On characteristics of an ordinary differential equation and a related inverse problem in epidemiology

Ralph Brinks

Institute of Biometry and Epidemiology

German Diabetes Center

Duesseldorf, Germany

In this work we examine the properties of a recently described ordinary differential equation that relates the age-specific prevalence of a chronic disease with the incidence and mortalities of the diseased and healthy persons. The equation has been used to estimate the incidence from prevalence data, which is an inverse problem. The ill-posedness of this problem is proven, too.

1 Introduction

Recently a novel ordinary differential equation (ODE) has been described that relates the age-specific prevalence of an irreversible disease with its incidence rate and the mortality rates of the diseased and the non-diseased persons (Brinks, 2011). This article is about properties of the ODE and its solutions.

Given the mortality rates and the age-specific prevalence, the ODE may be used to derive the incidence rate[1], which can be interpreted as an inverse problem. Inverse

---

[1]In this article the expressions *rate* and *density* are synonymously used.

problems are often examined with respect to ill- or well-posedness. A well-posed of an inverse problem in the sense of Jacques Hadamard means that a solution exists, that the solution is unique and stable (Hadamard, 1923). In this article the ill-posedness of the inverse problem is proven.

The article is organized as follows. Section 2 briefly reviews the derivation of the ODE. Then, some properties of the ODE and its solution are examined. In Section 3 the inverse problem is introduced and the ill-posedness is proven. Finally, in Section 4 the results and its consequences are discussed.

2 The ODE: Derivation and Properties

A popular framework for studying relations between prevalence and incidence of a disease is the simple model consisting of three states as depicted in Figure 1: *Normal, Disease* and *Death*, (Keiding, 1991; Murray and Lopez, 1994). This model goes back at least until the 1950s (Fix and Neyman, 1951). In general, the transition densities from one state to another depend on calendar time t and age a. The transition density from *Disease* to *Death* may also depend on the duration d of the disease.

People in the population under consideration can get a disease at incidence density i, they can die either after having got the disease at age-specific mortality rate m_1 or without having the disease at mortality rate m_0. The numbers of the individuals in the *Normal* and in the *Disease* state are denoted by S (susceptibles) and C (cases). Generally, these numbers as well as the mortality and incidence densities depend on the calendar time t. Sometimes, these quantities are assumed to be independent of time t, which is referred to as the time-homogeneous case (Keiding, 1991).

Assuming time-homogeneity and disease related mortality m_1 to be independent of duration d, Murray and Lopez (1994) described transitions along the paths in Figure 1 as a set of ODEs. Henceforth, beside time-homogeneity we additionally assume that the population is closed (no migration) and has a constant birth rate. Furthermore, the age-specific functions i, m_0, m_1 are non-negative and continuous in $[0, \omega]$ for some $\omega > 0$. Henceforth, ω is considered as the minimal age when all members (diseased and non-diseased) of the population are deceased[2]. Then, Equation (1) describes the change rates of the numbers S and C of normal and diseased individuals, respectively.

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}a} &= -\left(i(a) + m_0(a)\right) \cdot S \\
\frac{\mathrm{d}C}{\mathrm{d}a} &= i(a) \cdot S - m_1(a) \cdot C
\end{aligned} \tag{1}$$

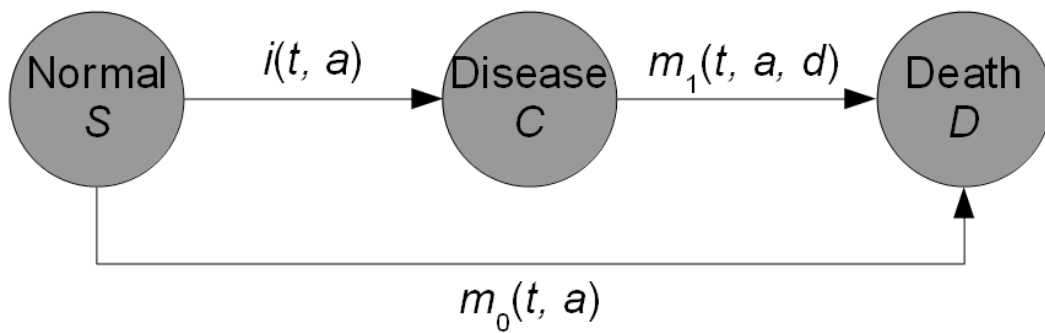[2]For example by choosing $\omega = \inf\{a > 0|\ C(a) + S(a) < 0.5\}$.

Figure 1: Three states model of normal, diseased and dead subjects. Transition densities may depend on calender time t, age a, and duration d of the disease.

The resulting set of ODEs is linear and of first order. Due to the simple structure of the ODEs for given age-specific incidence and mortality rates i, m_0 and m_1, the analytical solution of the corresponding initial value problem with initial conditions $S(0) = S_0 \geq 0$, $C(0) = C_0 \geq 0$, $S_0 + C_0 > 0$ can be calculated easily:

$$S(a) = S_0 \cdot \exp\left(-\int_0^a i(\tau) + m_0(\tau)\mathrm{d}\tau\right)$$

$$C(a) = \exp\left(-\int_0^a m_1(\tau)\mathrm{d}\tau\right) \cdot \left\{C_0 + \int_0^a i(\tau)S(\tau)\exp\left(\int_0^\tau m_1(t)\mathrm{d}t\right)\mathrm{d}\tau\right\}. \tag{2}$$

Obviously, from $S_0 + C_0 > 0$ it follows that $S(a) \geq 0$, $C(a) \geq 0$ and $S(a) + C(a) > 0$ for all $a \in [0, \omega]$.

Usually, for a population under consideration the overall (general) mortality density m is observed or reported in life-tables. The mortality density m is a convex combination of the mortality density m_0 of the normals and the mortality density m_1 of the diseased:

$$m(a) = p(a) \cdot m_1(a) + \{1 - p(a)\} \cdot m_0(a)$$
$$= m_0(a) \cdot \{p(a) \cdot (R(a) - 1) + 1\}, \tag{3}$$

where $R(a)$ is the relative risk, $R = \frac{m_1}{m_0}$. In this expression p is the prevalence of the disease, which for a specific age a and $S(a) + C(a) > 0$ can be written as

$$p(a) = \frac{C(a)}{S(a) + C(a)}. \tag{4}$$

Equation (3) allows the application of the ODE-system (1) in the case when m_0 and m_1 are unknown. In epidemiology this typically is the case. Then the ODE-system (1) becomes non-linear and does not have an analytical solution anymore.

Interestingly, the two-dimensional system can be reduced to a one-dimensional ODE, which is stated in the following Theorem. This can easily by derived from the quotient rule and Equation (1).

Theorem 2.1. *Let mortality densities* $m, m_0 \in C^0\left([0, \omega]\right)$ *and* $S, C \in C^1\left([0, \omega]\right)$ *with* $S(a) + C(a) > 0$ *for all* $a \in [0, \omega]$, *then* $p = \frac{C}{S+C}$ *is differentiable in* $[0, \omega]$ *and it holds*

$$\frac{\mathrm{d}p}{\mathrm{d}a} = (1 - p) \cdot (i - (m - m_0)). \tag{5}$$

4

Table 1: Types of the ODE (5) depending on the given mortality.

Given mortality	Right hand side of Eq. (5)	Type of the ODE
m, m_0	$(1-p) \cdot [i - (m - m_0)]$	Linear
m, m_1	$(1-p) \cdot \left[i - p \cdot \frac{m_1 - m}{1-p}\right]$	Linear
m_0, m_1	$(1-p) \cdot [i - p \cdot (m_1 - m_0)]$	Riccati
m_0, R	$(1-p) \cdot [i - p \cdot m_0 \cdot (R-1)]$	Riccati
m_1, R	$(1-p) \cdot \left[i - p \cdot m_1 \cdot \frac{R-1}{R}\right]$	Riccati
m, R	$(1-p) \cdot \left[i - m \cdot \frac{p \cdot (R-1)}{p \cdot (R-1)+1}\right]$	Abelian

Depending on the type of information about the mortality densities, the ODE (5) changes its type (Table 1), which is important when solving the ODE. In case the ODE is linear, an easy analytical solution exists. If the ODE is of Riccati or Abelian type (Kamke, 1983), a general analytical solution does not exist. An extensive monograph about Riccati equations is (Reid, 1972).

The fractions $\frac{R-1}{R}$ and $\frac{p \cdot (R-1)}{p \cdot (R-1)+1}$ in the last two rows, are very well known in epidemiology. These are the *exposition attributable risk* (EAR) and the *population attributable risk* (PAR), respectively, (Woodward, 2005).

Next it is examined, if the solutions of the one-dimensional ODE (5) are epidemiologically meaningful, i.e. $p(a) \in [0,1]$ for all $a \in [0, \omega]$. For the system (1) this is clear: the age-specific prevalence $p(a) = \frac{C(a)}{S(a)+C(a)}$ given by the solutions (2) are meaningful for all $a \in [0, \omega]$. However, it is not obvious that solutions p of (5) are between 0 and 1. For the special case that $m_0 = m_1$ – this case is called *non-differential mortality* – it can be proven directly. Then the solution of (5) with initial condition $p(0) = p_0 \in [0,1]$ is

$$p(a) = 1 - (1 - p_0) \cdot \exp\left(-\int_0^a i(\tau) \mathrm{d}\tau\right),$$

and the epidemiological meaningfulness follows immediately. In case of differential mortality ($m_0 \neq m_1$) epidemiological meaningfulness cannot not be proven directly, because it has to include all the different cases of the right hand side of (5) in Table 1. Instead of a direct proof we use the correspondence between Equations (1) and (5). Let $N(a) := C(a) + S(a)$ denote the number of persons alive at age a, then it holds $N(a) > 0$ for all $a \in [0, \omega]$. We augment Equation (5) by another ODE in N with m defined in Equation (3):

5

$$\frac{\mathrm{d}p}{\mathrm{d}a} = (1 - p) \cdot (i - (m - m_0))$$
$$\frac{\mathrm{d}N}{\mathrm{d}a} = -m \cdot N. \tag{6}$$

Then we have the following correspondence between the ODE-systems (1) and (6):

Theorem 2.2. *(A) If $S(a), C(a)$ are solutions of (1), then $p(a) := \frac{C(a)}{S(a)+C(a)}$ and $N(a) := C(a) + S(a)$ are solutions of (6).*

(B) If $p(a), N(a)$ are solutions of (6), then $S(a) := \{1 - p(a)\} \cdot N(a)$ and $C(a) := p(a) \cdot N(a)$ are solutions of (1).

Proof. This is an easy exercise in calculus. □

From Theorem 2.2 (B) it follows that $p(a) \in [0, 1]$ for all $a \in [0, \omega]$: If $p(a)$ is a solution of (6), then it has a representation $p(a) = \frac{C(a)}{N(a)}$. Since $0 \le C(a) \le N(a)$, the solution p is epidemiologically meaningful.

Remark 2.1. *From Theorem 2.2 (A) it is obvious that ODE-system (1) implies $\frac{\mathrm{d}N}{\mathrm{d}a} = -m \cdot N$. This is the defining equation of a* stationary population*, which is a population with a special type of age distribution (Preston and Coale, 1982). Hence, (1) is valid only for stationary populations. Since most populations are non-stationary, this is a heavy limitation. However, it can be shown that (5) holds true in general populations as long as certain restrictions on the migration rate are fulfilled. Details are not subject of this work and will be elaborated in a subsequent paper.*

3 The Inverse Problem

A key application for the ODE (5) is the derivation of the age-specific incidence rate $i(a)$ from $p(a)$ if the mortalities (or any equivalent information in the first column of Table 1) are known. In epidemiology, typically incidences rates are surveyed in follow-up studies, which may be very lengthy and costly. If the model assumptions for ODE (5) hold true, the equation can be solved for $i(a)$. Beside mortality information, the age course of the prevalence has to be known, which can be obtained from relatively cheap cross-sectional studies. An example is shown in (Brinks, 2011).

In such an application with given mortalities, we conclude from an effect (the prevalence) to the underlying cause (the incidence), which can be interpreted as an *inverse problem* (Tarantola, 2005). The inverse problem is opposed to the *direct problem* of inferring from the incidence (i.e. the cause) to the prevalence (the effect) by ODE

(5). Now we show that the inverse problem is ill-posed in the sense of Hadamard (1923). For given (sufficiently smooth) mortalities and $p_0 \in [0, 1]$ define the operator $\wp : C^0([0, \omega]) \to C^1([0, \omega]), i \mapsto p$, such that $p(0) = p_0$ and p is the solution of (5). To show that the inverse problem is ill-posed we prove that $\wp^{-1} : p \mapsto \frac{dp/da}{1-p} + m - m_0$ is discontinuous. It is sufficient to show this for the special case of non-differential mortality $(m = m_0)$. Let $C^k([0, \omega])$, $k = 0, 1$, be equipped with the maximum norm $\|\cdot\|$. Choose $p \in C^1([0, \omega])$, $\epsilon > 0$ and define $p_{\epsilon,n}(a) := p(a) + \epsilon \cdot \sin(n \cdot a)$. Then, it is $\|p - p_{\epsilon,n}\| \le \epsilon$ and

$$\|\wp^{-1}(p) - \wp^{-1}(p_{\epsilon,n})\| = \left\| \frac{dp}{da} \frac{1}{1-p} - \frac{dp_{\epsilon,n}}{da} \frac{1}{1-p_{\epsilon,n}} \right\|$$

$$= \left\| \frac{\frac{dp}{da} \epsilon \sin(n\cdot) + \epsilon\, n \cos(n\cdot)(1-p)}{(1-p)(1-p-\epsilon \sin(n\cdot))} \right\|.$$

For $\epsilon_n := n^{-1/2}$ and $p(a) \ne 1$ the term $\epsilon_n\, n \cos(na)(1 - p(a))$ is unbounded as $n \to \infty$, which implies that $\wp^{-1}$ is discontinuous and the inverse problem is ill-posed.

4 Discussion

By extending the framework of (Murray and Lopez, 1994) for studying the relation between prevalence and incidence, it had been found that prevalence, incidence and mortality are linked by a simple one-dimensional ODE. In this article it has been shown that the solutions of this ODE are epidemiologically meaningful. Depending on the type of mortality information available, the ODE changes its type, which has implications about existence of general analytical solutions. In many epidemiologically relevant cases, an analytical solution does not exist, and numerical treatment has to used instead.

An important application of the ODE is the derivation of age-specific incidences from the age distribution of the prevalence. This article shows that this question can be interpreted as an inverse problem. Furthermore, the inverse problem is ill-posed. The proof of the ill-posedness shows that an additive high frequency distortion ($\epsilon \sin(n\cdot)$) of the prevalence may lead to an unbounded inaccuracy in the derived incidence. However, high frequency distortions might be unlikely in real chronic diseases. Hence, the consequences in practical epidemiology are unclear so far.

In the discussed ODE model, several assumptions have been made. The ODE is valid only if incidence and mortality rates are independent from calendar time. Due to changes in medical progress, hygiene, nutrition and lifestyle, mortality does undergo secular trends. Thus, it is appropriate to formulate Equation (5) as a partial differential

equation, which is subject of a subsequent paper. Moreover, in real diseases mortality of the diseased persons depend on the duration of the disease. An example is diabetes, where the relative mortality over the diabetes duration is U-shaped (Carstensen et al., 2008). Duration dependency obfuscates the relation between prevalence, incidence and mortality (Keiding, 1991). Results as easy as presented here are unlikely not be expected.

A last note refers to the term *chronic disease*: In this article, *chronic* means *irreversible*, i.e. there is no way back from the *Disease* to the *Normal* state. However, most of the results presented here remain true, if there is remission back to state *Normal*. Then, the fundamental ODE (5) has an additional term that depends on the remission rate (Brinks, 2011).

References

Brinks, R. (2011). A new method for deriving incidence rates from prevalence data and its application to dementia in Germany. `http://arxiv.org/abs/1112.2720v1`

Carstensen, B., Kristensen, J.K., Ottosen, P., and Borch-Johnsen, K. (2008) The Danish National Diabetes Register: trends in incidence, prevalence and mortality. *Diabetologia* **51 (12),** 2187–96.

Fix, E. and Neyman, J. (1951). *Human Biology* **23 (3),** 205–241.

Hadamard, J. (1923). *Lectures on the Cauchy Problem in Linear Partial Differential Equations* New York: Yale University Press.

Kamke, E. (1983). *Differentialgleichungen.* Stuttgart: Teubner.

Keiding, N. (1991). Age-specific incidence and prevalence: a statistical perspective. *Journal of the Royal Statistical Society A* **154,** 371–412

Murray, C. J. L and Lopez, A. D. (1994). Quantifying disability: data, methods and results *Bulletin of the WHO* **72 (3),** 481–494

Preston, S. H. and Coale, A. J. (1982). Age structure, growth, attrition, and accession: a new synthesis, *Population Index* **48 (2)** 217–59

Reid, W. T. (1972). *Riccati Differential Equations* New York: Academic Press.

Tarantola, A. (2005). *Inverse Problem Theory* Philadelphia: Society for Industrial and Applied Mathematics.

Woodward, M. (2005). *Epidemiology. Study Design and Data Analysis* Boca Raton: Chapman & Hall/CRC.