modelling pre-symptomatic infectiousness in coviD-19

|  |  |
| --- | --- |
| *Dr. Russell Cheng* | *Dr. Christopher Dye* |
|  |  |
| University of Southampton | University of Oxford  |
| Highfield, SO17 5BJ | Mansfield Road, OX1 3SZ,  |
| United Kingdom | United Kingdom |
| cheng@btinternet.com | Chrisdye56@gmail.com |
|  |  |
| *Dr. John Dagpunar* | *Dr. Brian Williams* |
| University of Southampton | SACEMA |
| Highfield, SO17 1BJ | Stellenbosch University |
| United Kingdom | South Africa |
| jdagpunar@hotmail.com | williamsbg@me.com |

ABSTRACT

This paper considers SEPIR, an extension of the well-known SEIR continuous simulation compartment model. Both models can be fitted to real data as they include parameters that can be estimated from the data. SEPIR deploys an additional pre-symptomatic (also called asymptomatic) infectious stage not in SEIR but known to exist in COVID-19. This stage can also be fitted to data. We focus on how to fit SEPIR to a first wave. Both SEPIR and the existing SEIR model assume a homogeneous mixing population with parameters fixed. Moreover neither includes dynamically varying control strategies deployed against the virus. If either model is to represent more than just a single wave of the epidemic, then the parameters of the model would have to be time dependent. In view of this we also consider how reproduction numbers can be calculated to investigate the longer term overall result of an epidemic.

**Keywords**: Differential equation epidemic models, Parametric models, Effective Reproduction Number, Asymptomatic transmission

# INTRODUCTION

A parametric SEIR model has been used by the authors in Dye et al. (2020) to compare the first wave of the COVID-19 epidemics in different European countries. In Dye et al. (2020) this model is fitted to data using the method of maximum likelihood estimation rather than perhaps the more widely-used Bayesian Markov-chain. The compartmental structure of the SEIR model is standard and does not include a specific compartment to represent the pre-symptomatic (also asymptomatic) infectiousness stage known to occur in those infected by COVID-19. We describe the SEIR and SEPIR models in Sections 2 and 3, focusing on the models themselves rather than on the effect of the epidemics on the affected countries. We discuss the fitting of these models to data, focusing on use of the maximum-likelihood method of estimation which produces (point) estimates of parameter values, as this gives an unequivocal specific model representation of the epidemic. In Section 3.2 we give a numerical example based on the first wave of the COVID-19 epidemic in Switzerland.

 An important aspect of the basic maximum-likelihood method is that the parameters values are assumed to be unknown but fixed in value. Similarly, in the Bayesian case, the distributions of the parameters are not only unknown, but are assumed to be fixed. However different strategies varying over time have to be deployed in trying to contain a fast moving epidemic like that produced by COVID-19. This means that the model parameters do not remain constant but have to be time dependent if the trajectory of the epidemic is to be correctly reproduced. Note that use of the models based on SEIR can be used in examining more than one wave, see for example Dagpunar (2020).

 For reasons of space and simplicity we focus in this paper on ‘first wave’ behaviour ot the SEPIR model, leaving for elsewhere discussion of situations where time-varying parameters might be used.

 We do however discuss how progress of an epidemic is summarised by the *effective* reproduction number *Rt*, a dynamically varying version of *R*0, the (basic) reproduction number. Theoretically, *R*0 is unequivocally defined in terms of the idealized epidemic infecting a homogeneously mixed population with no control measures and assuming every member of the population is susceptible.

 However when monitoring the progress of an epidemic *Rt* is more useful, and in lay terms, is generally referred to as the reproduction number. It can still be defined to be the expected number of persons infected by an infective individually, but is made time dependent because of changes in the management of the epidemic and in the susceptible proportion . We consider how to calculate *Rt* in Section 4 which also examines how *R*0 itself can be calculated for the SEIR and SEPIR models as these can both be regarded as what are called SInRmodels, which include *n* multiple infectious stages, as described in Ma and Earn (2006), who discuss calculation of *R*0 in these.

# The seir model

The SEIR model has been described in the Supplementary Materials of Dye et al. (2020), but for ease of comparison with SEPIR model we give the description again here. The model is of a homogeneously mixed population with four compartments representing those who are susceptible, exposed, infectious and recovered (SEIR), as shown in Figure 1.

I

S

E

R

**Figure 1**. *The SEIR model. The compartments denote those in the*

*population that are Susceptible, Exposed, Infected and Recovered.*

The variables *S*, *E*, *I* and *R* satisfy the ordinary differential equations:

 (1)

 (2)

 (3)

 (4)

A convenient recent reference is Ma (2020) who uses a slightly different notation. Also, to highlight deaths due to the virus we divide those that recover well and those that die due to the virus. Thus the infectious individuals are divided into two compartments as illustrated Figure 2.

R\Z

Z

*I*

I

 Z

**Figure 2**. *Adjustment of the SEIR model where R is divided into two compartments, R \ Z, those that recover and Z, those that die; where is the proportion that recover.*

 More elaborate models can and have been developed. For example, see Dagpunar (2020) who extends R into additional compartments representing different outcomes of hospitalization

 The SEIR model of Figures 1 and 2 are assumed to depend on certain parameters, initially assumed unknown. Fitting the model to data, is simply the process of estimating the parameters, either directly using data obtained from observing the epidemic, or from information obtained from other sources. Once the parameter values are estimated, the behaviour of the SEIR model is completely specified. The parameters are defined in Dye et al. (2020). To avoid repetition and confusion they are not discussed directly here as we shall be discussing the SEPIR model where a very similar set of parameters will be fully defined.

 However we do point out here the time-delay parameter used to modify equation (4) to:

 (5)

We denote by the vector of parameters, where *m* is the number of parameters. In Dye et al. (2020), In the SEPIR model of Section 3, . With given, the four differential equations (1), (2), (3) and (5) can be solved by numerical integration to give the trajectories

 for (6)

where *t* is the day and *M* is the number of days of interest. We used the standard method of Maximum Likelihood (ML), as given for example in Cheng (2017), to estimate parameter values.

 Here we outline the approach used to estimate the parameters from a sample of observed daily deaths. Let the sample of observed number of daily deaths be denoted by

 (7)

where is the number of deaths on day *t* and is the number of days observed. If the observations were made without error and the right parameter values are correct for , then the death trajectory would match the observed deaths **Z** in (7) and the model would then be successful in explaining deaths.

 To include statistical uncertainty in the model we assume instead

 (8)

where is random error. For simplicity the are assumed to be normally and independently distributed (NID) with standard deviation i.e.

 ) ∼ NID(0,²), so that ∼ NID(0,²) (9)

Note that is treated as a parameters so is included as a component of **.**

The logarithm of the distribution of the sample is then

 (10)

where is the random argument, and the parameters are fixed. In ML estimation (MLE), this is turned on its head so that **Z** is simply the known sample of observations now regarded as fixed and we write L as ) calling it the (log)likelihood to indicate that it is now treated as a function of . The ML estimator is simply the value of at which L(|) is maximized. i.e.

 = {(L(|)}. (11)

Nelder-Mead numerical search for the maximum was used. This goes through different *i*=1, 2, 3,… comparing the different ) to find , the best .

 To simplify the description of the estimation process, only fitting to deaths data, **Z** as in (7) has been described, but the method extends straightforwardly to include other data samples. For example

 (12)

where is the number of prevalent active cases on day *t*. Fitting simultaneously to both **Y** and **Z** can be carried out by adding to the right-hand side of (10) a corresponding set of terms for **Y**

Each step of the Nelder-Mead optimization is summarized as follows. The trajectory of each of the variables *S*, *E*, *I*, *R*, as given in Equation (6) is calculated, for simplicity using Euler step-wise integration of thedifferential equations (1) – (4) , but with step-length 1/8th of day as a step length of 1 is quite inaccurate. These are essentially scale invariant so we can standardise the equations taking

 .

This choice of 1 for the standardising constant makes all four variables *fractions*. Initial values are S(0, ) = 1, and E(0, ) a small quantity, these denoted by *s*0 and *e*0 respectively. However these are subsequently adjustable, so are treated as parameters. Also we set *I*(0, ) = *R*(0, ) = 0.

 With the state variable trajectories obtained, the likelihood is then calculated. This requires values of all the parameters, in particular *N*, the population size, which is treated as a parameter. This is appears in calculating the differences used in the likelihood to compare the model deaths with the observed cumulative deaths . Thus *N* is taken into account and can be changed in selecting the next .

# The sePir model

**3.1 Structure of the SEPIR Model**

In the SEPIR model we introduce an extra compartment to the SEIR model in Fig. 1 changing it to Fig 3:

I

S

E

R

P

**Figure 3**. *The SEPIR model. The compartment P denotes those who are infectious but are pre-symptomatic whilst I denotes that are infectious and symptomatic*.

The original Icompartment is now split into two with its first compartment, P, comprising those infectious who are pre, i.e. asymptomatic, and the second, I, comprising those infectious who display symptoms. The ordinary differential equations (1), (2) and (3) in the SEIR model are replaced by the differential equations (13), (14), (15) and (16), with (4) and (5) remaining unchanged. There are two terms in going from S to E, comprising: those infected by someone in P, with transmission rate , and those infected by someone in I, with transmission rate The reciprocal is the mean period someone spends in state (compartment) P whilst is the mean period spent in I.

 (13)

 (14)

 (15)

 (16)

We treat the quantities as parameters to be estimated. However we include six further parameters , and These are all listed and defined in columns 1 and 2 of Table 1.

 Some of the parameters can be given fixed predetermined values with the others fitted to data by Maximum Likelihood as described in the SEIR model in Section2.

**3.2 Switzerland: A Numerical Example**

Column 3 gives the parameter values when SEPIR was fitted estimating all 11 parameters by maximum likelihood using *M* = 109 days of data based on daily observations starting on 15 Feb 2020. Two series: Daily New Cases and Daily Deaths were used. The values of all the parameters are of interest. The parameter values for Switzerland are given in Table 1. We highlight two aspects.

 Firstly, the SEPIR model, because the differential equations are scale invariant, can be standardised so that the population size is 1. However we allow the population size to be variable with the size estimated by allowing rescaling, when we maximize the likelihood. The estimated population size of 36,700 is remarkably small compared with the actual population size of 8.2 million. The main reason for the difference is that the model does not include a mechanism of epidemic control which we know took place in every country to prevent the spread of infection. The SEPIR model, which assumes a homogeneously mixed population can only allow for this by changing the population size. Moreover without examining regional records it may be that the outbreak in Switzerland was mainly confined to parts nearest Italy, the first European country to be badly affected by COVID.

**Table 1**. *Parameters of the SEPIR model with estimates for Switzerland*.

|  |  |  |
| --- | --- | --- |
| Symbol | Definition | Estimated value and 95%confidence interval |
|  | pre-symptomatic transmission rate  | 0.44 (0.435– 0.454) |
|  | symptomatic transmission rate | 0.137 (0.128– 0.147) |
|  | mean latent period in compartment E | 1.19 (1.13– 1.26) |
|  | mean pre-symptomatic period | 4.40 (4.00– 4.62)  |
|  | mean symptomatic period | 14.0 (13.8– 14.3) days |
|  | number of days from start of epidemic before observations began | 30 days (too small to measure) |
|  | initial proportion of individuals exposed | 6.6 (5.6– 8.2) E-07 |
|  | numerical size of exposed population  | 3.67 (3.60 – 3.74) E+04 |
|  | standard deviation of observational error | 103 (82 – 106) |
|  | probability of an infective recovered well | 0.943 (0.942 – 0.945) |
|  | mean time between the end of infectiousness and recovering well or death  | 3.0 (2.8 – 4.0) days |

 Secondly we examine whether the SEPIR model gives any indication of the extent of the pre-symptomatic stage.

 To do this, we first summarize what is already known about this stage by reporting the findings of He at al. (2020) who investigated the case histories of 77 infector-infectee pairs in each of which an infectious person, the infector, goes on to infect a susceptible person, the infectee.

 Citing the mean incubation period as 5.2 days, He et al, (2020) estimate the serial interval to be 5.8 days. From this they infer that infectiousness starts 2.3 days after the onset of infection and peaks just 0.7 days before symptom onset, giving an estimated proportion of infections of 44% as occurring before the onset of infector symptoms. Infectiousness then declined within 7 days. Figure 4 is a schematic showing the infector-infectee relationship.

 The estimate of He et al. (2020) that the proportion of individuals infected pre-symptomatically is 44% means, in our case, that the proportion ()/() should therefore be this value at least approximately. From Table 1, the value is 50.4%. This is in accord with the higher pre-symptomatic infection proportions given by Tapiwa et al (2020): 48% in Singapore and 62% in Tianjin. The practical consequences of this finding is evident, with elaborate track and tracing required to identify pre-symptomatic infections.

 

Serial Interval Probability Distribution of a Secondary Infection

**Figure 4**: *Infector-Infectee Relationship as described by He et al (2020)*.

 

**Figure 5**: *SEPIR Active Cases and Cumulative Deaths fitted to Swiss*

*Data. Horizontal axis is days with day #1 = 15th February 2020*.

 The quality of the fit achieved by the SEPIR model is illustrated in Figure 5 where the Active Cases and Cumulative Deaths curves obtained by fitting the model to both data sets simultaneously are plotted with their corresponding data.

 The SEPIR model was then repeatedly fitted to independent parametric bootstrap replications of the actual observed active cases and cumulative deaths data. As described in Subsection 4.1.3 of Cheng (2017), confidence intervals for the parameters can be obtained from the bootstrap parameter estimates. For illustration, the resulting 95% confidence intervals for each of the fitted parameter estimates are reported in Column 3 of Table 1, using 500 bootstraps.

 Charts of fitted SEPIR trajectories provide an easily understood way to display results. For example the fitted SEPIR cumulative deaths trajectory (red) is displayed in Figure 6 together with the observations (black). The method described in Section 4.3 of Cheng (2017) can be used to provide a confidence band for any model trajectory. For example we have a bootstrap cumulative deaths trajectory corresponding to each bootstrap sample. These are plotted (in grey) in Figure 6 giving a bundle of trajectories, with 95% confidence limits (green and blue). Only 250 bootstrap are depicted.

 

 **Figure 6:** *SEPIR Fitted Cumulative deaths trajectory (red) for Swiss data obtained*

 *from 109 observations (black). Upper (green) and lower (blue) confidence limits.*

# *Rt* the Effective Reproduction Number

The SEIR and SEPIR models are both SInR models with n multiple infectious stages as defined in Equations (6a)-6(d) in Ma and Earn (2006), whose Equation (7) gives formulas for the Reproduction Number, *R*0, in the models. (The Reproduction Number *R*0 is simply, but precisely, defined as the number of susceptible individuals that an infectious person will go onto infect when the epidemic first starts, assuming that the population is homogeneously mixed.) For the SEPIR model, using the parameters already appear in Equations (13)-(16), we have, in our Swiss example, that

 = 3.85. (17)

which seems not implausible.

 As mentioned in the Introduction, in practice *Rt*, the effective reproduction number, is more useful as, throughout the epidemic, it can be continually used to gauge how well control strategies are working. The theoretical basis underlying the calculation *Rt*is well described by Ma (2020). We have

 , (18)

where *c*(*t*) is the incidence curve of new cases at time *t* and *w*(*u*) is the *serial interval probability distribution* of a secondary infection; so that *w*(*u*)*du* is the probability that an infectious individual (the infector)) infects someone else (the infectee) in the time period (*u*, *u+du*). This probability distribution of the time between the infections of an infector-infectee pair is depicted in Figure 4.

 The denominator in Equation (18) measures how the new cases at time *t* arise from those infected prior to time *t*. The epidemic clearly is rising or falling depending on whether the numerator is larger or small than the denominator, with equilibrium when they are equal. Thus *Rt* has the critical reproduction property of *R*0 but moreover is dynamic, so that it can be used to gauge the progress of the epidemic as this develops.

It turns out that the formula (18) is quite robust so that the serial interval distribution does not have to be estimated all that accurately. In fact Germany, early during its first COVID-19 epidemic wave, used the simple denominator *c*(*t*-4). Cori et al. (2013), though using a Bayesian approach, examined various empirically obtained serial interval distributions drawn from different epidemics. In Dye et al. (2020) the authors used a discretized and shifted gamma distribution to represent the serial interval distribution *w*(*u*) that is shown as the red curve in Figure 4, calculating the denominator as

 . (19)

Figure 7 depicts calculated using this formula for Switzerland when c(t) is a daily 7-day moving average of new cases.



**Figure 7:** Chart of the effective *Rt* calculated using the formula in Equation (19)

 for Switzerland where is the 7-day moving average of new cases.

An important point is that how the epidemic ultimately ends depends on *R*0, not *Rt*. Until the arrival of vaccines all controls, lockdown, hand-washing, social distancing and so on, affect only Thus *Rt*, which depends on this temporary varies as the controls vary, and so is a simple gauge of how current controls are doing. However once controls are removed, returns to its original value as in *R*0, so that the epidemic returns, causing another wave. As determined by Kermack and McKendrick (1927) for the SIR model, the ultimate end, when a given number of susceptibles have been infected, is determined by *R*0; the remaining uninfected susceptibles, satisfying the so-called final-size relation:

 (20)

 This final-size relation also applies to the SEIR and SEPIR models as these are both SInR models with multiple infectious stages as defined in Equations (6a)-6(d) in Ma and Earn 2006, whose Equation (7) gives formulas for *R*0 in the models.

 Vaccination works differently by moving susceptibles directly to the *R* compartment so that the effective population shrinks. This is most clearly seen if immunization takes place before the epidemic so that a proportion *p* of the population is immunized (see Brauer et al. 2019). Then *R*0 would shrink to . (Here is the in Brauer et al. 2019 which assumes mass action incidence where an infectee makes contacts sufficient to transmit infection in unit time.) is less than 1 if

 ,

when there would be no epidemic. Thus not all of the population needs immunization to prevent the epidemic. For example if over two thirds of the population would need immunization. This clearly shows that the only long-lasting control is vaccination, At time of writing, the race is on in the UK, with covid-19 continuing to infect large numbers, but with vaccination rates being accelerated to counter this. Also the new variants suggest that a basic reproduction number of *R*0 as high as 5 is not out of the question. That would mean that under no Non-Pharmaceutical Interventions, 80 % of the population would need vaccination before the growth rate became negative. With such a high *R*0, almost the entire population would need vaccination for the disease to be eliminated, and even this assumes that new cases are not imported from other countries. It seems likely then that Covid-19 will be with us for a long time and effective test, trace, and isolation will be needed in the long term.

# closing remarks

In conclusion the SEPIR model is an extension of the well-known SEIR model. Both are particular cases of the more general SInR model with multiple infectious stages. For Covid-19, the SEPIR compartment, P, is used to represent those infected pre-symptomatically. Compared with the SEIR model this is an important improvement, as shown in our example. This latter is based on real data, and clearly shows the large part played by pre-symptomatic infection in the case of Covid-19. The practical implication is that Covid-19 control strategies need to recognise and deal with pre-symptomatic transmission to be truly successful. Our formulation of the SEPIR model includes adjustable parameters which can either be given or fitted to observed data. An important aspect of our parametrisation is that it allows estimation of an initial susceptible proportion that is less than unity rather than supposing that the susceptibles comprise the country’s whole population, as is usually assumed. This relaxes the assumption of homogeneity of virus transmission throughout the whole population, as this assumption may not be reasonable, especially in the early stages of an epidemic where the number of individuals infected is initially small.

 It should be noted that our transmission rates, though estimated, are supposed constant rather than time dependent. This latter would be needed to model changing management of the epidemic. This could explain why the estimates of some of the biological parameters are rather different from those observed in some other studies.

 We end with two caveats.

 Firstly we have not yet examined in detail the robustness of the maximum likelihood optimization used to fit the model. In our numerical example we chose the first wave of the epidemic in Switzerland because the data corresponded well to the characteristics of the SEPIR model. However even in this example alternative good fits can be achieved with combinations of parameter values different from those reported in Table 1. Thus, in practice, comparison with parameter estimates obtained in other ways should always be made where possible to assess when the estimates obtained can be relied on.

 Secondly, the simplicity of models such as SEIR or SEPIR. means that the practical usefulness of using them on their own, in isolation, is limited. The models are idealizations of the way the epidemic behaves and of population behaviour. Thus control policies are not modelled nor their influence on population behaviour. Indeed lack of homogeneity of population behaviour is an important factor that has to be addressed in implementing control policies because these latter have to recognize the issues they give rise to, for the population as a whole to be willing to follow them. At this time of writing, resurgence of the virus has taken place and more virulent virus strains have appeared, but balanced by the availability of vaccines, This all requires a national control policy which is fair. On this basis, it is not unfair as at time of writing and as adopted by present national policy to go in “earlier and harder and stay longer than might be thought necessary” even in areas with low prevalence. The alternative of delaying, all too easily is likely to ultimately inflict on such areas similar damage to that currently experienced by high tier areas

 Ideally a detailed model allowing for local differences is required, but seems unrealistic given the speed of changing events. . However, less complicated models like the present SEPIR model may be helpful in informing decision makers who may otherwise only have time to use just simple common sense in making their decisions.

**REFERENCES**

Brauer F, Castillo-Chavez C, Feng Z. (2019) Simple Compartmental Models for Disease Transmission. In: *Mathematical Models in Epidemiology. Texts in Applied Mathematics*, vol **69**. Springer, New York, NY. <https://doi.org/10.1007/978-1-4939-9828-9_2>

Cheng R C H (2017). *Non-Standard Parametric Statistical Inference*. Oxford University Press, Oxford.

Cori A, Ferguson N M, Fraser C and Cauchernez S (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology* **178(9)**: 1505-1512.

Dagpunar J S (2020). Sensitivity of UK Covid-19 deaths to the timing of suppression measures and their relaxation. *Infectious Disease Modelling* **5:** 525-535 <https://doi.org/10.1016/j.idm.2020.07.002>

Dye C, Cheng R C H, Dagpunar J and Williams B G (2020). The scale and dynamics of COVID-19 epidemics across Europe. [*R*](file:///C%3A%5CMyDocsA-N%5CA-Epidemics%5CSW21%20Paper%5CR)*. Soc. Open Sci.* . **7**: 201726 [https://doi.org/10.1098/rsos.201726](%20https%3A/doi.org/10.1098/rsos.201726)

He L, Lau E H Y, […], Leung G M (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26**: 672-675.

Kermack W O and McKendrick A G (1927). A contribution to the mathematical theory of epidemics, *Proc. R. Soc. Lond. Ser. A* **115**: 700-721.

Ma J and Earn D J D (2006). Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology* **68**: 679-702.

Ma J (2020). Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling* **5**: 129-141.

Tapiwa G, Kremer C, Chen D, Torner A, Faes C, Wallinga J and Hens N (2020). Estimating the gerneration interval for coronavirus disease (COVID-19) based on symptom onset, March 2020. *Euro Surveill*. **25(17**): pii=2000257.

author biographIES

**RUSSELL CHENG** retired from the University of Southampton in 2007 where he had been Head of the Operational Research Group, having held previous positions at Cardiff University and the University of Kent at Canterbury. <https://www.southampton.ac.uk/maths/about/staff/rchc.page>

**CHRISTOPHER DYE** FRS, FMedSci, has held positions at the London School of Hygiene and Tropical Medicine, the World Health Organization and Gresham College London. He is currently Professor of Epidemiology at Oxford University <http://en.wikipedia.org/wiki/Christopher_Dye>

**JOHN DAGPUNAR** retired from Edinburgh University in 2008. He is Visiting Research Fellow in Mathematical Sciences at the University of Southampton. His research interests are in simulation, financial mathematics, health studies, and reliability.

<https://www.southampton.ac.uk/maths/about/staff/jd2y15.page>

**BRIAN WILLIAMS** is Senior Research Fellow at the South African Centre for Epidemiological Modelling and Analysis (SACEMA) having held the position of Epidemiologist at the World Health Organisation from which he retired in 2008.