Mathematical Biology

Structural identifiability analysis of age-structured PDE epidemic models

Marissa Renardy[1] (ID) **· Denise Kirschner[1] · Marisa Eisenberg[2,3]**

Abstract

Computational and mathematical models rely heavily on estimated parameter values for model development. Identifiability analysis determines how well the parameters of a model can be estimated from experimental data. Identifiability analysis is crucial for interpreting and determining confidence in model parameter values and to provide biologically relevant predictions. Structural identifiability analysis, in which one assumes data to be noiseless and arbitrarily fine-grained, has been extensively studied in the context of ordinary differential equation (ODE) models, but has not yet been widely explored for age-structured partial differential equation (PDE) models. These models present additional difficulties due to increased number of variables and partial derivatives as well as the presence of boundary conditions. In this work, we establish a pipeline for structural identifiability analysis of age-structured PDE models using a differential algebra framework and derive identifiability results for specific age-structured models. We use epidemic models to demonstrate this framework because of their wide-spread use in many different diseases and for the corresponding parallel work previously done for ODEs. In our application of the identifiability analysis pipeline, we focus on a Susceptible-Exposed-Infected model for which we compare identifiability results for a PDE and corresponding ODE system and explore effects of age-dependent parameters on identifiability. We also show how practical identifiability analysis can be applied in this example.

✉ Marissa Renardy
renardy@umich.edu

Marisa Eisenberg
marisae@umich.edu

[1] Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, USA

[2] Department of Epidemiology, University of Michigan, Ann Arbor, USA

[3] Department of Mathematics, University of Michigan, Ann Arbor, USA

Keywords Epidemiology · Age structure · Partial differential equations ·
Tuberculosis · Modeling

Mathematics Subject Classification 35R30: Inverse problems (undetermined
coefficients, etc.) for PDE 92D30: Epidemiology

1 Introduction

Identifiability analysis addresses the question of whether the parameters of a mathematical model can be uniquely identified, or otherwise, given the system observations. There are two general types of identifiability analysis: structural (assuming data to be noiseless and arbitrarily fine-grained) and practical (using real data). These analyses are crucial to interpreting biologically relevant predictions from computational and mathematical models that rely on parameter values that are estimated from experimental datasets. Previous studies have shown that unidentifiable models can lead to vastly different predictions for different parameter values that are able to produce the same observed data (Kao and Eisenberg 2018). In the context of biological models, parameter identifiability gives increased confidence in model predictions that are based on estimated parameters. This is particularly relevant for predictions applied to topics such as diseases, treatment outcomes, vaccination efficacy, and future infection dynamics. Specifically, the use of epidemiological models has become increasingly necessary and useful for predicting disease outcomes, as we have seen in the case of COVID-19, and these efforts should be as accurate and confident as possible to produce the most useful predictions. Thus, we focus our work in the context of epidemic models due to their use in many diseases and to compare with corresponding parallel work previously done for ODEs regarding structural identifiability.

Many theoretical results and algorithms for structural identifiability of linear and non-linear ordinary differential equations (ODEs) models have been derived (Miao et al. 2011; Audoly et al. 2001; Ljung and Glad 1994; Pohjanpalo 1978; Chappell et al. 1990; Cobelli and DiStefano 1980; Chis et al. 2011) and have been applied to study an array of diseases. There are models for the spread of cholera (Eisenberg et al. 2013), vector-borne diseases (Kao and Eisenberg 2018; Zhu et al. 2018; Tuncer et al. 2016), and general infectious diseases (Tuncer and Le 2018; Evans et al. 2005, 2004). Methods for structural identifiability are often analytical in nature (e.g. the commonly used differential algebra approach (Audoly et al. 2001; Ljung and Glad 1994)), but may also be numerical, or some blend of the two (Jacquez and Greif 1985; Jacquez and Perry 1990; Raue et al. 2009). Numerical approaches to structural identifiability typically provide insight at the local, rather than global, level.

ODE models are not always sufficient for representing disease dynamics when, for example, age is an important factor. In many diseases, different age groups may respond differently to diseases or disease interventions based on differences in susceptibility (Gardner 1980), mortality (Castillo-Chavez et al. 1991), and contact rates (Mossong et al. 2008; Glasser et al. 2012). Age preferences have been shown to have a significant effect on mixing patterns that drive disease spread (Valle et al. 2007). This type of information can be incorporated into a mathematical model for epidemics, for example,

by explicitly representing age-structured distributions of the susceptible and infected subpopulations. Age-structured models have been used to evaluate and optimize age-targeted vaccination strategies for several diseases (Harris et al. 2019; Keeling and White 2011; Hao et al. 2019; Castillo-Chavez and Feng 1998; Shim et al. 2006), and it has been shown that population heterogeneity and non-random mixing can have a substantial effect on the prediction of disease outbreaks and herd immunity (Glasser et al. 2016; Feng et al. 2015).

In a continuous setting, age structure is introduced by creating state variables that are functions of both age and time, and also by introducing transport terms to represent constant aging of a population. These model formulations result in PDEs, rather than ODEs, making them more biologically accurate; however, such models also typically more complex to analyze. There are other PDE formulations of epidemic models such as age of infection models (Thieme and Castillo-Chavez 1993; Inaba and Sekine 2004) or spatially explicit models (Bertuzzo et al. 2009; Wu 2008). The focus of this paper will be specifically on age-structured models. While there are alternative approaches to incorporating age structure, such as to use compartmentalization to stratify populations (Chow et al. 2011; Ferguson et al. 1996) or to use a stochastic discrete representation such as an agent-based model (Ajelli et al. 2010; Bian 2004), we focus here on the PDE formulation. Importantly, the issues of both structural and practical parameter identifiability for these types of models still needs to be elucidated.

While structural identifiability analysis has been extensively applied in the context of ODE epidemic models, it has not been widely explored for age-structured PDE models (Evans et al. 2004, 2005). PDE models present additional difficulties due to the increased number of variables and derivatives as well as the presence of boundary conditions. Identifiability of age-structured PDE models was first considered by Perasso et al. (2011) for a simple SI-type model and was again considered by Perasso and Razafison (2016) for a population dynamics model including only birth and death. These analytical results have not yet been extended to more comprehensive PDE epidemic models, though some identifiability results have been shown numerically (Tuncer et al. 2016). For these types of problems, modelers have relied on practical identifiability analyses which are dependent on either available data or data that has been synthesized for the purpose of analysis.

In this paper, we derive analytical identifiability results for more complex age-structured epidemic models using a differential algebra framework that is simpler to apply than the previous (aforementioned) previous approaches. We apply this methodology to a compartmental epidemic Susceptible-Exposed-Infected (SEI) model both with and without age structure explicitly modeled, and we compare the identifiability results under different conditions such as the presence of immigration in the system or for age-dependent model parameters. To corroborate our structural identifiability results, we also perform a practical identifiability analysis for one version of the model. While we focus on age-structured models, the methods we present herein can be applied to other PDE formulations, with certain assumptions about smoothness and uniqueness of solutions.

2 Methods

2.1 Identifiability framework

To consider parameter identifiability of a model, we must first define the model structure. From an identifiability or parameter estimation perspective, a model is composed of two pieces: (1) a state model or set of state equations, which describe the underlying system of interest, and (2) a measurement model or set of *output equations*, which describe the observation or measurement of interest (i.e., the quantities of interest for which one either has or plans to collect data). In this study, we assume the following general first-order partial differential equation (PDE) model structure, which is commonly used in age-structured PDE models in biology and health:

$$
\begin{aligned}
\frac{\partial \mathbf{x}}{\partial t} + \frac{\partial \mathbf{x}}{\partial a} &= f(\mathbf{x}, t, a, \mathbf{p}, \mathbf{u}) \\
\mathbf{x}(0, a) &= f_0(\mathbf{x}, a, \mathbf{p}) \\
\mathbf{x}(t, 0) &= f_1(\mathbf{x}, t, \mathbf{u}, \mathbf{p}) \\
\mathbf{y} &= g(\mathbf{x}, t, a, \mathbf{p})
\end{aligned}
\tag{1}
$$

where $\mathbf{x}$ is a vector of state variables describing the system of interest, $\mathbf{p}$ is the vector of parameters, t and a are two independent variables. The model may include initial and boundary conditions that are generic (unknown), written in terms of input parameters, or known. In the following definitions, we assume generic initial and boundary conditions. Non-generic initial conditions can be addressed as detailed in Pia Saccomani et al. (2003), and we assume that non-generic boundary conditions can be addressed analogously. Here we consider time and age as the independent variables, although the approach given here will also work for more general variables, as well as for additional derivatives of further independent variables on the left hand side. Additionally, $\mathbf{y}$ represents the vector of n_y model outputs (measured variables) and $\mathbf{u}$ represents any known inputs (e.g. forcing functions) to the system, if they exist, which are assumed to be arbitrarily differentiable (or at least sufficiently differentiable for determining the input-output equations, described below). Finally, f and g are rational functions of the model variables and parameters such that the model can be written equivalently as differential polynomials by clearing denominators.

In structural identifiability analysis, it is assumed that the inputs and outputs are known (observed) perfectly – in this case, we assume that u, and y are known for all times and all ages and that there is no noise in the data. Similarly, the independent variables t and a are assumed to be known. Thus, the resulting identifiability results can be interpreted as the upper limit of what can possibly be identified with perfect data collection methods. However, structural identifiability is a necessary condition for identifiability in the realistic data case, making it an important first step in the parameter estimation process.

A model such as (1) can be thought of as a map from parameter space (e.g. for $\mathbf{p} \in \mathbb{R}^{n_p}$) to output space (the space of trajectories for $\mathbf{y}$), often termed the model map $\Phi : \mathbf{p} \mapsto \mathbf{y}$. With this framing, the structural identifiability question becomes

that of whether or not the model map Φ is injective. In particular, we define structural identifiability as follows:

Definition 1 For a model of the form (1), an individual parameter p in $\mathbf{p}$ is *uniquely (or globally) structurally identifiable* if for almost every value $\mathbf{p}^*$ and almost all initial conditions, the equation $\mathbf{y}(\mathbf{x}, t, a, \mathbf{p}^*) = \mathbf{y}(\mathbf{x}, t, a, \mathbf{p})$ implies $p = p^*$. A parameter p is said to be *non-uniquely (or locally) structurally identifiable* if for almost any $\mathbf{p}^*$ and almost all initial conditions, the equation $\mathbf{y}(\mathbf{x}, t, a, \mathbf{p}^*) = \mathbf{y}(\mathbf{x}, t, a, \mathbf{p})$ implies that p has a finite number of solutions.

Definition 2 Similarly, a model of the form (1) is said to be *uniquely* (respectively, *non-uniquely*) *structurally identifiable* for a given choice of output $\mathbf{y}$ if every parameter is uniquely (respectively, non-uniquely) structurally identifiable, i.e. for almost every value $\mathbf{p}^*$ and almost all initial conditions, the equation $\mathbf{y}(\mathbf{x}, t, a, \mathbf{p}^*) = \mathbf{y}(\mathbf{x}, t, a, \mathbf{p})$ has only one solution, $\mathbf{p} = \mathbf{p}^*$ (respectively, finitely many solutions). Equivalently, a model is uniquely (respectively, non-uniquely) structurally identifiable for a given output if and only if the map Φ is injective (respectively, finite-to-one) almost everywhere, i.e. if there exists a unique set of parameter values $\mathbf{p}^*$ which yields a given trajectory $\mathbf{y}(\mathbf{x}, t, a, \mathbf{p}^*)$ almost everywhere.

Definition 3 A model of the form (1) is said to be *structurally unidentifiable* for a given choice of output $\mathbf{y}$ if it is not uniquely or non-uniquely identifiable, namely if at least one parameter is structurally unidentifiable, i.e. for a set of $\mathbf{p}^*$ and initial conditions of nonzero measure, the equation $\mathbf{y}(\mathbf{x}, t, a, \mathbf{p}^*) = \mathbf{y}(\mathbf{x}, t, a, \mathbf{p})$ has infinitely many solutions. Equivalently, a model is structurally unidentifiable for a given output if and only if the map Φ is infinite-to-one on a set of positive measure.

We note that because some specific parameter values or initial conditions may render an otherwise identifiable model unidentifiable (e.g. if all parameters or initial conditions are zero), we define structural identifiability generically, i.e., for almost all parameter values and initial conditions (Audoly et al. 2001; Pia Saccomani et al. 2003). Because the definitions apply for almost all parameter values and initial conditions, there may be some initial conditions of $\mathbf{x}$, or trajectories of $\mathbf{x}$ (corresponding to a measure zero set of parameters and initial conditions) for which an otherwise identifiable model may be unidentifiable. We additionally note that the system's initial and boundary conditions may also depend on unknown parameters.

2.2 Overview of the differential algebra approach

To determine structural identifiability of a model with defined outputs, we extend the differential algebra approach (Ljung and Glad 1994; Audoly et al. 2001) to the age-structured PDE case. This approach relies on converting the model system into a system of input-output equations, that is, a set of n_y monic polynomial equations in terms of the known variables $\mathbf{y}$, $\mathbf{u}$, and their derivatives, with rational coefficients in the parameters $\mathbf{p}$. The input-output equations represent an implicit form of the model map, and as a set of differential equations, they are input-output equivalent to the original system. This means that for the same inputs and parameter values, the

equations generate the same output trajectories (Eisenberg 2019). We then consider the map from the input parameter space to the coefficients of the resulting monic polynomial.

A key insight of the differential algebra approach in the rational ODE case is that injectivity of the model map can be evaluated by considering the coefficient map, i.e. the map from parameters to coefficients of the input-output equations (Audoly et al. 2001; Ljung and Glad 1994). In the rational ODE case, the coefficient map typically provides complete identifiability information for the model. If this map is injective, then the model is uniquely structurally identifiable. If the model is unidentifiable, this map determines the identifiable combinations of parameters. For example, while some model parameters may not be independently identifiable, the product or sum of two or more model parameters may be identifiable. This may enable the model to be reparameterized in a way that allows the model to become structurally identifiable (Chappell and Gunn 1998; Cole et al. 2010; Meshkat et al. 2009, 2011, 2014; Meshkat and Sullivant 2014). It should be noted that a set of identifiable combinations is not necessarily unique – there may be many sets of parameter combinations that are equivalent. For example, the sets of identifiable combinations $\{ab, bc\}$ and $\{ab, a/c\}$ are equivalent since $(ab)/(bc) = a/c$.

The differential algebra approach for ODE models typically uses characteristic sets to calculate the input-output equations (Audoly et al. 2001; Ljung and Glad 1994). Characteristic sets are a differential analog of the Gröbner basis (Ritt 1950) based on differential polynomial psuedodivision. For more details on the characteristic set approach, the reader is referred to (Audoly et al. 2001; Ljung and Glad 1994; Pia Saccomani et al. 2003; Hong et al. 2020; Eisenberg 2019).

This methodology has previously been applied to ODE systems, where derivatives are taken with respect to only one variable. In this work, we generalize this methodology to PDE systems of the form in (1); in particular, we consider age-structured systems where derivatives are taken with respect to both time and age. We show that the same framework can be applied to these systems, with the additional consideration of boundary conditions. In age-structured systems, a boundary condition must be specified at $a = 0$, usually in the form of a Dirichlet boundary condition. Thus, in addition to performing identifiability analysis on the model equations, one must also consider the equations evaluated at $a = 0$ to explicitly include information at the boundary. This is analogous to methods for determining identifiability of initial conditions in ODE systems (Pia Saccomani et al. 2003).

2.3 Extending the differential algebra approach to age structured PDE models

In this section, we discuss our generalization of the differential algebra approach to the PDE case. For PDE models, the primacy of the differential ideal generated by the model and measurement equation differential polynomials is not necessarily guaranteed (at least, to our knowledge it has not been shown to be the case), meaning that the characteristic set may not be unique. Thus, instead of a characteristic set based approach, we will use a simpler, substitution-based approach based on previous work by Eisenberg (2019) (similar to that used in Eisenberg et al. (2013)). In this approach,

we use basic substitution (which, as long as one is careful not to divide by quantities that could be zero, etc., does not change the underlying solutions to the system) to eliminate the unobserved variables, resulting in a monic, reduced system of differential equations which is input-output equivalent to the original system (Eisenberg 2019). This input-output equivalent system has equivalent identifiability properties to the original, allowing us to use it to evaluate identifiability of a certain class of PDE systems.

As is typical in ODE systems, once we have a set of input-output equations (i.e. monic differential polynomials only in terms of the known variables $\mathbf{y}$, $\mathbf{u}$ and their derivatives with respect to either independent variable and the parameters $\mathbf{p}$), we assess identifiability based on the coefficients of the monomial terms. As in the ODE case, this requires that the input-output equations be in some sense minimal or fully reduced, with the monomials of the input-output equations linearly independent (Ovchinnikov et al. 2021; Bearup et al. 2013). Given these reduced input output equations, we assume that our observed trajectory contains sufficiently many algebraically independent points that we can successfully identify the coefficients of the input-output equations. In other words, from a given observed solution of $\mathbf{y}$ over age and time, we assume the observed solution contains an arbitrary number of distinct points at which the input-output equations can be evaluated to solve for the input-output coefficients (in particular, more distinct points than the number of coefficients). In practice most trajectories meet this criterion, although some care must be taken to avoid systems or measure zero sets of initial conditions for which the solutions are at steady state, or have key parameters equal to zero, etc. (Hong et al. 2020). With the coefficient-identifiability assumption (Eisenberg 2019), the rest of the differential algebra approach follows as with the ODE case, and is described below.

Procedurally, the process of performing identifiability analysis for PDEs is almost identical to the ODE case, though substitutions may be made more complicated by the presence of an additional derivative. To obtain input-output equations from model equations, a typical first step is to solve the output equations for state variables and, using substitution, eliminate these state variables and their derivatives from the model equations. For simplicity, all nonzero terms in the model equations are moved to the right-hand side. Remaining state variables can then be eliminated by solving one or more of the model equations. The particular substitutions required will of course vary depending on the model and output equations. The goal is to obtain a system of polynomial equations in terms of the outputs and their derivatives, with all state variables eliminated, whose coefficients are in terms of the model parameters. In the case that square roots arise, for example, they must be eliminated by multiplying by the conjugate.

Once input-output polynomial equations are obtained, we divide each equation by the coefficient of its leading term to obtain a monic polynomial (see below for details about the ranking of terms). From these monic polynomials, we obtain the set of polynomial coefficients that defines the coefficient map. This map, as in the ODE case, can be used to determine complete identifiability information for the model equations.

Note that thus far, we have not considered any information about the initial or boundary conditions—an important consideration of identifiability, In particular, analysis of

the boundary conditions is a key component of identifiability analysis for PDEs that does not occur in the ODE case. In the case of age-structured systems, there is a single boundary condition at $a = 0$. Thus, we treat the boundary condition as similar to an initial condition for ODEs (see e.g. the methods in Pia Saccomani et al. (2003) for the ODE case). If information is known about the boundary conditions, such as values or functional forms, this information can be substituted into the model equations evaluated at $a = 0$ to obtain *boundary condition equations*. Since the values of model outputs are known at $a = 0$, the same identifiability analysis as above can be applied to boundary condition equations. This may reveal additional identifiability information about the model or any parameters that are present in the boundary conditions.

We note that, as in the initial condition case, if identifiability is evaluated without considering boundary conditions, this implicitly assumes that the boundary conditions are generic and unknown. However, there are some important aspects of the boundary conditions which are more subtle than the ODE case of initial conditions, and may warrant further exploration, such as that boundary conditions still may vary over time. Exploring how different boundary condition types may affect identifiability would be an important direction for future work.

Our procedure for performing structural identifiability analysis, which can be applied to both ODE and age-structured PDE systems, is outlined in Fig. 1. In our structural identifiability analyses herein, we utilize Mathematica to perform the required algebraic manipulations and evaluate map injectivity. All Mathematica notebooks are provided in our code repository at https://github.com/epimath/age-pde-identifiability.

2.4 Practical identifiability analysis

Practical identifiability was evaluated using the Fisher Information Matrix (FIM) and profile likelihoods (Jacquez and Perry 1990; Raue et al. 2009; Venzon and Moolgavkar 1988; Eisenberg and Hayashi 2014). Here, we follow the methodology laid out in (Eisenberg and Hayashi 2014).

For ODE models when evaluating identifiability, the FIM is computed by $F = X^T X$, where X is the sensitivity matrix defined by $X_{(i,j)} = \frac{\partial y}{\partial p_i}(t_j)$ for parameters $p_1, ..., p_n$ and time points $t_1, ..., t_m$. This form is a simplified form of the FIM for the case of normally distributed measurement error, but it is frequently used for evaluating structural identifiability due to its easily interpretable form and clear connection to the sensitivity matrix X (the two matrices share the same rank, and identifiable parameter combinations can be naturally understood through the sensitivity matrix as compensation between parameters will be reflected in dependences in the columns). For an age structured system, we define $F = X^T X$ where the sensitivity matrix is defined by $X_{(i,j)} = \frac{\partial y}{\partial p_i}((a, t)_j)$ where $(a, t)_j$ for $j = 1, ..., m$ are pairs of age and time points that form a grid. Here, y denotes the observable model output. For multiple model outputs (e.g., y_1 and y_2), the matrix X can be defined by concatenation (i.e., $X = [X_1; X_2]$). The rank of F indicates the number of identifiable parameters and parameter combinations. If F is singular, the model is structurally unidentifiable; if F is close to singular (meaning a large condition number or a determinant near zero, e.g. below machine precision), the model may be practically unidentifiable. The FIM can
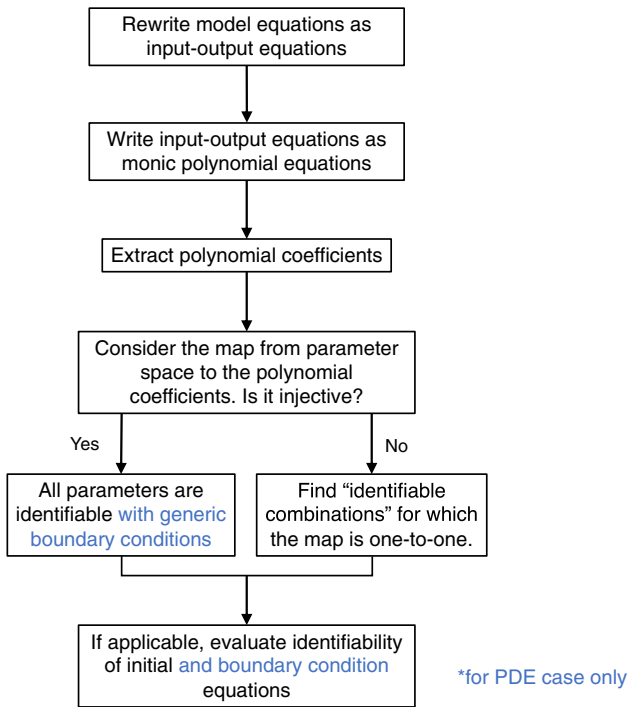
Fig. 1 Flowchart describing the methodology for identifiability analysis of ODE and age-structured PDE models based on using a differential algebra approach
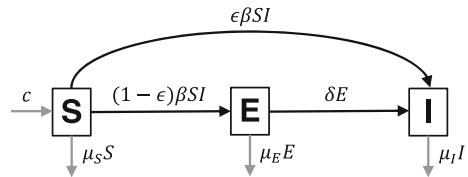
also be used to determine which parameter subsets are identifiable or unidentifiable (Eisenberg and Hayashi 2014).

We compute profile likelihoods by fixing a particular parameter p_i and fitting all remaining parameters $p_{j \neq i}$ to maximize a likelihood function. We define the likelihood by assuming constant, normally distributed measurement error with standard deviation equal to 10% of the mean of the data, and mean equal to the model trajectory. Maximizing the likelihood function is equivalent to minimizing a cost function based on the negative log likelihood (in this case equivalent to least squares). This is performed at many values across the range of values for p_i. The profile likelihood for p_i is given by the maximum values of the likelihood function across the range of values for p_i. If the profile likelihood is flat, p_i is structurally unidentifiable; if it is sufficiently shallow, p_i is practically unidentifiable. For more details on profile likelihood approaches to identifiability, the reader is referred to (Raue et al. 2009; Venzon and Moolgavkar 1988).

In our analyses herein, we solved the age-structured PDEs numerically using Method of Lines, using a first order finite difference approximation of the spatial derivative, and using Matlab's ODE solver *ode45*.

Parameter	Definition
c	birth rate
β	transmission rate
ϵ	probability of primary infection
δ	reactivation rate
μ_S, μ_E, μ_I	death rates

3 SEI model for TB epidemiology

In several case studies below, we consider a simple mathematical model of the epidemiology of tuberculosis (TB). TB is transmitted via the aerosolized bacterium *Mycobacterium tuberculosis* from person to person via lung inhalation. TB is unique from many infectious diseases in that most people develop a *latent* infection, meaning that they present no symptoms and cannot transmit the disease, and the infection can remain latent for an extended period of time. Thus, our model consists of three mutually exclusive compartments: Susceptible (S), Exposed/latent (E), and Infectious (I). Through contact with infectious individuals, susceptible individuals (called susceptibles) may progress to the exposed class (i.e. latent infection) or the infectious class (i.e. primary infection). This is slightly different from the standard SEI model, in that some individuals bypass the exposed class and immediately become infectious, but is similar to what has been used to model tuberculosis epidemics previously (Ozcaglar et al. 2012; Guzzetta et al. 2011; Renardy and Kirschner 2019; Vynnycky and Fine 1997). Individuals who are exposed/latent may progress to the infectious class where they can transmit bacteria. We assume all individuals are born into the susceptible class, and each compartment has a compartment-specific death rate.

A schematic for this model is shown in Fig. 2. This schematic can be translated into a mathematical or computational model by representing it as a system of ordinary or partial differential equations (as has been done by, e.g., Guzzetta et al. (2011); Renardy and Kirschner (2019); Vynnycky and Fine (1997)), a stochastic model, or an individual-based model (as has been done by, e.g., Guzzetta et al. (2011); Renardy and Kirschner (2020)). Here, we will first consider the model as a system of ODEs and then extend the results to age-structured PDEs.

4 Case study 1: density-dependent transmission

For simplicity, we first demonstrate the identifiability analysis framework on a model with density-dependent transmission, i.e. transmission is proportional to SI rather than $S\frac{I}{N}$. In this case, the per-individual rate of contact increases as the population density

increases. This is not necessarily a realistic model for TB, since TB is often spread only among close contacts, but simplifies the differential algebra for the purposes of demonstration.

In order to perform parameter identifiability analysis, we must first define what the outputs (i.e., the quantities for which we could obtain reliable, biologically relevant time series data) of the specific model are. For all of the following examples, we will assume that we can observe a fraction of the exposed individuals, i.e. $y_E = k_E E$ where k_E is the reporting/detection rate. This type of data could be obtained by, for example, contact tracing to identify individuals who have been exposed to TB disease, or testing a proportion of the population for exposure. This is a useful practice even in areas with a low TB burden such as the US, since detection and treatment of latent infection is crucial for TB elimination (LoBue and Mermin 2017). Since latent disease is asymptomatic, it is highly unlikely that all exposed/latent individuals will be identified, and thus we assume instead that only a fraction k_E of the exposed individuals are identified.

In addition, we may be able to observe all or some of either all active infections (prevalence) or new active infections (incidence). The relevant output function depends on the reporting and surveillance data for the region of interest. For example, if all infectious cases are monitored and reported then these data would represent total prevalence. If not all cases are detected due to misdiagnosis, for example, then the data would represent only a fraction of the total prevalence. If new infectious cases are reported upon diagnosis but then lost to follow-up, then incidence is an observable output while prevalence is not. Again, since it is possible that not all infected individuals will be diagnosed, we may only be able to observe a fraction of incidence. We present here results for observing prevalence over time, but the identifiability analysis of model parameters for incidence would follow similarly.

We assume that $y_I = k_I I$, i.e. that we are able to obtain data for a fraction of the total number of infections at any given time. We assume no other information about the system parameters or initial conditions. The case of incidence is more complex due to nonlinearities, but the methodology presented here can still be applied, and we provide such an example in our code repository.

4.1 ODE SEI model with density-dependent transmission

Though identifiability analysis for ODE systems is well established, we include the technical details here for completeness and for comparison when we move to the PDE system. The model can be written as a system of ODEs:

$$
\begin{aligned}
\frac{dS}{dt} &= c - \beta S I - \mu_S S \\
\frac{dE}{dt} &= (1-\epsilon)\beta S I - \delta E - \mu_E E \\
\frac{dI}{dt} &= \epsilon \beta S I + \delta E - \mu_I I
\end{aligned}
\tag{2}
$$

We consider the case of generic (unknown) initial conditions. We do not intiially assume a fixed population, and thus the birth rate and death rates are independent. We begin by rewriting the system of Eq. (2) in terms of the outputs y_E and y_I. For ease of notation, we adopt the notation $y' = \frac{dy}{dt}$. Using the substitutions $E = \frac{y_E}{k_E}$ and $I = \frac{y_I}{k_I}$, we obtain the following equations:

$$S' = c - \beta S \frac{y_I}{k_I} - \mu_S S$$

$$\frac{y_E'}{k_E} = (1 - \epsilon)\beta S \frac{y_I}{k_I} - \delta \frac{y_E}{k_E} - \mu_E \frac{y_E}{k_E}$$

$$\frac{y_I'}{k_I} = \epsilon \beta S \frac{y_I}{k_I} + \delta \frac{y_E}{k_E} - \mu_I y_I.$$

It now remains to eliminate S from these equations so that we obtain equations in terms of solely the model parameters and outputs. This can be done by solving for S in the equation for $\frac{dy_I}{dt}$:

$$S = \frac{k_E(y_I' + \mu_I y_I) - \delta k_I y_E}{\beta k_E \epsilon y_I}.$$

We note that we divide by y_I in this example for simplicity; however the same input-output equations can be obtained without dividing, using Ritt's pseudodivision (Ritt 1950) (just with more involved algebraic computations required). This expression can then be substituted into the remaining two equations to obtain two input-output equations that are rational equations in terms of the outputs and their derivatives. By multiplying by a common denominator and moving terms to one side, these input-output equations can be rewritten as the following polynomial equations.

$$0 = k_I^2 \delta \mu_S y_E y_I + (ck_E k_I \beta \epsilon - k_E k_I \mu_I \mu_S) y_I^2 + k_I \beta \delta y_E y_I^2$$
$$\quad - k_E \beta \mu_I y_I^3 + k_I^2 \delta y_I y_E' - k_I^2 \delta y_E y_I' - k_I k_E \mu_S y_I y_I'$$
$$\quad - k_E \beta y_I^2 y_I' + k_E k_I y_I'^2 - k_E k_I y_I y_I''$$
$$0 = (-k_I \delta - k_I \epsilon \mu_E) y_E + (k_E \mu_I - k_E \epsilon \mu_I) y_I - k_I \epsilon y_E'$$
$$\quad + (k_E - k_E \epsilon) y_I'$$

Dividing the first equation by $k_E k_I$ and the second equation by $k_E - k_E \epsilon$ to form monic polynomials, we get the following set of polynomial coefficients.

$$\left\{ -1, 1, \frac{\beta}{k_I}, -\frac{\delta k_I}{k_E}, \frac{\delta k_I}{k_E}, -\frac{\beta \delta}{k_E}, \frac{k_I \epsilon}{k_E(\epsilon - 1)}, \frac{k_I(\delta + \mu_E \epsilon)}{k_E(\epsilon - 1)}, \right.$$
$$\left. \frac{\beta \mu_I}{k_I}, \mu_I, \mu_S, -\frac{\delta k_I \mu_S}{k_E}, \mu_I \mu_S - \beta c \epsilon \right\}$$

We consider this set of coefficients as a function of the model parameters, and we seek to determine whether this map is injective. If it is not, then we consider whether

it is injective as a function of some parameter combinations. We find that μ_S and μ_I are structurally identifiable. We then simplified the above set by removing duplicates up to sign (e.g. $\delta k_I/k_E$ and $-\delta k_I/k_E$ are redundant), subtracting off known quantities where we can (e.g. if $\mu_I \mu_S - \beta c\epsilon$ is identifiable and μ_I and μ_S are identifiable, then so is $\beta c\epsilon$, a simpler identifiable combination), and identifying coefficients that are combinations of previously established known quantities (e.g. $\delta k_I/k_E$ can be found by dividing coefficients β/k_I and $\beta\delta/k_E$). Through simplification and substitution of the above set, we find that other identifiable parameter combinations are

$$\left\{ \frac{\beta}{k_I}, \beta c\epsilon, \frac{\beta\delta}{k_E}, \frac{\delta(-1+\epsilon)}{\epsilon}, \mu_E + \delta \right\}.$$

If we further assume that $\mu_S = \mu_E$, then δ and ϵ also become identifiable while β, c, k_E, and k_I remain unidentifiable. However, the following combinations are identifiable: $\{\frac{\beta}{k_E}, \frac{\beta}{k_I}, \beta c\}$. Consequently, if any one of β, c, k_E, or k_I is known, then the rest become identifiable.

4.2 Age-structured PDE model with density-dependent transmission and constant parameters

Now that identifiability results are established for the ODE representation of the SEI model 2, we extend these results to a more complex PDE model that captures the same epidemiological setting (SEI), but includes age structure. We incorporate age structure into the model by adding a transport term to capture population aging. To simplify the model, we assume that transmission occurs between individuals of the same age, so that the transmission term is proportional to $S(t, a)I(t, a)$ for any given age a and time t. While this is a significant simplification, data has shown that people tend to interact primarily with others in their age group (Mossong et al. 2008). The model thus becomes

$$\begin{aligned}
\frac{\partial S}{\partial t} + \frac{\partial S}{\partial a} &= -\beta SI - \mu_S S \\
\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a} &= (1-\epsilon)\beta SI - \delta E - \mu_E E \\
\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a} &= \epsilon\beta SI + \delta E - \mu_I I
\end{aligned} \tag{3}$$

for $t > 0$ and $a \in [0, \infty)$. Note that all state variables are now functions of both time and age. The boundary conditions are $S(t, 0) = c$, $E(t, 0) = I(t, 0) = 0$, and we begin by assuming initial conditions are generic. Note that the birth rate is now incorporated into the boundary condition rather than into the differential equation for S. We begin by assuming that model parameters are constants, independent of age.

4.2.1 Identifiability of model equations with generic boundary conditions

We first consider identifiability of the model parameters without explicitly accounting for the presence of boundary conditions, meaning that we assume there is no *a priori* knowledge about the boundary conditions and that they cannot be observed. We call this case *generic* boundary conditions. To perform the identifiability analysis, we utilize the same framework as we did for the above ODE system analysis. That is, we rewrite the model system as a set of polynomial equations in terms of the outputs and their partial derivatives, and then consider the map from parameter space to the polynomial coefficients.

We adopt the notation $y^{(i,j)} = \frac{\partial^{i+j} y}{\partial t^i \partial a^j}$. Using the substitutions $E = \frac{y_E}{k_E}$ and $I = \frac{y_I}{k_I}$, as in the ODE case, we obtain the following equations:

$$S^{(1,0)} + S^{(0,1)} = -\beta S \frac{y_I}{k_I} - \mu_S S$$

$$\frac{y_E^{(0,1)}}{k_E} + \frac{y_E^{(1,0)}}{k_E} = (1-\epsilon)\beta S \frac{y_I}{k_I} - \delta \frac{y_E}{k_E} - \mu_E \frac{y_E}{k_E}$$

$$\frac{y_I^{(0,1)}}{k_I} + \frac{y_I^{(1,0)}}{k_I} = \epsilon \beta S \frac{y_I}{k_I} + \delta \frac{y_E}{k_E} - \mu_I y_I.$$

We now eliminate S to obtain equations in terms of only the model parameters and outputs. This can be done by solving for S in the third equation:

$$S = \frac{k_E \left(y_I^{(0,1)} + y_I^{(1,0)} + \mu_I y_I \right) - \delta k_I y_E}{\beta k_E \epsilon y_I}.$$

We assume here that generically $y_I > 0$ to avoid division by zero, although one can also obtain the same input-output equations without dividing by y_I by solving instead for $S y_I$ and substituting that into the equations (but this requires more extensive algebraic calculations and multiplication by y_I for the first equation). As before, this expression can then be substituted into the remaining two equations to obtain two input-output equations that are rational equations of the outputs and their derivatives. By multiplying by a common denominator and moving terms to one side, these input-output equations can be rewritten as the following polynomial equations.

$$\begin{aligned}
0 = {} & -k_E k_I \mu_S y_I^{(0,1)} y_I - k_E k_I \mu_S y_I^{(1,0)} y_I - k_E k_I y_I^{(0,2)} y_I \\
& - 2k_E k_I y_I^{(1,1)} y_I - k_E k_I y_I^{(2,0)} y_I + k_E k_I (y_I^{(0,1)})^2 \\
& + k_E k_I (y_I^{(1,0)})^2 + 2k_E k_I y_I^{(0,1)} y_I^{(1,0)} - k_E k_I \mu_I \mu_S y_I^2 \\
& - \beta k_E y_I^{(0,1)} y_I^2 - \beta k_E y_I^{(1,0)} y_I^2 - \beta k_E \mu_I y_I^3 \\
& + \delta k_I^2 y_E^{(0,1)} y_I + \delta k_I^2 y_E^{(1,0)} y_I - \delta k_I^2 y_E y_I^{(0,1)} \\
& - \delta k_I^2 y_E y_I^{(1,0)} + \delta k_I^2 \mu_S y_E y_I + \beta \delta k_I y_E y_I^2 \\
0 = {} & -k_E \epsilon y_I^{(0,1)} - k_E \epsilon y_I^{(1,0)} + k_E y_I^{(0,1)} + k_E y_I^{(1,0)}
\end{aligned}$$

$$+ k_E \mu_I y_I - k_E \mu_I \epsilon y_I - k_I \epsilon y_E^{(0,1)} - k_I \epsilon y_E^{(1,0)}$$
$$- \delta k_I y_E - k_I \mu_E \epsilon y_E$$

We choose, without loss of generality, derivatives with respect to time to be higher ranked than those with respect to age for our ranking of terms within the polynomial, and let y_I be higher ranked than y_E, yielding the ranking $y_E < y_I < y_E^{(0,1)} < y_I^{(0,1)} < y_E^{(1,0)} < y_I^{(1,0)} < y_E^{(0,2)} < y_I^{(0,2)} < y_E^{(1,1)} < y_I^{(1,1)} < \dots$. Dividing the first equation by $k_E k_I$ and the second equation by $k_E - k_E \epsilon$ to form monic polynomials, we get the following set of polynomial coefficients.

$$\left\{ -2, -1, 1, 2, \frac{\beta}{k_I}, -\frac{k_I \delta}{k_E}, \frac{k_I \delta}{k_E}, \frac{k_I \epsilon}{k_E(\epsilon - 1)}, \frac{k_I(\delta + \mu_E \epsilon)}{k_E(\epsilon - 1)}, \right.$$
$$\left. \frac{\beta \mu_I}{k_I}, \mu_I, \mu_S, -\frac{\delta k_I \mu_S}{k_E}, \mu_I \mu_S - \frac{\beta \delta}{k_E} \right\}$$

We find that, as was the case for the ODE model, μ_S, and μ_I are structurally identifiable. Through simplification and substitution, we find that other identifiable parameter combinations are

$$\left\{ \frac{\beta}{k_I}, \frac{\delta k_I}{k_E}, \frac{\delta(\epsilon - 1)}{\epsilon}, \mu_E + \delta \right\}.$$

If $\mu_S = \mu_E$ then δ and ϵ become identifiable as well as the combinations $\left\{ \frac{\beta}{k_E}, \frac{\beta}{k_I} \right\}$.

4.2.2 Incorporating non-generic boundary conditions

In our model, the boundary conditions are in fact not generic, but rather are known to be $S(t, 0) = c$, $E(t, 0) = I(t, 0) = 0$, where c represents the birth rate. Here, the boundary condition contains an unknown model parameter, c. To determine the identifiability of the boundary condition parameter c, we perform the same analysis as above but with the model equations evaluated at $a = 0$ (boundary condition equations). However, to avoid reducing equations to zero, we let the boundary conditions be set to an extremely small but assumed known quantity that should not affect the model simulations significantly but enables us to evaluate structural identifiability in this case (e.g. reflecting a small amount of congenital tuberculosis in the population (Cantwell et al. 1994)). Then, noting that $S(t, 0)$ is constant and thus $\frac{\partial S}{\partial t}(t, 0) = 0$, the relevant equations are

$$\frac{\partial S}{\partial a}(t, 0) = -\beta c I(t, 0) - \mu_S c$$
$$\frac{\partial E}{\partial t}(t, 0) + \frac{\partial E}{\partial a}(t, 0) = (1 - \epsilon)\beta c I(t, 0) - \delta E(t, 0) - \mu_E E(t, 0) \qquad (4)$$
$$\frac{\partial I}{\partial t}(t, 0) + \frac{\partial I}{\partial a}(t, 0) = \epsilon \beta c I(t, 0) + \delta E(t, 0) - \mu_I I(t, 0)$$

By transforming the system (4) to a set of polynomial equations we obtain the following polynomial coefficients:

$$\left\{ 1, -\frac{k_I \delta}{k_E}, -\frac{ck_E \beta - ck_E \beta \epsilon}{k_I}, \delta + \mu_E, \mu_I - c\beta\epsilon \right\}$$

Since μ_I is identifiable from the model equations, we thus find that $\beta c\epsilon$ is an identifiable combination, which is consistent with the ODE model. Thus, the full set of identifiable parameters and combinations for this model is

$$\left\{ \mu_S, \mu_I, \frac{\beta}{k_I}, \beta c\epsilon, \frac{\delta k_I}{k_E}, \frac{\delta(\epsilon - 1)}{\epsilon}, \mu_E + \delta \right\}$$

which is equivalent to the set of identifiable combinations for the ODE model.

4.2.3 Identifiability of initial conditions

With generic initial conditions, we do not gain any additional identifiability results. However, if we assume that certain information is known, such as the total population size N at the initial time, then further model parameters can become identifiable. Here, we assume that the initial population size as well as its derivatives with respect to time and age are known for all ages (noting that the age-specific population could potentially be determined from census data, and derivatives perhaps roughly estimated from multiple years of census data). Under these conditions, we may solve for the initial susceptible population $S(0, a)$ as $S(0, a) = N(0, a) - E(0, a) - I(0, a)$, where N represents the total population size. Similarly, since we assume the derivatives of N are known at $t = 0$, we may solve for the derivatives of S at $t = 0$ in terms of the derivatives of E and I. It is thus straightforward to solve for S and its derivatives at $t = 0$ in terms of the outputs y_E and y_I.

To determine parameter identifiability from the initial condition with this additional knowledge, we evaluate the model equations at $t = 0$ and perform identifiability analysis analogous to the previous analysis for the boundary conditions. We obtain the following set of polynomial coefficients:

$$\left\{ 1, \frac{k_I}{k_E}, \frac{\beta}{k_E}, \frac{\beta}{k_I}, -\frac{\delta k_I}{k_E}, -\frac{\beta k_E(\epsilon - 1)}{k_I^2}, \frac{\beta k_E(\epsilon - 1)}{k_I}, \frac{\beta\epsilon}{k_E}, \right.$$
$$\left. \frac{\beta\epsilon}{k_I}, \frac{\beta - \beta\epsilon}{k_I}, \delta + \mu_E, \mu_I - \beta\epsilon, \frac{k_I \mu_S}{k_E}, \mu_S - \beta, -k_I(\mu_S + 2) \right\}$$

From these coefficients, we find that $\beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, k_E$, and k_I are all identifiable. Combining this with the previously established identifiability of $c\beta\epsilon$ from the boundary equations, we find that *all parameters are identifiable* if the initial population size and its derivatives are known.

5 Case study 2: frequency-dependent transmission

While density-dependent transmission is convenient for simplicity, it is not generally a realistic model for TB. Rather, most models for TB epidemiology use frequency-dependent transmission models wherein the transmission term is proportional to $S\frac{I}{N}$. In this section we present identifiability results for the SEI model with frequency-dependent transmission, assuming that our output functions are $y_E = k_E E$ and $y_I = k_I I$. In addition to the ODE and constant-parameter PDE models, we will also consider variations including immigration and age-dependent death rates. The full details of the analysis are cumbersome and thus are not provided here, but can be found in the code repository.

5.1 ODE SEI model with frequency-dependent transmission

For completeness, we include here identifiability results for the ODE version of the model. This model can be written as the following system of equations:

$$\frac{dS}{dt} = c - \beta S \frac{I}{N} - \mu_S S$$
$$\frac{dE}{dt} = (1 - \epsilon)\beta S \frac{I}{N} - \delta E - \mu_E E \tag{5}$$
$$\frac{dI}{dt} = \epsilon \beta S \frac{I}{N} + \delta E - \mu_I I$$

where $N(t) = S(t) + E(t) + I(t)$. We find that the set of identifiable parameters and parameter combinations for this model is

$$\left\{ \beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, ck_I, \frac{k_E}{k_I} \right\}.$$

5.2 Age-structured PDE SEI model with frequency-dependent transmission and constant parameters

$$\frac{\partial S}{\partial t} + \frac{\partial S}{\partial a} = -\beta S \frac{I}{N} - \mu_S S$$
$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a} = (1 - \epsilon)\beta S \frac{I}{N} - \delta E - \mu_E E \tag{6}$$
$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a} = \epsilon \beta S \frac{I}{N} + \delta E - \mu_I I$$

with boundary conditions $S(t, 0) = c$, $E(t, 0) = I(t, 0) = 0$. From the model equations, we find that the parameters and combinations $\{\beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, \frac{k_E}{k_I}\}$ are identifiable. We further find from the boundary equations that ck_I is identifiable. Thus, the full set of identifiable parameters and combinations for this model is

$$\left\{ \beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, ck_I, \frac{k_E}{k_I} \right\}$$

which is consistent with the ODE model. If the initial age distribution of the total population and its derivatives were known (as in Sect. 4.2.3 for the previous case study), then all parameters in the model would be identifiable.

5.3 SEI model with frequency-dependent transmission and immigration

The models considered thus far have included influx of new individuals only via birth. However, it may be necessary to also consider immigration when studying epidemics in fluid populations. If immigration is included in the model, with the assumptions that the immigration rate is constant over time and all immigrants are susceptible, the ODE system becomes

$$
\begin{aligned}
\frac{dS}{dt} &= m + c - \beta S \frac{I}{N} - \mu_S S \\
\frac{dE}{dt} &= (1 - \epsilon)\beta S \frac{I}{N} - \delta E - \mu_E E \\
\frac{dI}{dt} &= \epsilon \beta S \frac{I}{N} + \delta E - \mu_I I
\end{aligned}
\tag{7}
$$

where m is the immigration rate. If we assume that the age distribution $\theta(a)$ of immigrants is known, then the corresponding age-structured PDE system becomes

$$
\begin{aligned}
\frac{\partial S}{\partial t} + \frac{\partial S}{\partial a} &= m\theta(a) - \beta S \frac{I}{N} - \mu_S S \\
\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a} &= (1 - \epsilon)\beta S \frac{I}{N} - \delta E - \mu_E E \\
\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a} &= \epsilon \beta S \frac{I}{N} + \delta E - \mu_I I
\end{aligned}
\tag{8}
$$

where $\theta(a)$ is some known function such that $\theta(0) = 0$ and $\int_0^\infty \theta(a)da = 1$, with boundary conditions $S(t, 0) = c$ and $E(t, 0) = I(t, 0) = 0$.

The identifiability results for the ODE system (7) remain the same as before but with c replaced by $c + m$, i.e., the identifiable parameters and combinations are

$$\left\{ \beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, (c + m)k_I, \frac{k_E}{k_I} \right\}.$$

For the PDE system (8), we get the following identifiable parameters and combinations from the model and boundary equations:

$$\left\{ \beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, ck_I, mk_I, \frac{k_E}{k_I} \right\}.$$

Thus, in this case the identifiability results for the PDE model differ from those for the ODE model in that the combinations ck_I and mk_I are both identifiable, whereas in the ODE model only the sum $(c + m)k_I$ is identifiable. In other words, if all immigrants are susceptible then in an age-structured model it is possible to differentiate between birth and immigration, whereas in an ODE model it is not.

5.4 Age-structured PDE SEI model with frequency dependent transmission and age-dependent parameters

A major benefit of age-structured models is the ability to incorporate age-dependent parameters such as death rates, contact/transmission rates, and reactivation rates to make the model more realistic. Here, we consider the case of age-dependent death rates. There are several ways to introduce age dependence to the death rate: one could assume a specific functional form (e.g., piecewise-constant, linear, exponential), that the function belongs to a particular family (e.g., polynomials, splines), or that the function is completely arbitrary with some conditions on smoothness. In this section, we present identifiability results for four cases: (1) piecewise-constant death rate functions for all three compartments, (2) a single exponential death rate function for all compartments, (3) a single polynomial death rate function for all compartments, and (4) a single arbitrary function for all compartments. For simplicity, we will consider the SEI model without immigration.

5.4.1 PDE SEI model with piecewise-constant death rates case

We first consider the case that the death rates are piecewise-constant functions of age with discontinuities at ages $\{0 = a_0, a_1, a_2, ...a_n = A\}$, i.e. $\mu_j(a) = \mu_j^i$ for $a \in (a_i, a_{i+1})$ and $j = S, E, I$. This represents the case where death rates are known independently for different age groups. Then the system (6) becomes a sequence of PDE systems obeying the equations

$$\frac{\partial S}{\partial t} + \frac{\partial S}{\partial a} = -\beta S \frac{I}{N} - \mu_S^i S$$

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a} = (1 - \epsilon)\beta S \frac{I}{N} - \delta E - \mu_E^i E$$

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a} = \epsilon \beta S \frac{I}{N} + \delta E - \mu_I^i I$$

for $t > 0$ and $a \in (a_i, a_{i+1})$, $0 \le i < n$, with boundary conditions $S(t, 0) = c$, $E(t, 0) = I(t, 0) = 0$ and

$$S(t, a_i) = \lim_{a \to a_i^-} S(t, a)$$

$$E(t, a_i) = \lim_{a \to a_i^-} E(t, a)$$

$$I(t, a_i) = \lim_{a \to a_i^-} I(t, a)$$

for $0 < i < n$. We will assume that the locations of the discontinuities $a_1, ..., a_n$ are known, and that the sub-intervals are wide enough that the usual structural identifiability results hold.

Since the model equations remain essentially unchanged, structural identifiability of the model equations has already been established in Sect. 5.2. Thus, from the model equations we obtain that

$$\left\{ \beta, \delta, \epsilon, \mu_S^i, \mu_E^i, \mu_I^i, \frac{k_E}{k_I} \right\}$$

are identifiable for all i. From the boundary equations at $a = 0$ we again get that ck_I is identifiable, as in the constant parameter case. We gain no additional information about the parameter values from the boundary conditions at $a = a_i$ for $i > 0$.

5.4.2 PDE SEI model with exponential death rate case

One may also wish to use a continuous function to represent age-dependent death rates. One possibility is to use an exponential function $\mu(a) = \mu_0 \exp(\kappa a)$ so that death rate increases exponentially with age. For simplicity, we assume that death rates are the same for all three compartments. Then the model system can be written as

$$\frac{\partial S}{\partial t} + \frac{\partial S}{\partial a} = -\beta S \frac{I}{N} - \mu S$$

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a} = (1 - \epsilon)\beta S \frac{I}{N} - \delta E - \mu E$$

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a} = \epsilon \beta S \frac{I}{N} + \delta E - \mu I$$

$$\frac{d\mu}{da} = \kappa \mu$$

for $t > 0$ and $a \in [0, \infty)$, with boundary conditions $S(t, 0) = c$, $E(t, 0) = I(t, 0) = 0$, and $\mu(0) = \mu_0$. We write the system in this way rather than directly inserting $\mu(a) = \mu_0 \exp(\kappa a)$ so that we maintain a differential polynomial form.

To obtain an input-output equation from the model equations, we use the I equation to solve for S and the E equation to solve for $\mu(a)$ (note that these choices are somewhat arbitrary). Then by noting that $I = y_I/k_I$ and $E = y_E/k_E$ and substituting into the S and μ equations, we obtain rational equations in terms of the outputs and their derivatives. Thus, by multiplying by a common denominator, we obtain differential polynomials. We find that $\left\{ \beta, \delta, \epsilon, \kappa, \frac{k_E}{k_I} \right\}$ are the identifiable parameters and parameter combinations.

By performing similar analysis on the boundary equations, we find that μ_0 and ck_I are identifiable. Thus, the full set of identifiable parameters and combinations for this version of the model is

$$\left\{ \beta, \delta, \epsilon, \kappa, \mu_0, ck_I, \frac{k_E}{k_I} \right\}.$$

This shows that both parameters from the exponential death rate are identifiable. The identifiability of the remaining parameters is consistent with the results for the ODE model in Sect. 5.1.

5.4.3 PDE SEI model with polynomial death rate

Another option for a continuous death rate is to use a polynomial, $\mu(a) = \sum_{i=0}^{n} \kappa_i a^i$. For example, a polynomial may be fit to the available data for death rate by age group. Then the model equations become

$$\frac{\partial S}{\partial t} + \frac{\partial S}{\partial a} = -\beta S \frac{I}{N} - \left(\sum_{i=0}^{N} \kappa_i a^i \right) S$$

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a} = (1 - \epsilon) \beta S \frac{I}{N} - \delta E - \left(\sum_{i=0}^{N} \kappa_i a^i \right) E$$

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a} = \epsilon \beta S \frac{I}{N} + \delta E - \left(\sum_{i=0}^{N} \kappa_i a^i \right) I$$

for $t > 0$ and $a \in [0, \infty)$, with boundary conditions $S(t, 0) = c$ and $E(t, 0) = I(t, 0) = 0$. The age a is considered to be observable. By using the I equation to solve for S and substituting into the remaining equations, we obtain a set of input-output equations that are differential polynomials. We find that the corresponding monic polynomial contains the monomial terms $\kappa_i a^i y_I$ for all $i = 0, ..., n$. Thus, all coefficient parameters κ_i in $\mu(a)$ are identifiable and $\mu(a)$ can be uniquely recovered. We also find that $\left\{ \beta, \delta, \epsilon, \frac{k_E}{k_I} \right\}$ are identifiable from the model equations and ck_I is identifiable from the boundary equations.

5.4.4 PDE SEI model with arbitrary death rate function

The final and most general option we consider is to allow $\mu(a)$ to be an arbitrary function of a, i.e. the model equations become

$$\frac{\partial S}{\partial t} + \frac{\partial S}{\partial a} = -\beta S \frac{I}{N} - \mu(a) S$$

$$\frac{\partial E}{\partial t} + \frac{\partial E}{\partial a} = (1 - \epsilon) \beta S \frac{I}{N} - \delta E - \mu(a) E$$

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a} = \epsilon \beta S \frac{I}{N} + \delta E - \mu(a) I$$

for $t > 0$ and $a \in [0, \infty)$. In this case, we use the I equation to solve for S and then use the E equation to solve for $\mu(a)$. We obtain the following expression for $\mu(a)$ in

terms of the outputs and parameters:

$$
\begin{aligned}
\mu(a) &= -\frac{1}{k_E(\epsilon-1)y_I + k_I\epsilon y_E}\Big(\delta k_I y_E + k_I\epsilon y_E^{(1,0)} \\
&\quad + k_I\epsilon y_E^{(0,1)} + k_E(\epsilon-1)y_I^{(1,0)} + k_E(\epsilon-1)y_I^{(0,1)}\Big) \\
&= -\frac{1}{(\epsilon-1)y_I + \frac{k_I}{k_E}\epsilon y_E}\Big(\delta\frac{k_I}{k_E}y_E + \frac{k_I}{k_E}\epsilon y_E^{(1,0)} \\
&\quad + \frac{k_I}{k_E}\epsilon y_E^{(0,1)} + (\epsilon-1)y_I^{(1,0)} + (\epsilon-1)y_I^{(0,1)}\Big)
\end{aligned}
\tag{9}
$$

By substituting (9) into the remaining equation for S and performing the identifiability analysis, we again find that $\left\{\beta, \delta, \epsilon, \frac{k_I}{k_E}\right\}$ are identifiable. We note that the second derivatives of y_E and y_I appear in the input-output equations, and hence $\mu(a)$ must be at least twice differentiable. We then find that the right hand side of (9) consists entirely of observable quantities and identifiable parameter combinations; hence $\mu(a)$ is identifiable. Note that while we use the term "identifiable" here, what we actually mean is that $\mu(a)$ can be determined from observable quantities and identifiable parameter combinations. Further, we obtain from the boundary equations that ck_I is identifiable.

Identifiability results for all cases of the ODE and age-structured PDE model systems are summarized in Table 1. We note that in the cases where the same death rate is used for all compartments, the death rate function is always identifiable. In cases where the compartments have different death rates, the rates for the susceptible and infected compartments are identifiable. These results hold under the assumption that a fraction of the total prevalence and a fraction of all exposures can be observed at all times and ages. It has been shown previously that other output functions may lead to non-identifiability of death rates even in simple population models (Perasso and Razafison 2016).

5.5 Practical identifiability

The above structural analysis determines identifiable parameters and combinations under the assumption that data are noiseless and are available for arbitrarily many time and age points. In reality, epidemiological data are often noisy and are collected only at discrete time points and for discrete age groups. Thus, it is also necessary to perform practical identifiability analysis to determine which parameters and parameter combinations are still identifiable given more realistic data. The practical identifiability of model parameters given realistic data may differ from their structural identifiability.

As a proof of concept, we explore the practical identifiability of the constant-parameter model from Sect. 5.2 using profile likelihoods. We assume that data are collected monthly for 5-year age groups from age 0 to 100 over a span of 20 years. We synthetically generate data by choosing a baseline parameter set, simulating the model, and adding noise that is normally distributed with a standard deviation of 5%. Using these synthetic data, we perform parameter estimation using the Matlab function

Table 1 Summary of identifiable parameters and combinations for the ODE and age-structured PDE systems where the output functions are $y_E = k_E E$ and $y_I = k_I I$

Case	Identifiable parameters and combinations
Density-dependent ODE (Sect. 4.1)	$\left\{ \mu_S, \mu_I, \dfrac{\beta}{k_I}, \beta c \epsilon, \dfrac{\delta k_I}{k_E}, \dfrac{\delta(\epsilon - 1)}{\epsilon}, \mu_E + \delta \right\}$
Density-dependent PDE with constant parameters (Sect. 4.2)	$\left\{ \mu_S, \mu_I, \dfrac{\beta}{k_I}, \beta c \epsilon, \dfrac{\delta k_I}{k_E}, \dfrac{\delta(\epsilon - 1)}{\epsilon}, \mu_E + \delta \right\}$
Frequency-dependent ODE (Sect. 5.1)	$\left\{ \beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, c k_I, \dfrac{k_E}{k_I} \right\}$
Frequency-dependent PDE with constant parameters (Sect. 5.2)	$\left\{ \beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, c k_I, \dfrac{k_E}{k_I} \right\}$
Frequency-dependent ODE with immigration (Sect. 5.3)	$\left\{ \beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, (c + m) k_I, \dfrac{k_E}{k_I} \right\}$
Frequency-dependent PDE with immigration (Sect. 5.3)	$\left\{ \beta, \delta, \epsilon, \mu_S, \mu_E, \mu_I, c k_I, m k_I, \dfrac{k_E}{k_I} \right\}$
Frequency-dependent PDE with piecewise-constant death rates (Sect. 5.4.1)	$\left\{ \beta, \delta, \epsilon, c k_I, \dfrac{k_E}{k_I} \right\} \cup \left(\bigcup_{i=0}^{n-1} \{\mu_S^i, \mu_I^i, \mu_E^i\} \right)$
Frequency-dependent PDE with exponential death rate (Sect. 5.4.2)	$\left\{ \beta, \delta, \epsilon, \kappa, \mu_0, c k_I, \dfrac{k_E}{k_I} \right\}$
Frequency-dependent PDE with polynomial death rate (Sect. 5.4.3)	$\left\{ \beta, \delta, \epsilon, c k_I, \dfrac{k_E}{k_I} \right\} \cup \left(\bigcup_{i=0}^{n} \{\kappa_i\} \right)$
Frequency-dependent PDE with arbitrary death rate function (Sect. 5.4.4)	$\left\{ \beta, \delta, \epsilon, c k_I, \dfrac{k_E}{k_I}, \mu(a) \right\}$

fminsearch, and evaluate practical identifiability of the model parameters as detailed in Sect. 2.4.

We compute the Fisher Information Matrix (FIM) using the outputs y_E and y_I at the above-specified ages and time points. We find that for a wide variety of parameter sets within reasonable ranges, the rank of the FIM is eight, implying that there should be eight identifiable parameters and parameter combinations. This is consistent with our structural identifiability analysis performed in Sect. 5.2, which revealed a set of eight identifiable parameters and combinations. From the structural analysis, we see that all parameters except c, k_I, and k_E are structurally identifiable. However, from profile likelihoods we observe that the practical identifiability of these parameters does vary across the parameter space. Here, we demonstrate two baseline parameter sets that lead to qualitatively different identifiability profiles (Fig. 3). Practical identifiability varies across the parameter space, from clear minimum consistent with the structural identifiability results, to practical unidentifiability as the likelihood surface flattens
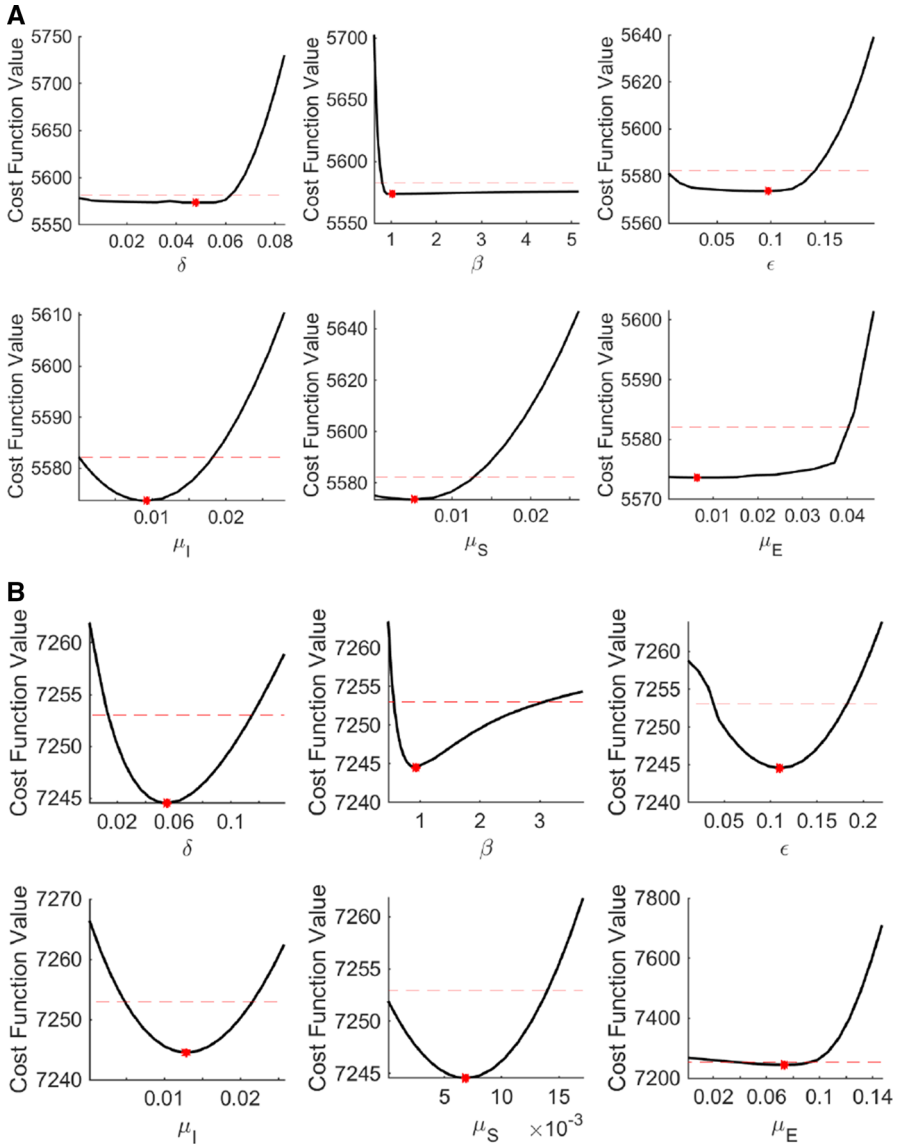
Fig. 3 Practical identifiability of the constant-parameter model from Sect. 5.2. Cost functions (black lines) based on likelihoods are shown for two sets of parameter values, $P_1 = \{c = 100, \beta = 1, \epsilon = 0.1, \delta = 0.05, \mu_S = 0.005, \mu_E = 0.005, \mu_I = 0.01, k_E = 0.5, k_I = 0.5\}$ (panel A) and $P_2 = \{c = 100, \beta = 1, \epsilon = 0.1, \delta = 0.05, \mu_S = 0.005, \mu_E = 0.075, \mu_I = 0.01, k_E = 0.75, k_I = 1\}$ (panel B). Practical identifiability varies across the parameter space, from clear minimum consistent with the structural identifiability results, to practical unidentifiability. Note that all model parameters are naturally bounded below by zero, and thus a lower bound on the confidence interval is only meaningful if it is greater than zero. The x axes in the plots above extend to zero for all parameters that do not have such a lower bound. Red dashed lines denote 95% confidence thresholds, and red dots denote local minima (corresponding to maximum likelihoods). See Sect. 2.4 for more details of the analysis

out around the best-fit value. For both parameter sets, c, k_E, and k_I are not practically identifiable.

Using the parameter set $P_1 = \{c = 100, \beta = 1, \epsilon = 0.1, \delta = 0.05, \mu_S = 0.005, \mu_E = 0.005, \mu_I = 0.01, k_E = 0.5, k_I = 0.5\}$, we find that certain structurally identifiable parameters are only practically identifiable on one side, that is, we can establish a reasonable lower or upper bound but not both. While the profile likelihoods do have a unique minimum, the profiles are flat and confidence intervals for these parameters are thus wide. In particular, the transmission rate β cannot be reasonably bounded from above, and ϵ, δ, μ_S, and μ_E cannot be distinguished from zero.

Using a slightly different parameter set $P_2 = \{c = 100, \beta = 1, \epsilon = 0.1, \delta = 0.05, \mu_S = 0.005, \mu_E = 0.075, \mu_I = 0.01, k_E = 0.75, k_I = 1\}$, in which reporting rates and death among the exposed population are increased, we find that all structurally identifiable parameters are also practically identifiable. While this parameter set is not necessarily biologically realistic (since $\mu_E > \mu_I$, for example), this demonstrates that there are regions in the parameter space where these parameters can be well estimated.

6 Summary and discussion

Creating mathematical and computational models can greatly impact the biological community through predictions that can be used and tested. Parameterization of models is a key factor in connecting them to the biology, and thus identifying parameters accurately is an important step of model development. Parameters are typically estimated from data; however, data may be sparse and noisy, and even perfect data does not guarantee unique estimability of parameters. Understanding if model parameters are identifiable lends credibility to the model system. If parameters are not identifiable, then the applicability of those models becomes more theoretical. While this may add value to understanding, it limits the potential for predictions to be translated in a meaningful way toward impacting the biological system. To this end, we focus on parameter identifiability in both ODE and PDE models. While there is a solid literature on structural identifiability analysis for ODE models, there is not a unified framework for studying identifiability in PDEs. Here, we have presented a methodology for performing structural identifiability analysis on age-structured epidemic models using a differential algebra framework. We have applied this methodology to explore the structural identifiability of a simple age-structured Susceptible-Exposed-Infected (SEI) epidemic model under various assumptions.

We found that when the age-structured model uses constant parameters and does not include immigration, then the identifiability results are identical to those for the corresponding ODE model. In the presence of immigration, however, differences in identifiability arise: the rates for birth and immigration appear only as a sum in the identifiable combinations of the ODE model, while they appear separately in those for the age-structured PDE model formulation. Further, we found that age-dependent death parameters in the PDE model are able to be uniquely recovered in at least the following three cases: piecewise-constant death rates, a single exponential death rate for all compartments, and a single polynomial death rate for all compartments.

These findings imply that identifiability results for corresponding ODE and PDE models are likely equivalent when parameters are constant, but that additional information can be gained from PDE models when parameters vary with the non-temporal variable(s). Rigorous derivation of necessary and sufficient conditions for different models to have identical identifiability results remains to be done, however based on our results herein, we conjecture that identifiability for the ODE and PDE forms of an age-structured model should be equivalent when there are not age-dependent parameters present in the model. Assuming data is available for all time and age points, we also expect age-dependent parameters to be identifiable in the PDE model if the corresponding constant parameters are identifiable in the ODE model. Additionally, future work to investigate how discretization of the PDE (e.g. converting age dynamics into linear 'chains' of compartments for age groups (Hurtado and Kirosingh 2019; Hurtado and Richards 2020)) impacts identifiability of the model would be a useful direction given the common use of compartmental representations of aging in biological models. Since the PDE system can be viewed as the limit of age-structured compartmental ODE models as the number of age compartments approaches infinity, we expect the identifiability results for the PDE system to also be a limiting case of those for compartmental ODE systems.

More generally, this work underscores some of the directions that will need to be explored to extend the differential algebra approach to the PDE case. The framework presented here could be extended to other PDE models, such as reaction-diffusion and advection-diffusion equations, in a variety of application areas. In particular, the assumption used here that age-interactions are only local (i.e. individuals interact with only people their own age) may be reasonable for reaction-diffusion models. But for more general models, this method will need to be extended to handle integrals across some variables—either to account for measurement processes that aren't age-structured or spatially resolved, or to account for non-local interactions. However, since this methodology relies on the ability to express the input-output equations as differential polynomials, there is a limit to the amount of complexity that it can handle. Extension of concepts such as Gröbner bases and characteristic sets (which enable automated generation of the input-output equations for ODEs) to the PDE case would be an important step in this direction. However, highly nonlinear models and/or output functions may require the use of other tools such as numerical methods. One such method is the FIM, which we have demonstrated in the context of an age-structured model with preferential mixing informed by real data. Additional tools for examining identifiability of ODE models, e.g. based on Lie derivatives and observability methods (Villaverde et al. 2018; Yates et al. 2009; Raue et al. 2014) may also be able to be translated to the PDE case. In applying this methodology to other classes of PDE models, care will need to be taken regarding required existence and uniqueness properties. Rigorous derivation of the necessary and sufficient conditions for this methodology to be applied remains a direction for future research. Similarly, expanding the types of PDE models considered will also require further consideration of how boundary conditions (particularly age or spatially varying boundary conditions) may affect model identifiability—either in the case where the boundary conditions are known and may improve model identifiability or in the case where they involve unknown parameters to be estimated. Finally, the approaches demonstrated here using PDEs may also open

up the possibility of examining identifiability of stochastic systems that can be characterized using PDEs (e.g. via Kolmogorov equations). We have previously begun to explore this possibility for a class of continuous time Markov models used in cancer modeling (in which case the Kolmogorov equations could be written as ODEs) (Brouwer et al. 2017).

A significant limitation of structural identifiability analysis is that it assumes the availability of perfect data. In practice, data may contain noise and bias, and may be sampled only at discrete time points. Thus, practical identifiability analysis must also be performed to gain a complete understanding of the parameter identifiability of a model given the available data. As demonstrated in our examples, practical identifiability of a model may vary not only with the quality of the data but also with the corresponding values of model parameters. Practical identifiability thus needs to be handled on a case-by-case basis for models and their corresponding parameter sets. Taken together, practical and structural identifiability provide rigor to the modeling process, adding to the applicability of model predictions.

A further limitation of this methodology for structural identifiability analysis is that it relies on the ability to eliminate system variables to obtain input-output equations. The required algebra to do so may be intractable for sufficiently complex models and/or outputs, and may necessitate extending concepts such as characteristic sets and differential Gröbner bases (Boulier 2006) to the PDE identifiability setting. Alternatively one might need to rely on numerical approaches to identifiability. For example, an age-structured model with non-local mixing would require integral terms in the state equations, making this approach substantially more difficult. Similarly, other concepts from the ODE case, such as the idea of "persistently exciting" inputs (Ljung and Glad 1994), will need to be translated to the PDE case as well. Structural identifiability analysis of such problems is a direction for future work.

References

Ajelli M, Gonçalves B, Balcan D, Colizza V, Hu H, Ramasco JJ, Merler S, Vespignani A (2010) Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. BMC Infect Dis 10(1):190. https://doi.org/10.1186/1471-2334-10-190

Audoly S, Bellu G, D'Angio L, Saccomani MP, Cobelli C (2001) Global identifiability of nonlinear models of biological systems. IEEE Trans Biomed Eng 48(1):55–65

Bearup DJ, Evans ND, Chappell MJ (2013) The input-output relationship approach to structural identifiability analysis. Comput Methods Programs Biomed 109(2):171–181

Bertuzzo E, Casagrandi R, Gatto M, Rodriguez-Iturbe I, Rinaldo A (2009) On spatially explicit models of cholera epidemics. J R Soc Interface 7(43):321–333. https://doi.org/10.1098/rsif.2009.0204

Bian L (2004) A conceptual framework for an individual-based spatially explicit epidemiological model. Environ Plann B Plann Des 31(3):381–395. https://doi.org/10.1068/b2833

Boulier F (2006) Differential elimination and biological modelling. In: Workshop D2. 2 of the Special Semester on Gröbner Bases and Related Methods, de Gruyter, vol 2, pp 111–139

Brouwer AF, Meza R, Eisenberg MC (2017) A systematic approach to determining the identifiability of multistage carcinogenesis models. Risk Anal 37(7):1375–1387

Cantwell MF, Shehab ZM, Costello AM, Sands L, Green WF, Ewing EP Jr, Valway SE, Onorato IM (1994) Congenital tuberculosis. New Engl J Med 330(15):1051–1054

Castillo-Chavez C, Feng Z (1998) Global stability of an age-structure model for TB and its applications to optimal vaccination strategies. Math Biosci 151(2):135–154. https://doi.org/10.1016/S0025-5564(98)10016-0

Castillo-Chavez C, Hethcote HW, Andreasen V, Levin SA, Liu WM (1991) Epidemiological models with age structure, proportionate mixing, and cross-immunity. J Math Biol 27(3):233–258. https://doi.org/10.1007/BF00275810

Chappell MJ, Gunn RN (1998) A procedure for generating locally identifiable reparameterisations of unidentifiable non-linear systems by the similarity transformation approach. Math Biosci 148(1):21–41. https://doi.org/10.1016/S0025-5564(97)10004-9

Chappell MJ, Godfrey KR, Vajda S (1990) Global identifiability of the parameters of nonlinear systems with specified inputs: A comparison of methods. Math Biosci 102(1):41–73

Chis OT, Banga JR, Balsa-Canto E (2011) Structural identifiability of systems biology models: a critical comparison of methods. PloS ONE 6(11):e27755

Chow L, Fan M, Feng Z (2011) Dynamics of a multigroup epidemiological model with group-targeted vaccination strategies. J Theor Biol 291:56–64. https://doi.org/10.1016/j.jtbi.2011.09.020

Cobelli C, DiStefano JJ (1980) Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. Am J Physiol Regulatory Integr Comp Physiol 239(1):R7–R24

Cole D, Morgan B, Titterington D (2010) Determining the parametric structure of models. Math Biosci 228(1):16–30. https://doi.org/10.1016/j.mbs.2010.08.004

Eisenberg M (2019) Input-output equivalence and identifiability: some simple generalizations of the differential algebra approach. arXiv

Eisenberg MC, Hayashi MA (2014) Determining identifiable parameter combinations using subset profiling. Math Biosci 256:116–126

Eisenberg MC, Robertson SL, Tien JH (2013) Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease. J Theor Biol 324:84–102. https://doi.org/10.1016/j.jtbi.2012.12.021

Evans N, Chappell M, Chapman M, Godfrey K (2004) Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems. Automatica 40:1947–1953. https://doi.org/10.1016/j.automatica.2004.06.002

Evans ND, White LJ, Chapman MJ, Godfrey KR, Chappell MJ (2005) The structural identifiability of the susceptible infected recovered model with seasonal forcing. Math Biosci 194:175–197. https://doi.org/10.1016/j.mbs.2004.10.011

Feng Z, Hill AN, Smith PJ, Glasser JW (2015) An elaboration of theory about preventing outbreaks in homogeneous populations to include heterogeneity or preferential mixing. J Theor Biol 386:177–187. https://doi.org/10.1016/j.jtbi.2015.09.006

Ferguson NM, Nokes DJ, Anderson RM (1996) Dynamical complexity in age-structured models of the transmission of the measles virus: Epidemiological implications at high levels of vaccine uptake. Math Biosci 138(2):101–130. https://doi.org/10.1016/S0025-5564(96)00127-7

Gardner ID (1980) The effect of aging on susceptibility to infection. Rev Infect Dis 2(5):801–810. https://doi.org/10.1016/j.mbs.2011.10.001

Glasser J, Feng Z, Moylan A, Valle SD, Castillo-Chavez C (2012) Mixing in age-structured population models of infectious diseases. Math Biosci 235:1–7. https://doi.org/10.1016/j.mbs.2011.10.001

Glasser J, Feng Z, Omer SB, Smith PJ, Rodewald LE (2016) The effect of heterogeneity in uptake of the measles, mumps, and rubella vaccine on the potential for outbreaks of measles: a modelling study. Lancet Infect Dis 16(5):599–605. https://doi.org/10.1016/S1473-3099(16)00004-9

Guzzetta G, Ajelli M, Yang Z, Merler S, Furlanello C, Kirschner D (2011) Modeling socio-demography to capture tuberculosis transmission dynamics in a low burden setting. J Theor Biol 289(1):197–205

Hao L, Glasser JW, Su Q, Ma C, Feng Z, Yin Z, Goodson JL, Wen N, Fan C, Yang H, Rodewald LE, Feng Z, Wang H (2019) Evaluating vaccination policies to accelerate measles elimination in China: a meta-population modelling study. International Journal of Epidemiology dyz058, https://doi.org/10.1093/ije/dyz058

Harris RC, Sumner T, Knight GM, Evans T, Cardenas V, Chen C, White RG (2019) Age-targeted tuberculosis vaccination in China and implications for vaccine development: a modelling study. Lancet Global Health 7:E209–E218. https://doi.org/10.1016/S2214-109X(18)30452-2

Hong H, Ovchinnikov A, Pogudin G, Yap C (2020) Global identifiability of differential models. Commun Pure Appl Math 73(9):1831–1879

Hurtado PJ, Kirosingh AS (2019) Generalizations of the 'linear chain trick': incorporating more flexible dwell time distributions into mean field ode models. J Math Biol 79(5):1831–1883

Hurtado PJ, Richards C (2020) Time is of the essence: Incorporating phase-type distributed delays and dwell times into ode models. arXiv:2008.01318

Inaba H, Sekine H (2004) A mathematical model for Chagas disease with infection-age-dependent infectivity. Math Biosci 1:39–69. https://doi.org/10.1016/j.mbs.2004.02.004

Jacquez JA, Greif P (1985) Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. Math Biosci 77(1–2):201–227

Jacquez JA, Perry T (1990) Parameter estimation: local identifiability of parameters. Am J Physiol 258(4 Pt 1):E727-36

Kao YH, Eisenberg MC (2018) Practical unidentifiability of a simple vector-borne disease model: implications for parameter estimation and intervention assessment. Epidemics 25:89–100. https://doi.org/10.1016/j.epidem.2018.05.010

Keeling MJ, White PJ (2011) Targeting vaccination against novel infections: risk, age and spatial structure for pandemic influenza in Great Britain. J R Soc Interface 8(58):661–670. https://doi.org/10.1098/rsif.2010.0474

Ljung L, Glad T (1994) On global identifiability for arbitrary model parameterization. Automatica 30(2):265–276

LoBue PA, Mermin JH (2017) Latent tuberculosis infection: the final frontier of tuberculosis elimination in the USA. Lancet Infect Dis 17(10):e327–e333. https://doi.org/10.1016/S1473-3099(17)30248-7

Meshkat N, Sullivant S (2014) Identifiable reparametrizations of linear compartment models. J Symb Comput 63:46–67, https://doi.org/10.1016/j.jsc.2013.11.002, https://www.sciencedirect.com/science/article/pii/S074771711300165X

Meshkat N, Eisenberg M, DiStefano JJ III (2009) An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner bases. Math Biosci 222:61–72. https://doi.org/10.1016/j.mbs.2009.08.010

Meshkat N, Anderson C, DiStefano JJ III (2011) Finding identifiable parameter combinations in nonlinear ODE models and the rational reparameterization of their input-output equations. Math Biosci 233:19–31. https://doi.org/10.1016/j.mbs.2011.06.001

Meshkat N, Kuo C, Distefano J (2014) On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: A novel web implementation. PloS ONE 9:e110261. https://doi.org/10.1371/journal.pone.0110261

Miao H, Xia X, Perelson A, Wu H (2011) On identifiability of nonlinear ode models and applications in viral dynamics. SIAM Rev 53(1):3–39. https://doi.org/10.1137/090757009

Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 5(3):e74

Ovchinnikov A, Pogudin G, Thompson P (2021) Parameter identifiability and input–output equations. Applicable Algebra in Engineering, Communication and Computing pp 1–18

Ozcaglar C, Shabbeer A, Vandenberg SL, Yener B, Bennett KP (2012) Epidemiological models of mycobacterium tuberculosis complex infections. Math Biosci 236(2):77–96. https://doi.org/10.1016/j.mbs.2012.02.003

Perasso A, Razafison U (2016) Identifiability problem for recovering the mortality rate in an age-structured population dynamics model. Inverse Probl Sci Eng 24(4):711–728. https://doi.org/10.1080/17415977.2015.1061522

Perasso A, Laroche B, Chitour Y, Touzeau S (2011) Identifiability analysis of an epidemiological model in a structured population. J Math Anal Appl 374:154–165. https://doi.org/10.1016/j.jmaa.2010.08.072

Pia Saccomani M, Audoly S, D'Angio L (2003) Parameter identifiability of nonlinear systems: the role of initial conditions. Automatica 39(4):619–632

Pohjanpalo H (1978) System identifiability based on the power series expansion of the solution. Math Biosci 41(1–2):21–33

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmuller U, Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25(15):1923–9

Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J (2014) Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 30(10):1440–1448

Renardy M, Kirschner D (2019) Evaluating vaccination strategies for tuberculosis in endemic and non-endemic settings. J Theor Biol 469:1–11

Renardy M, Kirschner D (2020) A framework for network-based epidemiological modeling of tuberculosis dynamics using synthetic datasets. Bull Math Biol 82(6):78

Ritt JF (1950) Differential algebra, vol 33. American Mathematical Society, Providence

Shim E, Feng Z, Martcheva M, Castillo-Chavez C (2006) An age-structured epidemic model of rotavirus with vaccination. J Math Biol 53(4):719–746. https://doi.org/10.1007/s00285-006-0023-0

Thieme HR, Castillo-Chavez C (1993) How may infection-age-dependent infectivity affect the dynamics of HIV/AIDS? SIAM J Appl Math 53(5):1447–1479. https://doi.org/10.1137/0153068

Tuncer N, Le TT (2018) Structural and practical identifiability analysis of outbreak models. Math Biosci 299:1–18. https://doi.org/10.1016/j.mbs.2018.02.004

Tuncer N, Gulbudak H, Cannataro VL, Martcheva M (2016) Structural and practical identifiability issues of immuno-epidemiological vector-host models with application to Rift Valley fever. Bull Math Biol 78:1796–1827. https://doi.org/10.1007/s11538-016-0200-2

Valle SYD, Hyman JM, Hethcote HW, Eubank SG (2007) Mixing patterns between age groups in social networks. Soc Netw 29(4):539–554

Venzon DJ, Moolgavkar SH (1988) A method for computing profile-likelihood-based confidence intervals. J R Stat Soc Ser C (Appl Stat) 37(1):87–94. https://doi.org/10.2307/2347496

Villaverde AF, Evans ND, Chappell MJ, Banga JR (2018) Input-dependent structural identifiability of nonlinear systems. IEEE Control Syst Lett 3(2):272–277

Vynnycky E, Fine PEM (1997) The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol Infect 119(2):183–201

Wu J (2008) Spatial structure: Partial differential equations models. In: Brauer F, van den Driessche P, Wu J (eds) Mathematical Epidemiology. Lecture Notes in Mathematics, vol 1945. Springer, Berlin, pp 191–203

Yates JW, Evans ND, Chappell MJ (2009) Structural identifiability analysis via symmetries of differential equations. Automatica 45(11):2585–2591

Zhu S, Verdiére N, Denis-Vidal L, Kateb D (2018) Identifiability analysis and parameter estimation of a chikungunya model in a spatially continuous domain. Ecol Complex 34:80–88. https://doi.org/10.1016/j.ecocom.2017.12.004

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