes of infection, date of symptom onset, laboratory diagnosis date and notification date, travel history, and close contacts.

Among the first questions that arise, and that can be answered with epidemiological surveillance data, are about the distribution of cases in the time scale, namely: what is the recent incidence of the disease and its trend? Are the number of new cases of the disease increasing, decreasing or stable?

A simple but essential way to study the course of the epidemic is by plotting the epidemic curve. This graph basically depicts the number of new cases by date of disease onset.

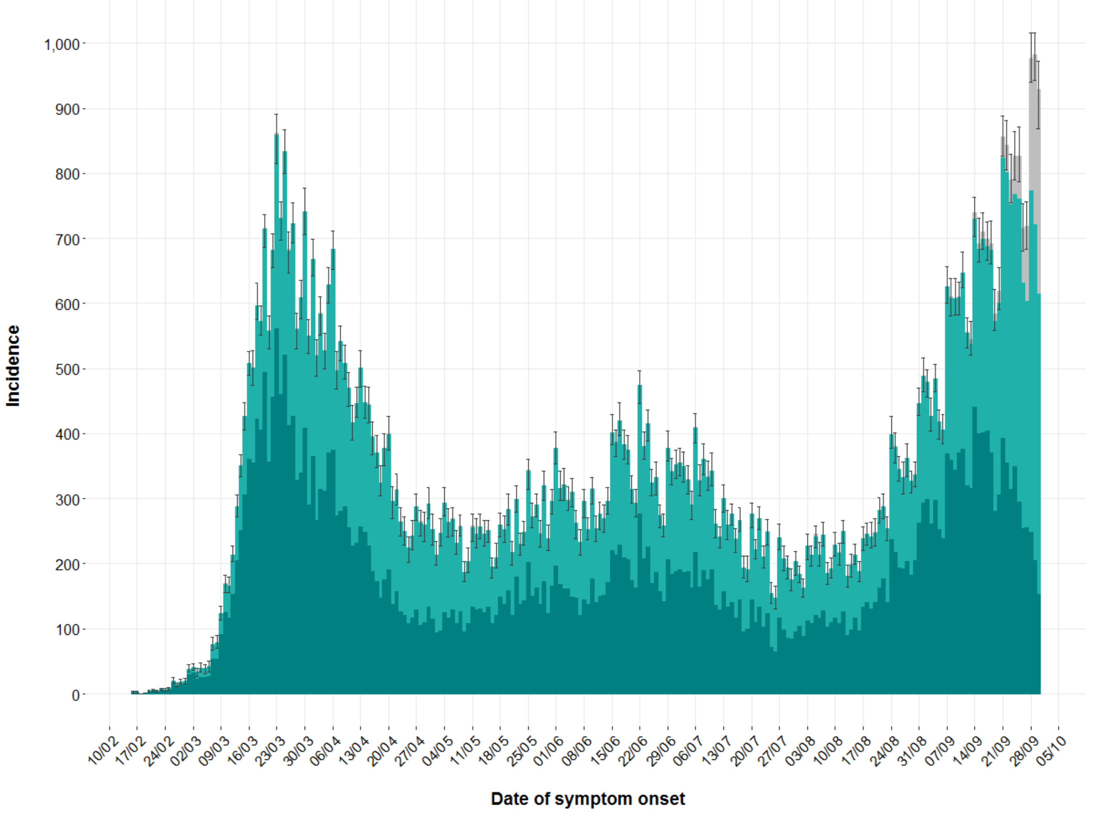

Figure 1: SARS-CoV-2 epidemic curve by date of symptom onset corrected for notification delay (Dark blue – Observed cases with date of symptom onset; Light blue – Observed cases with imputed date of symptom onset; Grey – Occurred but not yet reported cases (nowcasted)).

In actuality, the graph should be represented by date of infection, but for the majority of the cases this information is not available. The analysis of this figure is of utmost importance, given that allows the epidemiologist and public health officials to analyse the evolution of the infection transmission, its trends, growing and decreasing phases, the impact of public health measures or changes in the epidemiology of the infection.

The classical presentation of an epidemic curve is divided in four phases, the establishment phase, with a sporadic number of new cases, the exponential growing phase, the peak and the decreasing phase. But in reality, mainly with pandemics like the one we are facing with SARS-CoV-2 virus, there is a diversity in the presentation of the epidemic curve. As Adam Kurcharski, from London School of Hygiene and Tropical Medicine, referred in his recent book, The Rules of Contagion ([22]), "If you have seen one pandemic, you have seen one pandemic.". In reality, the form of the epidemic curve can present several growing, decreasing and endemic phases, depending on the presence of several factors, including public health measures, population behaviour, population susceptibility, virus changes or climate effects.

The quality of the surveillance data used in the production and analysis of an epidemic curve in real time can be affected by several factors. Surveillance data is collected hastily, since the priority of medical personnel is to isolate the cases and quarantine their contacts, in order to be able to interrupt the chains of transmission. So, collecting perfect data in these circumstances is very difficult, and usually, data quality, completeness and timeliness are substandard during pandemic emergencies. This problem is further exacerbated during the epidemic growth and epidemic peak phases, which is when this information is mostly needed.

As mentioned above, in order to draw a proper and useful epidemic curve the date of disease onset for each case is essential. Unfortunately, during an epidemic this is one of the variables that suffers more from incompleteness. Moreover, there is the need to draw conclusions on the epidemic course for the last few days or weeks recorded. As can be easily understood, before an infected individual is accounted for in the surveillance system, several other events must occur, namely: disease onset, which is when the individual starts developing symptoms; medical visit, which is usually the time it takes an individual to seek health care after developing symptoms; diagnosis, when the infection has laboratory confirmation; and ultimately, notification, which is when the case is introduced in the surveillance system database.

Naturally, from disease onset to the notification of the case to the surveillance system, several days can go by. Not only due to the disease and reporting process described above but also due to other factors, such as the awareness of the population, preparedness of the surveillance system, overload of the professionals involved in the reporting process during the growth or peak phases, weekends when most medical sites are closed, personal and regional characteristics, among others ([15]).

The time between disease onset and case notification, precludes the monitoring of the epidemic in real time, translated in epidemic curves showing the last days/weeks always with fewer cases than the day before, suggesting a declining trend in incidence, when in fact, as more data is available, it might actually show it was increasing.

The notification delay varies across different surveillance systems because of the intrinsic nature of the reporting process implemented for each disease and country, and also due to changes in time. During the COVID-19 emergency, the time between disease onset to notification delays have been described as 8 days on average, also 95% of the cases with onset of symptoms in one specific day take 15 days to be notified ([30, 12]). In Portugal, the median time between disease onset and notification decreased from 8 to 4 days throughout the course of the epidemic.

Here the statistical and epidemiological methods can be used to our advantage ([23]). Imputation methods can be used to estimate missing dates of disease onset, and nowcasting techniques can be applied to estimate the occurred-but-not-yet-reported cases in the last weeks of the epidemic curve ([23, 25, 19]). In order to be applied, both these methods need the distribution of the diagnosis and/or notification delay.

Several approaches have been proposed to impute date of symptom onset and to nowcast the recent days in public health setting ([15]). Nowcasting methods have also been developed in both frequentist and Bayesian framework ([19]).

There are several constraints that difficult the real time implementation of nowcast procedures such as the distribution of the delay from disease onset to notification not being invariant in time, place and individual characteristics. During the first weeks of the epidemic, the delay between disease onset and notification is usually longer, given that the first traced cases are usually detected at a later date. This means that several weeks of epidemic data are necessary in order to be able to estimate a complete distribution of the delay. So, for the first weeks of the epidemic, the delay distribution is truncated, and we need to recur to other sources of information, like historical data from other epidemics in order to estimate the notification delay distribution and nowcast the most recent days of the epidemic curve.

Besides the level of incidence and its distribution in time, public health officials usually need to know the level of transmission of the infection in the community. The reproduction number (R) of the infection is the measure used to monitor the transmission of infectious diseases. During the COVID-19 emergency it has been given a very important role in decision making, such as being used to implement or lift lockdown measures. This index measures the average number of secondary cases a typical infectious case gives origin to. When R is above one it means that the epidemic is in a growing phase, and each generation of infected individuals, gives rise to a new generation with a higher number of infected individuals. Otherwise if R is below one, the next generation of infected individuals is smaller than the previous one, which means that the size of the epidemic converges to 0, leading to its extinction. If R is equal to or close to 1, then the number of new infected individuals in each subsequent generation is approximately constant, which corresponds to an endemic or stable phase of disease spread.

It is important to estimate the reproduction number during the first phase of a pandemic, when all the population is susceptible and there are no control measures in place, in order to have a natural measure of the infection transmissibility in the population in study. This parameter is called R_0 ("R naught") and has a very crucial role in infectious disease epidemiology. It can be interpreted as the average number of infectious cases resulting from one infectious case, after its introduction in a completely susceptible population ([31]).

R_0 is used in mathematical modelling studies to set up the base scenario of an epidemic, allowing to find the level of public health measures needed to bring it below one. Or it can be used as reference to be compared with the effective reproduction number in order to evaluate changes in the infection transmission, due to the implementation of public health measures or due to the natural course of the epidemic.

R_0 can be estimated from transmission chains data, which are usually difficult to obtain, measuring the distribution of the secondary cases each initial case gave origin to. It can also be estimated from the epidemic curve growth in the initial phase of the epidemic, with additional information on the generation time distribution of the infection ([31]).

The generation time of infection corresponds to the time elapsed between the times of infection of infector and infectee individuals. Given that it is difficult to collect data on infection dates, this distribution is substituted by the serial time distribution, which is given by the time interval between symptom onset of the infector and the symptom onset of the infectee. Regarding the COVID-19 infection, the serial time is being described with a mean from 3.1 to 7.5 according to the study published ([2]). This distribution plays also a very relevant role in the epidemic transmission, given that it represents the time elapsed between the disease onset in two generations of infected individuals. The shorter the serial time, the quicker the infection spread.

Besides the R_0, another reproduction number that has been recurrently used by several public health agencies and governments to monitor the SARS-CoV-2 transmission, is the time dependent effective reproduction number. Firstly proposed by Wallinga and Teunis in 2004 ([32]), to measure the average number of secondary cases originated by the incident infectious cases in time t, it has been showed to be an excellent tool to monitor transmission over time and evaluate in a real-time manner the impact of public health measures and changes in the infection epidemiology. Nevertheless, for surveillance proposes, where date-of-onset data is always incomplete in the last few recorded weeks, several researchers have recommended the use of Cori *et al.* method ([8]) to estimate the time dependent effective reproduction number. In this method the R_t represents the average number of secondary cases that originated the new cases observed at time t.

As stated by other researchers, today's challenge is how to include the imputation and nowcasting methods uncertainty into the estimate of R_t.

In Portugal the R_t has been estimated from the epidemic curve after imputation and nowcasting ([20]), using a method developed by Antunes *et al.* in 2014 ([3]) for daily mortality monitoring. The R_t has been estimated using the Cori *et al.* method ([8]). This analysis did not include a method for error propagation from the imputation and nowcasting to the estimate of the reproduction number.

3. SPECIAL STUDIES

Special studies are specifically designed to measure epidemiological parameters of infection or disease and to identify risk factors of infection or disease severity, including the need for critical care and case-fatality.

One of the most relevant studies that the WHO recommends, at the start of the emergency of a new infectious agent with pandemic potential, is the First Few X Cases study (FFX) ([29]). This study should be implemented as quickly as the first cases are identified, and is aimed at estimating key epidemiological indicators that are essential both for surveillance and for modelling the transmission and impact of the disease. FFX study aims at describing the clinical presentation of the disease and the routes of infection, the secondary infection and clinical attack rate, the serial interval, proportion of asymptomatic infections, the basic reproduction number, the incubation period, and eventually preliminary estimates of infection and disease-severity ratios (case-severity and case-fatality ratios).

Early estimates of these parameters are essential to the design and selection of the more effective public health measures, however, during emergency situations, it might not be feasible to carry out such studies. During the COVID-19 emergency, Portugal included the FFX protocol in the list of studies to implement ([26]), but once the number of cases started to increase exponentially, its implementation was not possible. Since only a few countries were able to conduct this study ([5]), the majority of the estimates of key epidemiological parameters came from the secondary analysis of data from routine epidemiological surveillance systems.

According to the different phases of the pandemic, and with the increase in the number of cases, other studies can be implemented. Such as, cohort or case-control studies aimed at identifying risk or protective factors of infection, disease or disease severity ([26]). These studies are also crucial for the implementation of targeted public health measures, but also to feed the mathematical models that aim to produce impact scenarios in the health care sector.

Other special studies of high importance are population-based surveys. These are cross-sectional studies that aim to measure seroprevalence of antibodies against the infectious agent, or surveys focused on evaluating the knowledge about the infection and disease, the adoption of preventive measures (hand washing, social distancing or use of mask) or the impact of the universal public health measures, such as lockdown of specific areas or countries, on the socioeconomic indicators or even on mental health. A very specific survey that some countries have implemented before and during the COVID-19 emergency, aims at measuring the profile of the contacts between individuals, according to their age and setting (home, school, workplace and general community). These are very important to feed the mathematical models in order to include age heterogeneity in the contacts between individuals and the observed impact of social distancing measures on these contact matrices.

Finally, it is important to mention those studies dedicated at measuring the effectiveness or impact of the public health measures, vaccines or treatments. These are essential to adjust the set of intervention and treatments. During the COVID-19 emergency several studies were performed to evaluate the impact of the lockdown measures, using quasi-experimental designs like interrupted time series analysis. Moreover, complex randomized clinical trials aimed at identifying effective treatments against COVID-19 disease and their complications were also implemented.

In all these studies the role of the statistical methods is undeniable in the several phases of the study development. These encompass the design, field implementation and data validation, analysis, reporting and interpretation of results. One of the lessons learned during the COVID-19 emergency is that the development of these studies needs generic scientific protocols developed and approved during inter-pandemic phases, and most importantly, dedicated teams for their implementation.

4. MATHEMATICAL MODELS OF DISEASE TRANSMISSION

Epidemic models for the spread of infectious diseases date back to the beginning of the twentieth century. Pioneer mathematical techniques were proposed to describe the dynamics of disease transmission and these are key tools to lay out proper mitigation and suppression measures to deal with an epidemic ([13]). The susceptible-infected-recovered (SIR) models developed by Kermack and McKendrick (1927) were the first mathematical models developed to the transmission dynamics of infectious diseases. These compartmental models have been further extended and adapted to an assortment of different diseases ([31]). Nowadays, these models are seldom used in their original form, since they are too simplistic to account for the inherent heterogeneity of disease transmission.

The SIR and SEIR (Susceptible, Exposed, Infectious and Removed) models have been improved to account for several crucial factors, such as the heterogeneity of human contact ([16, 31]), geographical distribution ([4]) and disease susceptibility, as well as accounting for the underlying uncertainty in disease transmission, i.e. stochastic models ([35, 34]). These extended models have been developed for different goals:

a) assessing the epidemic preparedness of health systems ([27, 7]), by taking into consideration the susceptibility and contact pattern of individuals, as well as the infectivity of the pathogen;

b) evaluate the impact of intervention measures that aim to reduce mortality and healthcare demand during an epidemic ([13]);

c) account for the geographical distribution of the population, i.e. metapopulation models ([4, 33]);

d) study the seasonality of disease prevalence ([10]);

e) evaluate the necessary herd immunity vaccination threshold ([17]).

Since the beginning of the SARS-CoV-2 epidemic, a number of different working groups have employed SIR-like deterministic and stochastic models to evaluate the spread of the disease. The authors in [33] used a deterministic SEIR model to nowcast and forecast the national spread of SARS-CoV-2 in China, by creating scenarios for the transmissibility reduction and mobility reduction associated with the measures implemented in Wuhan, China. A SEIR-type stochastic model was developed by the authors of [35] to estimate key latent epidemiological parameters and states, such as the proportion of asymptomatic individuals, and the strength of the contact tracing. The authors in [9] employed the use of a healthy-asymptomatic-sick-dead model to assess the relaxation of social-distancing measures in Germany, a similar approach was also employed by the authors in [24] for Spain.

One of the biggest challenges that epidemiologists and modellers face during this epidemic is to foresee the duration of different epidemic phases. The first phase consisted in the containment of imported cases and identification of transmission chains, after which, an exponential growth of incident cases is expected, hence, the next phase consisted in the introduction of suppression measures by enforcing strict social-distancing. Lately, several countries having been phasing-out social-distancing measures, which might result in higher disease transmission during next winter along with the transmission of other seasonal respiratory viruses. For each of these phases several measures are necessary to prevent, control and mitigate COVID-19's impact, which need to be evaluated and simulated with best available modelling tools.

Other very important discussion for which mathematical modelling has provided very important insights and discussion is the future of the pandemic and its end, such as the herd immunity threshold, with estimates ranging from around 10% to 70% ([14, 6, 1]). Furthermore, should we expect that SARS-CoV-2 will be present for several years and became a seasonal respiratory virus ([21]) together with influenza and other respiratory virus ([14, 6, 1])?

5. FINAL REMARKS

The emergence of SARS-CoV-2 and the COVID-19 pandemic was really a black swan event, although international and national institutions have been preparing for new influenza pandemic, in truth, the majority of the countries were not prepared for all the implications of this event.

Data and information about the epidemic course, the risk factors for infection and disease, the effectiveness of public health measures and treatments, and the future scenarios of the course of the pandemic are among the most wanted pieces of knowledge by all the sectors of society.

These can be obtained through the development of surveillance systems, special epidemiological studies and mathematical modelling, areas that must be developed together in comprehensive and timely manner. Data from surveillance systems is important to feed special studies and mathematical modelling. Special studies allow the estimation of relevant parameters to feed the mathematical models and to estimate transmission parameters in real time (R_t). The public health measures that are evaluated prospectively through mathematical modelled scenarios, can be retrospectively evaluated using surveillance systems data or special studies.

Epidemiologists, statisticians and mathematical modellers have been probably among the most needed professionals during this phase, due to the amounts of data and information that needs to be collected, analysed, reported and interpreted. Data and information is demanded by decision makers with higher levels of certainty and lowest timeliness.

On the side of those that must make decisions, the information and knowledge that public health officials and decision-makers (at the highest level) receive in real time to decide is overwhelming. During this emergency, decision makers needed to deal with complex epidemiological concepts, with high levels of uncertainty. This placed very high pressure in public institutions responsible for delivering the information needed for decision, but also on decision makers and public health authorities that are confronted with uncertain information and are asked to make correct decisions ([18]).

A reflection should be made about several of these issues in the light of the COVID-19 pandemic experience, from the point of view of data and information. These reflections should include the anticipation of data, studies and information needed during the different phases of a pandemic; methods and models available for analysis in order to develop a public available data analysis toolkit and clear rules for data disclosure and availability. The collaboration between public health institutes and the science academies should also be enhanced for the development of these three areas. Finally, efforts should be allocated in the translation of data and information for decision makers in order to contribute to more informed decisions, with a focus on the population health and welfare.

REFERENCES

[1] Aguas, R.; Corder, R.M.; King, J.G.; Gonçalves, G.; Ferreira, U.M. and Gomes, M.G.M. (2020). Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics, *medRxiv*, `https://doi.org/10.1101/2020.07.23.20160762`.

[2] Ali, S.T.; Li, W.; Lau, E.H.Y.; Xu, X.-K.; Du, Z.; Wu, Y.; Leung, G. and Colwing, B.J. (2020). Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions, *Science*, **369**(6507), 1106–1109.

[3] Antunes, L.; Antunes, M. and Nunes, B. (2014). *Modelação do atraso na notificação dos óbitos no sistema de vigilância diária da mortalidade*, Master's thesis, FCUL, Lisbon, Portugal.

[4] Apolloni, A.; Poletto, C.; Ramasco, J.J.; Jensen, P. and Colizza, V. (2014). Metapopulation epidemic models with heterogeneous mixing and travel behaviour, *Theoretical Biology and Medical Modelling*, **11**(1), 3.

[5] Boddington, N. (2020). COVID-19 in Great Britain: epidemiological and clinical characteristics of the first few hundred (FF100) cases: a descriptive case series and case control analysis, *medRxiv*, `https://doi.org/10.1101/2020.05.18.20086157`.

[6] Britton, T.; Ball, F. and Trapman, P. (2020). A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2, *Science*, **360**(6505), 846–849.

[7] Chitnis, N.; Hyman, J.M. and Valle, S.Y.D. (2017). Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases, *Mathematical Biosciences and Engineering*, **15**(5), 1475–1497.

[8] Cori, A.; Ferguson, N.; Fraser, C. and Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics, *American Journal of Epidemiology*, **178**(9), 1505–1512.

[9] Donsimoni, J.R.; Glawion, R.; Plachter, B.; Weiser, C. and Wälde, K. (2020). Should contact bans be lifted in Germany? A quantitative prediction of its effects, *medRxiv*, `https://doi.org/10.1101/2020.04.10.20060301`.

[10] Dorélien, A.M.; Ballesteros, S. and Grenfell, B.T. (2013). Impact of birth seasonality on dynamics of acute immunizing infections in Sub-Saharan Africa, *PLoS ONE*, **8**, 10.

[11] ECDC: Event Background-COVID-19, `https://www.ecdc.europa.eu/en/novel-coronavirus/event-background-2019`.

[12] Estimation of the current development of the SARS-CoV-2 epidemic in Germany – nowcasting Epidemiologisches Bulletin. 17/2020 Robert Koch Institute. `https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/17_20.pdf`.

[13] Ferguson, N.M.; Laydon, D.; Nedjati-Gilani, G.; Imai, N.; Ainslie, K.; Baguelin, M.; Bhatia, S.; Boonyasiri, A.; Cucunubá, Z.; Cuomo-Dannenburg, G.; Dighe, A.; Dorigatti, I.; Fu, H.; Gaythorpe, K.; Green, W.; Hamlet, A.; Hinsley, W.; Okell, L.C.; van Elsland, S.; Thompson, H.; Verity, R.; Volz, E.; Wang, H.; Wang, Y.; Walker, P.G.T.; Walters, C.; Winskill, P.; Whittaker, C.; Donnelly, C.A.;

Riley, S. and Ghani, A.C. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand, *Imperial College of London*, Report 9, `https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/`.

[14] Gomes, M.G.M.; Corder, G.; King, J.; Langwig, K.E.; Souto-Maior, C.; Carneiro, J.; Gonçalves, G.; Penha-Gonçalves, C.; Ferreira, M.U. and Aguas, R. (2020). Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold, *medRxiv*, `https://doi.org/10.1101/2020.04.27.20081893`.

[15] Held, L.; Hens, N.; O'Neill, P. and Wallinga, J. (2019). *Handbook of Infectious Disease Data Analysis*, Chapman and Hall/CRC, New York.

[16] Hens, N. (2012). *Modeling Infectious Disease Parameters Based On Serological and Social Contact Data: A Modern Statistical Perspective*, Springer, New York.

[17] Hethcote, H.W. (2013). The mathematics of infectious diseases, *SIAM Review*, **42**(4), 599–653, `https://doi.org/10.1137/s0036144500371907`.

[18] Hilderink, H.B.M. (2020). The corona crisis and the need for public health foresight studies, *European Journal of Public Health*, **30**(4), 616.

[19] Höhle, M. and van der Heiden, M. (2014). Bayesian nowcasting during the STEC O104:H4 outbreak in Germany, *Biometrics*, **70**, 993–1002.

[20] INSA: Covid-19 curva epidémica e parâmetros de transmissibilidade, `http://www.insa.min-saude.pt/category/areas-de-atuacao/epidemiologia/covid-19-curva-epidemica-e-parametros-de-transmissibilidade/`.

[21] Kissler, S.; Tedijanto, C.; Goldstein, E.; Grad, Y.H. and Lipsitch, M. (2020). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period, *Science*, **368**(6493), 860–868.

[22] Kurcharki, A. (2020). *The Rules of Contagion: Why Things Spread – and Why They Stop*, Profile Books Ltd, London.

[23] Lipsitch, M.; Finelli, L.; Heffernan, R.T.; Leung, G.M. andRedd, S.C. (2011). Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1, *Biosecur. Bioterror.*, **9**, 89–115.

[24] Lopez, L. and Rodo, X. (2020). The end of the social confinement in Spain and the COVID-19 re-emergence risk, *medRxiv*, `https://doi.org/10.1101/2020.04.14.20064766`.

[25] Nunes, B.; Natário I. and Carvalho, M.L. (2013). Nowcasting influenza epidemics using non-homogeneous hidden Markov models, *Stat. Med.*, **32**(15), 2643–2660.

[26] Plano Nacional de Preparação e Resposta à Doença por novo coronavírus (COVID-19), `https://www.dgs.pt/documentos-e-publicacoes/plano-nacional-de-preparacao-e-resposta-para-a-doenca-por-novo-coronavirus-covid-19-pdf.aspx`.

[27] Saunders-Hastings, P.; Hayes, B.Q.; Smith, R. and Krewski, D. (2017). National assessment of Canadian pandemic preparedness: employing InFluNet to identify high-risk areas for inter-wave vaccine distribution, *Infectious Disease Modelling*, **2**(3), 341–352.

[28] SINAVE: Sistema Nacional de Vigilância Epidemiológica, `https://www.dgs.pt/servicos-on-line1/sinave-sistema-nacional-de-vigilancia-epidemiologica.aspx`.

[29] The First Few X (FFX) Cases and contact investigation protocol for 2019-novel coronavirus (2019-nCoV) infection, `https://www.who.int/publications/i/item/the-first-few-x-(ffx)-cases-and-contact-investigation-protocol-for-2019-novel-coronavirus-(2019-ncov)-infection`.

[30] Tsang, T.K.; Wu, P.; Lin, Y.; Lau, E.H.Y.; Leung, G.M. and Cowling, B.J. (2020). Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study, *Lancet Public Health*, **5**, e289–e296.

[31] Vynnycky, E. and White, R.G. (2019). *An Introduction to Infectious Disease Modelling*, Oxford Univ. Press, Oxford.

[32] Walling, J. and Teunis, P. (2014). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures, *American Journal of Epidemiology*, **160**(6), 509–516.

[33] Wu, J.T.; Leung, K. and Leung, G.M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, *The Lancet*, **395**(10225), 689–697.

[34] Yan, P. and Chowell, G. (2019). *Quantitative Methods for Investigating Infectious Disease Outbreaks*, Springer, Cham.

[35] Zhang, Y.; You, C.; Cai, Z.; Sun, J.; Hu, W. and Zhou, X.H. (2020). Prediction of the COVID-19 outbreak based on a realistic stochastic model, *medRxiv*, `https://doi.org/10.1101/2020.03.10.20033803`.