

Supplementary material to the article

# Occam's razor gets a new edge: the role of symmetries in model selection

Johannes Borgqvist<sup>†‡§¶</sup> and Sam Palmer<sup>†</sup>.

---

<sup>†</sup>Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, United Kingdom

<sup>‡</sup> Research fellow at the Wenner-Gren Foundations, Sweden

<sup>§</sup> Junior Research Fellow (JRF) at Linacre College, University of Oxford, United Kingdom

<sup>¶</sup>E-mail: [Johannes.Borgqvist@maths.ox.ac.uk](mailto:Johannes.Borgqvist@maths.ox.ac.uk)

# Contents

<b>1</b>	<b>Finding the symmetries of the three candidate models</b>	<b>2</b>
1.1	Derivation of the ODE models . . . . .	2
1.2	Mathematical theory of symmetries of a single first order ODE . . . . .	3
1.3	Derivation of the symmetries . . . . .	6
1.3.1	The IM-I . . . . .	6
1.3.2	The PLM . . . . .	10
1.3.3	The IM-II . . . . .	15
<b>2</b>	<b>Summary of the calculated symmetries</b>	<b>18</b>
<b>3</b>	<b>Processing of the data and the curve fitting</b>	<b>18</b>
<b>4</b>	<b>Analytical validation of the <math>t</math>-directional generators of the PLM and the IM-II</b>	<b>23</b>
<b>5</b>	<b>Numerical validation of the <math>t</math>-directional symmetries of the PLM and the IM-II</b>	<b>25</b>
<b>6</b>	<b>Although two models have very good fits to the same time series data it is essential to perform statistical tests</b>	<b>26</b>

## List of Figures

1	Action of the non-trivial symmetries of the PLM . . . . .	20
2	Action of the non-trivial symmetry of the IM-II . . . . .	21
3	Action of the non-trivial symmetries of the IM-I . . . . .	22
4	Processing of the data . . . . .	22
5	Numerical validation of the symmetries . . . . .	27
6	Both the PLM and the IM-II fit colon cancer data well . . . . .	28

## List of Tables

1	ODE formulations of the three models. . . . .	4
2	Summary of the symmetries . . . . .	19
3	The fit of the two candidate models to the colon cancer data . . . . .	29
4	The variance test suggests that both models are indistinguishable when fitted to two different data sets. . . . .	30

# 1 Finding the symmetries of the three candidate models

The aim is to find the symmetries of three models predicting risk profiles over time. The focus of the article is the following three models [1]:

$$R(t) = Ae^{\alpha t}, \quad (1)$$

$$R(t) = At^\gamma, \quad (2)$$

$$R(t) = \frac{A}{\exp(e^{-\alpha(t-\tau)}) - 1}. \quad (3)$$

More precisely, the comparison in the main article is between (2) and (3) as these two models fit the data the best. Despite the fact that the model in (1) is excluded from the symmetry test due to its bad fit its symmetries are also calculated in this supplementary document.

These models are referred to as the IM-I in (1), the *power law model* (PLM) in (2) and the IM-II in (3). The idea is to use the symmetry based framework for model selection presented in [2] on these models and the data presented in [1] in order to deduce which model that best predicts the risk profile. To this end, it is necessary to re-write these models as *ordinary differential equations* (ODEs) and then solve the so called *linearised symmetry condition* [3] in order to find the symmetries of the models at hand. Subsequently, the ODE models will be derived, then the theory of symmetries of first order ODEs will be described briefly and then the symmetries for each of the three models will be derived.

## 1.1 Derivation of the ODE models

As symmetry methods are well-established for differential equations, it is advantageous to formulate the candidate models in this way. The corresponding ODE formulation of each model is derived by differentiating the curves in (1) to (3) and then substituting the equation for the actual curve into the equation for the derivative. For (1), we have

$$\frac{dR}{dt} = \frac{d}{dt} (Ae^{\alpha t}) = \underbrace{\alpha Ae^{\alpha t}}_{=R} = \alpha R$$

so

$$\frac{dR}{dt} = \alpha R.$$

For (2), we have

$$\frac{dR}{dt} = \frac{d}{dt} (At^\gamma) = \gamma At^{\gamma-1} = \gamma \underbrace{(At^\gamma)t^{-1}}_{=R} = \gamma \frac{R}{t}$$

so

$$\frac{dR}{dt} = \gamma \frac{R}{t}.$$

Lastly, for (3) we have

$$\begin{aligned}
\frac{dR}{dt} &= \frac{d}{dt} \left[ A \left( \exp(e^{-\alpha(t-\tau)}) - 1 \right)^{-1} \right] \\
&= -A \left( \exp(e^{-\alpha(t-\tau)}) - 1 \right)^{-2} \exp(e^{-\alpha(t-\tau)}) \left( -\alpha e^{-\alpha(t-\tau)} \right) \\
&= \frac{\alpha}{A} \underbrace{\left( \frac{A}{\exp(e^{-\alpha(t-\tau)}) - 1} \right)^2}_{=R^2} \exp(e^{-\alpha(t-\tau)}) e^{-\alpha(t-\tau)} \\
&= \frac{\alpha}{A} R^2 e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)})
\end{aligned}$$

so

$$\frac{dR}{dt} = \frac{\alpha}{A} R^2 e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}).$$

An important classification of an ODE is whether or not it is autonomous. This term implies that there is no dependence on the variable, e.g. the time  $t$ , in the right hand side of the ODE. As can be seen by comparing the ODE formulations it is possible to conclude that the three candidate models are qualitatively different (Tab 1). This qualitative difference is reflected in the symmetries as a well-known symmetry that characterises the autonomous family of ODEs which is not a symmetry of non-autonomous ODEs is the famous *translation symmetry* which will be described later in the text.

Throughout the text, the right hand side of the ODE will be denoted by  $\omega(t, R)$  and this will be referred to as the *reaction term*. More precisely, for the three models predicting the risk profiles the ODEs will be written on the following form:

$$\frac{dR}{dt} = \omega(t, R). \quad (4)$$

Before the symmetries of the three models are derived, a brief presentation of the theory of symmetries are presented in order to properly define the involved concepts.

## 1.2 Mathematical theory of symmetries of a single first order ODE

In order for the symmetry based model selection framework proposed in [2] to work, it is necessary to find symmetries that are *unique* to each model. Before these calculations are conducted, the notion of a symmetry as a mathematical object will be presented.

A symmetry denoted by  $\Gamma$  is an operator that maps a solution curve  $\gamma = (t, R(t))$  to another solution curve  $\hat{\gamma} = (\hat{t}, \hat{R})$  [3]. Mathematically, this can be expressed as follows:  $\Gamma : \gamma \mapsto \hat{\gamma}$ . A more rigorous definition of these operators requires the introduction of some notation. If  $B = \{t \in \mathbb{R}\} \simeq \mathbb{R}$  denotes the so called *base space* which is the space of the variable and  $F = \{y(t) \in \mathbb{R}\} \simeq \mathbb{R}$  denotes the so called *fibre* being the space of the dependent variable or the function then the solutions curves  $\gamma = (t, R(t))$  reside in the so called *total space*  $E = B \times F \simeq \mathbb{R}^2$  being the product space of the two previously mentioned spaces. Now,

Table 1: *ODE formulations of the three models.* In the three columns from left to right the following is presented: the original model, the ODE formulation of the model and the characterisation of the ODE as either autonomous or non-autonomous.

<i>Model</i>	<i>Curve</i>	<i>ODE</i>	<i>Characterisation</i>
IM-I	$R(t) = Ae^{\alpha t}$	$\frac{dR}{dt} = \alpha R$	Autonomous
PLM	$R(t) = At^\gamma$	$\frac{dR}{dt} = \gamma \frac{R}{t}$	Non-autonomous
IM-II	$R(t) = \frac{A}{\exp(e^{-\alpha(t-\tau)}) - 1}$	$\frac{dR}{dt} = \frac{\alpha}{A} R^2 e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)})$	Non-autonomous

a pointwise symmetry  $\Gamma$  maps every point on a solution curve to another point on another solution curve which can be summarised as follows  $\Gamma : E \mapsto E$ . In the context of this article, we are interested in symmetries that are so called  $\mathcal{C}^\infty$  diffeomorphisms (an isomorphism of smooth manifolds) which are parametrised by a single parameter  $\epsilon$ . In particular, this means that the symmetry is continuous in the parameter  $\epsilon$ , it is also invertible and it can be represented as a Taylor series around  $\epsilon = 0$  [3]. From now on, when we say that we are looking for a symmetry it will be understood to be of this type, and we will denote these by

$$\Gamma(\epsilon) : (t, R) \mapsto (\hat{t}(\epsilon), \hat{R}(\epsilon)).$$

Due to the continuity of these symmetries, it is possible to obtain a local description of their action.

The key for finding symmetries is the local representation of them in terms of their Taylor series. For a given symmetry  $\Gamma(\epsilon) : (t, R) \mapsto (\hat{t}(\epsilon), \hat{R}(\epsilon))$ , the Taylor expansion of the transformed coordinates around  $\epsilon = 0$  is the following

$$\begin{aligned}\hat{t}(\epsilon) &= t + \xi(t, R)\epsilon + \mathcal{O}(\epsilon^2) \\ \hat{R}(\epsilon) &= R + \eta(t, R)\epsilon + \mathcal{O}(\epsilon^2)\end{aligned}$$

where

$$\xi(t, R) = \left. \frac{d\hat{t}}{d\epsilon} \right|_{\epsilon=0} \quad \text{and} \quad \eta(t, R) = \left. \frac{d\hat{R}}{d\epsilon} \right|_{\epsilon=0}.$$

These latter two terms  $(\xi, \eta)$  are referred to as the tangents of the symmetry and they define the vector field

$$X = \xi(t, R)\partial t + \eta(t, R)\partial R \quad (5)$$

which is called the *infinitesimal generator of the Lie group* as the set of symmetries form a one parameter Lie group [3]. By defining the exponential map as follows

$$e^{\epsilon X} = \sum_{j=0}^{\infty} \frac{\epsilon^j}{j!} X^j$$

the symmetry  $\Gamma : (t, R) \mapsto (\hat{t}(\epsilon), \hat{y}(\epsilon))$  is generated by the infinitesimal generator of the Lie group in the following manner

$$\Gamma(\epsilon) : (t, R) \mapsto (e^{\epsilon X} t, e^{\epsilon X} R) . \quad (6)$$

In other words, it is enough to calculate the tangents to obtain the generator  $X$  in (5) which subsequently can be used to obtain the symmetries through the exponential map according to (6). Thus, the task of finding symmetries revolves around finding the tangents  $\xi$  and  $\eta$  in the infinitesimal generator of the Lie group  $X$ .

The tangents can be found by solving the *linearised symmetry condition*. For a first order ODE [3], this condition corresponds to solving a two dimensional PDE which looks as follows:

$$\frac{\partial \eta}{\partial t} + \left( \frac{\partial \eta}{\partial R} - \frac{\partial \xi}{\partial t} \right) \omega(t, R) - \frac{\partial \xi}{\partial R} \omega(t, R)^2 = \xi \frac{\partial \omega}{\partial t} + \eta \frac{\partial \omega}{\partial R}. \quad (7)$$

Here,  $\omega$  corresponds to the reaction term of the ODE in question as in (4). So the tangents  $(\xi, \eta)$  are found by solving the linearised symmetry condition in (7), and then the symmetries are obtained using the exponential map as in (6). Furthermore, it is possible to classify the action of a symmetry in a similar way to the classification of the ODEs in terms of their qualitative behaviour (Tab 1). The reason why this classification is necessary is on account of the fact that symmetries that map solutions curves to *different* solution curves are required for the model selection procedure to work.

The so called *reduced characteristic* denoted by  $\overline{Q}$  is used to classify symmetries. For a first order ODE, this is defined as follows: [3]

$$\overline{Q}(X) = \overline{Q}(t, R) \Big|_{\xi, \eta \text{ defined by } X} = \eta(t, R) - \omega(t, R)\xi(t, R). \quad (8)$$

If  $\overline{Q}(X) = 0$  for a specific solution  $\gamma = (t, R)$ , the solution is *trivial* under the action of the symmetry. This implies that the symmetry maps points along the *same* solution curve. Moreover, if  $\overline{Q}(X) \equiv 0$  then the symmetry itself is *trivial* meaning that it does not transform any point for any solution curve (for an illustration of the geometrical interpretation of the characteristic see [4]). In order for the symmetry framework to work, we require at least one unique and *non-trivial* symmetry for each model, i.e.  $\overline{Q}(X) \neq 0$  for the particular symmetry

$X$  at hand, meaning that points on a solution curve are transformed to points on *another* solution curve for each model. Hence, after the linearised symmetry condition is solved for each of the three models, the calculated symmetries will be characterised using the reduced characteristic in (8).

### 1.3 Derivation of the symmetries

For the purpose of selecting a model, we are looking for three non-trivial symmetries  $\Gamma_1(\epsilon)$ ,  $\Gamma_2(\epsilon)$  and  $\Gamma_3(\epsilon)$  that are *unique* to each of the candidate models in (1), (2) and (3) respectively. Subsequently, the derivation of each of these three symmetries is presented in chronological order.

#### 1.3.1 The IM-I

For the sake of notation, let's denote the tangents in the infinitesimal generator of the Lie group for the IM-I by  $\xi_3(t, R)$  and  $\eta_3(t, R)$  respectively. The reaction term in (1) is given by  $\omega_3(t, R) = \alpha R$ . For the partial derivatives in the linearised symmetry condition, we have

$$\frac{\partial \omega_3}{\partial t} = 0, \quad \frac{\partial \omega_3}{\partial R} = \alpha \quad \text{and} \quad \omega_3^2 = \alpha^2 R^2.$$

Plugging in these expressions into (7) yields

$$\frac{\partial \eta_3}{\partial t} + \left( \frac{\partial \eta_3}{\partial R} - \frac{\partial \xi_3}{\partial t} \right) \alpha R - \frac{\partial \xi_3}{\partial R} \alpha^2 R^2 = \eta_3 \alpha$$

and by organising this as a polynomial in  $R$ , the above expression can be written as follows:

$$\left[ \frac{\partial \eta_3}{\partial t} - \alpha \eta_3 \right] R^0 + \left[ \alpha \left( \frac{\partial \eta_3}{\partial R} - \frac{\partial \xi_3}{\partial t} \right) \right] R + \left[ \alpha^2 \frac{\partial \xi_3}{\partial R} \right] R^2 = 0.$$

Now, since the monomials  $\{R^0, R^1, R^2\}$  in the above polynomial are linearly independent it follows that the following so called *determining equations* [3] must hold:

$$\frac{\partial \eta_3}{\partial t} - \alpha \eta_3 = 0, \tag{9}$$

$$\alpha \left( \frac{\partial \eta_3}{\partial R} - \frac{\partial \xi_3}{\partial t} \right) = 0, \tag{10}$$

$$\alpha^2 \frac{\partial \xi_3}{\partial R} = 0. \tag{11}$$

From (9), it follows that  $\eta_3$  must be of the form

$$\eta_3(t, R) = C_1(R) e^{\alpha t}.$$

From (11), it follows that the tangent in the time direction is univariate implying that  $\xi_3(t, R) = \xi_3(t)$ . Plugging in these two expressions into (10) results in the following equation

$$C_1'(R) e^{\alpha t} = \frac{\partial \xi_3}{\partial t}$$

and since  $\xi_3$  is independent of  $R$  it must hold that  $C_1'(R) = C_2$  is a constant and hence

$$C_1(R) = C_2 R + C_3$$

for two arbitrary constants  $C_2, C_3$ . Moreover, the solution to the above equation is given by the following calculations

$$\begin{aligned} \frac{\partial \xi_3}{\partial t} &= C_2 e^{\alpha t} \\ \implies \xi_3(t) &= C_4 + \frac{C_2}{\alpha} e^{\alpha t} \end{aligned}$$

for an additional arbitrary constant  $C_4$ . In summary, the tangents are given by

$$\begin{aligned} \xi_3(t, R) &= C_4 + \frac{C_2}{\alpha} e^{\alpha t}, \\ \eta_3(t, R) &= (C_2 R + C_3) e^{\alpha t}. \end{aligned}$$

Now, by plugging in these expressions into (5), it is possible to organise the expressions according to the arbitrary constants  $C_2, C_3$  and  $C_4$  respectively. It follows that the most general infinitesimal generator of the Lie group  $X_1$  for the IM-I in (1) is given by

$$\begin{aligned} X_3 &= \left( C_4 + \frac{C_2}{\alpha} e^{\alpha t} \right) \partial_t + ((C_2 R + C_3) e^{\alpha t}) \partial_R \\ &= C_2 \left[ \underbrace{\frac{1}{\alpha} e^{\alpha t} \partial_t + R e^{\alpha t} \partial_R}_{=X_{3,0}} \right] + C_4 \left[ \underbrace{\partial_t}_{=X_{3,1}} \right] + C_3 \left[ \underbrace{e^{\alpha t} \partial_R}_{=X_{3,2}} \right] \\ &= C_2 X_{3,0} + C_4 X_{3,1} + C_3 X_{3,2}, \\ \implies X_3 &= C_2 X_{3,0} + C_4 X_{3,1} + C_3 X_{3,2}. \end{aligned}$$

Note that the generator  $X_3$  in fact corresponds to a set of generators. This set of generators is called the *Lie Algebra* which is in fact a special type of vector space<sup>1</sup> [3]. Thus, the set of generators is spanned by  $\{X_{3,0}, X_{3,1}, X_{3,2}\}$  which forms a basis of the Lie algebra of the model in (1) and thus we have the following three generators to choose from:

$$X_{3,0} = \frac{1}{\alpha} e^{\alpha t} \partial_t + R e^{\alpha t} \partial_R, \quad (12)$$

$$X_{3,1} = \partial_t, \quad (13)$$

$$X_{3,2} = e^{\alpha t} \partial_R. \quad (14)$$

---

<sup>1</sup>A Lie Algebra is a vector space that is also equipped with an operation referred to as the Lie bracket and this type of vector space is closed under the Lie Bracket in addition to the standard requirements of a vector space.



A non-trivial generator is chosen for the model selection procedure. Thus, each of these candidate generators are subsequently analysed by means of the reduced characteristic  $\overline{Q}$  in (8). For (12), the following holds

$$\overline{Q}(X_{3,0}) = \eta_{3,0} - \omega_3 \xi_{3,0} = R e^{\alpha t} - (\alpha R) \left( \frac{1}{\alpha} e^{\alpha t} \right) = 0$$

and hence the generator  $X_{3,0}$  is *trivial*. For (13), the following holds

$$\overline{Q}(X_{3,1}) = \eta_{3,2} - \omega_3 \xi_{3,2} = 0 - (\alpha R) (1) = -\alpha R \neq 0$$

and hence the generator  $X_{3,1}$  is *non-trivial*. For (14), the following holds

$$\overline{Q}(X_{3,2}) = \eta_{3,2} - \omega_3 \xi_{3,2} = e^{\alpha t} - (\alpha R) (0) = e^{\alpha t} \neq 0$$

and hence the generator  $X_{3,2}$  is *non-trivial*. Thus, both symmetries stemming from the two non-trivial generators  $X_{3,1}$  and  $X_{3,2}$  will be calculated next.

The generator  $X_{3,1}$  is so well-known that it has a name and it is called the *translation generator*. The reason why it is well-known is because it is common to all autonomous models. In order to find the corresponding symmetry, it is necessary to plug in the translation generator  $X_{3,1}$  in (13) into the exponential map according to (6).

For the transformed coordinate  $\hat{t}(\epsilon)$ , it is possible to apply the generator  $X_{3,1}$  recursively on  $t$ . It follows that

$$\begin{aligned} X_{3,1}^0(t) &= t, \\ X_{3,1}^1(t) &= \partial t(t) = 1, \\ X_{3,1}^2(t) &= X_{3,1}(X_{3,1}(t)) = X_{3,1}(1) = \partial t(1) = 0, \\ X_{3,1}^3(t) &= X_{3,1}(X_{3,1}(t)^2) = X_{3,1}(0) = \partial t(0) = 0, \\ &\vdots \end{aligned}$$

and thus it is only the first two terms in the exponential map that are non-zero. Note that the zeroth power of any generator corresponds to the trivial generator which leaves the point it transforms unchanged which is why the very first term satisfies  $X_{3,1}^0(t) = t$ . By (6), it follows that

$$e^{\epsilon X_{3,1}} t = \sum_{j=0}^{\infty} \frac{\epsilon^j}{j!} X_{3,1}^j(t) = t + \epsilon.$$

Similarly, for the transformed coordinate  $\hat{R}(\epsilon)$ , the generator  $X_{3,1}$  is applied recursively on  $R$  which gives

$$\begin{aligned} X_{3,1}^0(R) &= R, \\ X_{3,1}^1(R) &= \partial t(R) = 0, \\ X_{3,1}^2(R) &= X_{3,1}(X_{3,1}(R)) = X_{3,1}(0) = \partial t(0) = 0, \\ X_{3,1}^3(R) &= X_{3,1}(X_{3,1}(R)^2) = X_{3,1}(0) = \partial t(0) = 0, \\ &\vdots \end{aligned}$$

and thus all terms are zero for this coordinate with the exception of the zeroth trivial term  $X_{3,1}^0$ . By (6), the second coordinate is given by

$$e^{\epsilon X_{3,1}} R = \sum_{j=0}^{\infty} \frac{\epsilon^j}{j!} X_{3,1}^j(R) = R.$$

In conclusion, the translation symmetry  $\Gamma_{3,1}(\epsilon)$  associated to the IM-I in (1) is given by

$$\Gamma_{3,1}(\epsilon) : (t, R) \mapsto (t + \epsilon, R).$$

For most generators including the generator  $X_{3,2} = e^{\alpha t} \partial R$ , the corresponding symmetry is calculated by using a change of coordinates. More precisely, instead of calculating the symmetry using the exponential map it is beneficial to use a coordinate change from  $(t, R)$  to, say,  $(s, r)$  where the latter coordinates have the same properties as the translation generator in  $X_{3,1}$ . These coordinates are called the *canonical coordinates* [3] and they satisfy the following criteria

$$\Gamma_{3,2} : (s, r) \mapsto (s + \epsilon, r).$$

In other words, the transformed canonical coordinates satisfy  $(\hat{s}(\epsilon), \hat{r}(\epsilon)) = (s + \epsilon, r)$  which is equivalently formulated in terms of the infinitesimal generator of the Lie group by the equation  $X_{3,2} = \partial_s$ . The coordinate change from  $(t, R)$  to  $(s, r)$  is characterised by the following equations [3]

$$\begin{aligned} \xi_{3,2}(t, R) \frac{\partial r}{\partial t} + \eta_{3,2}(t, R) \frac{\partial r}{\partial R} &= 0, \\ \xi_{3,2}(t, R) \frac{\partial s}{\partial t} + \eta_{3,2}(t, R) \frac{\partial s}{\partial R} &= 1. \end{aligned}$$

In the case of the generator  $X_{3,2}$ , the tangents are given by  $(\xi_{3,2}, \eta_{3,2}) = (0, e^{\alpha t})$  which results in the equations

$$\begin{aligned} e^{\alpha t} \frac{\partial r}{\partial R} &= 0, \\ e^{\alpha t} \frac{\partial s}{\partial R} &= 1. \end{aligned}$$

The solution to the first equation is that  $r$  is a univariate function according to  $r(t, R) = r(t)$  so we might as well choose  $r = t$  implying that the symmetry will not move points in the  $t$ -direction which is in fact the case as  $\xi_{3,2} = 0$ . The second equation yields the canonical coordinate

$$s = R e^{-\alpha t} = R e^{-\alpha r}.$$

By solving this equation for  $R$  it follows that

$$R = s e^{\alpha r}.$$

Now, the transformed coordinate  $\hat{t}$  is defined in terms of the canonical coordinate  $r$  as follows

$$\hat{t}(\epsilon) = \hat{r}(\epsilon) = r = t.$$

The transformed coordinate  $\hat{R}$  is defined in terms of both canonical coordinates  $s$  and  $r$  and is given by the following calculations

$$\begin{aligned}\hat{R}(\epsilon) &= \hat{s}(\epsilon)e^{\alpha\hat{r}(\epsilon)} \\ &= (s + \epsilon)e^{\alpha r} \\ &= \underbrace{se^{\alpha r}}_{=R} + \epsilon e^{\alpha r} = R + \epsilon e^{\alpha t} \\ \implies \hat{R}(\epsilon) &= R + \epsilon e^{\alpha t}.\end{aligned}$$

Consequently, the infinitesimal generator of the Lie group given by

$$X_{3,2} = e^{\alpha t} \partial R$$

generates the following non-trivial symmetry

$$\Gamma_{3,2}(\epsilon) : (t, R) \mapsto (t, R + \epsilon e^{\alpha t}).$$

### 1.3.2 The PLM

Denote the tangents in the infinitesimal generator of the Lie group for the PLM by  $\xi_1(t, R)$  and  $\eta_1(t, R)$  respectively. The reaction term in (1) is given by

$$\omega_1(t, R) = \frac{\gamma R}{t}.$$

For the partial derivatives in the linearised symmetry condition, we have

$$\frac{\partial \omega_1}{\partial t} = -\frac{\gamma R}{t^2}, \quad \frac{\partial \omega_1}{\partial R} = \frac{\gamma}{t} \quad \text{and} \quad \omega_1^2 = \frac{\gamma^2 R^2}{t^2}.$$

Plugging in these derivatives into the linearised symmetry condition yields the following equation

$$\frac{\partial \eta_1}{\partial t} + \left( \frac{\partial \eta_1}{\partial R} - \frac{\partial \xi_1}{\partial t} \right) \frac{\gamma R}{t} - \frac{\partial \xi_1}{\partial R} \frac{\gamma^2 R^2}{t^2} = -\xi_1 \frac{\gamma R}{t^2} + \eta_1 \frac{\gamma}{t}$$

which can be organised according to the different powers of  $R$

$$\left[ \frac{\partial \eta_1}{\partial t} - \eta_1 \frac{\gamma}{t} \right] R^0 + \left[ \left( \frac{\partial \eta_1}{\partial R} - \frac{\partial \xi_1}{\partial t} \right) \frac{\gamma}{t} + \xi_1 \frac{\gamma}{t^2} \right] R^1 - \left[ \frac{\partial \xi_1}{\partial R} \frac{\gamma^2}{t^2} \right] R^2 = 0.$$

Since the monomials  $\{R^0, R^1, R^2\}$  are linearly independent, the following determining equations must hold

$$\frac{\partial \eta_1}{\partial t} - \eta_2 \frac{\gamma}{t} = 0, \quad (15)$$

$$\left( \frac{\partial \eta_1}{\partial R} - \frac{\partial \xi_1}{\partial t} \right) \frac{\gamma}{t} + \xi_1 \frac{\gamma}{t^2} = 0, \quad (16)$$

$$\frac{\partial \xi_1}{\partial R} \frac{\gamma^2}{t^2} = 0. \quad (17)$$

By (17), the tangent  $\xi$  is a univariate function of the variable  $t$ , i.e.  $\xi(t, R) = \xi(t)$ . By (15), the following calculations hold

$$\begin{aligned} \frac{d\eta_1}{dt} &= \frac{\gamma \eta_1}{t}, \\ \implies \frac{d\eta_1}{\eta_1} &= \gamma \frac{dt}{t}, \\ \implies \ln(\eta_1) &= \gamma \ln(t) + \tilde{C}_1(R) = \ln(t^\gamma) + \tilde{C}_1(R), \\ \implies \eta_1(t, R) &= \exp \left( \underbrace{\tilde{C}_1(R)}_{=C_1(R)} \right) t^\gamma = C_1(R) t^\gamma, \end{aligned}$$

where  $C_1(R)$  is an arbitrary function of  $R$ . By plugging in these two expressions, into (16) the following equation must hold

$$\left( C_1'(R) t^\gamma - \frac{\partial \xi_1}{\partial t} \right) \frac{\gamma}{t} + \xi_1 \frac{\gamma}{t^2} = 0.$$

and since  $\xi_2$  is independent of  $t$  it must follow that  $C_1'(R) = C_2$  and  $C_1(R) = C_2 R + C_3$  for two arbitrary integration constants  $C_2, C_3$ . Moreover, the solution to the last ODE is given by the following calculations:

$$\begin{aligned}
C_2 t^\gamma - \frac{\partial \xi_1}{\partial t} + \frac{\xi_1}{t} &= 0 \\
\implies \frac{\partial \xi_1}{\partial t} &= \frac{\xi_1}{t} + C_2 t^\gamma \\
\left\{ \text{Multiply with integrating factor : } \exp(-\ln(t)) = \frac{1}{\exp(\ln(t))} = \frac{1}{t} \right\} \\
\implies \frac{1}{t} \frac{\partial \xi_1}{\partial t} &= \frac{\xi_1}{t^2} + C_2 t^{\gamma-1} \\
\implies \frac{1}{t} \frac{\partial \xi_1}{\partial t} - \frac{\xi_1}{t^2} &= C_2 t^{\gamma-1} \\
\implies \frac{d}{dt} \left( \frac{1}{t} \xi_1(t) \right) &= C_2 t^{\gamma-1} \\
\implies \frac{1}{t} \xi_1(t) &= \int C_2 t^{\gamma-1} dt = \frac{C_2}{\gamma} t^\gamma + C_4 \\
\implies \xi_1(t) &= \frac{C_2}{\gamma} t^{\gamma+1} + C_4 t
\end{aligned}$$

where  $C_4$  is an additional arbitrary coefficient. Thus, the tangents for the model in (2) are given by

$$\begin{aligned}
\xi_1(t) &= \frac{C_2}{\gamma} t^{\gamma+1} + C_4 t, \\
\eta_1(t, R) &= (C_2 R + C_3) t^\gamma.
\end{aligned}$$

Similarly, as with the previous generator it is possible to organise the most general form of the infinitesimal generator of the Lie group  $X_1$  for the PLM in (2) with respect to the integration constants  $C_2$ ,  $C_3$  and  $C_4$  respectively. This organisation is summarised in the following calculations

$$\begin{aligned}
X_1 &= \left[ \frac{C_2}{\gamma} t^{\gamma+1} + C_4 t \right] \partial_t + [(C_2 R + C_3) t^\gamma] \partial_R \\
&= C_2 \left[ \underbrace{\frac{t^{\gamma+1}}{\gamma} \partial_t + R t^\gamma \partial_R}_{=X_{1,0}} \right] + C_4 \left[ \underbrace{t \partial_t}_{=X_{1,1}} \right] + C_3 \left[ \underbrace{t^\gamma \partial_R}_{X_{1,2}} \right] \\
&= C_2 X_{1,0} + C_4 X_{1,1} + C_3 X_{1,2}.
\end{aligned}$$

In similarity with the previous model, the Lie algebra of the PLM in (2) is spanned by the

set of basis generators  $\{X_{2,1}, X_{2,2}, X_{2,3}\}$  listed below

$$\begin{aligned} X_{1,0} &= \frac{t^{\gamma+1}}{\gamma} \partial_t + Rt^\gamma \partial_R, \\ X_{1,1} &= t \partial_t, \\ X_{1,2} &= t^\gamma \partial_R. \end{aligned}$$

Similarly as before, these three generators will now be classified as either trivial or non-trivial. For the first generator, the reduced characteristic  $\overline{Q}$  in (8) is given by

$$\overline{Q}(X_{1,0}) = \eta_{1,0} - \omega_1 \xi_{1,0} = Rt^\gamma - \left( \frac{\gamma R}{t} \right) \left( \frac{t^{\gamma+1}}{\gamma} \right) = 0$$

and hence the generator  $X_{1,0}$  is trivial. For the second generator, it follows that

$$\overline{Q}(X_{1,1}) = \eta_{1,1} - \omega_1 \xi_{1,1} = 0 - \left( \frac{\gamma R}{t} \right) (t) = \gamma R \neq 0$$

and hence the generator  $X_{1,1}$  is non-trivial. For the third generator, it follows that

$$\overline{Q}(X_{1,2}) = \eta_{1,2} - \omega_1 \xi_{1,2} = t^\gamma - \left( \frac{\gamma R}{t} \right) (0) = t^\gamma \neq 0$$

and hence the generator  $X_{1,2}$  is non-trivial. Thus, the symmetries generated by  $X_{1,1}$  and  $X_{1,2}$  respectively will be calculated next.

Just as the translation generator, the generator  $X_{1,1} = t \partial_t$  has a name which is the *scaling generator*. Its symmetry can be calculated using the exponential map as in (6). By applying the generator recursively to  $t$ , it follows that

$$\begin{aligned} X_{1,1}^0(t) &= t \\ X_{1,1}^1(t) &= \underbrace{t \partial_t(t)}_{=1} = t \\ X_{1,1}^2(t) &= X_{1,1}(X_{1,1}^1(t)) = \underbrace{X_{1,1}(t)}_{=t} = t \\ X_{1,1}^3(t) &= X_{1,1}(X_{1,1}^2(t)) = \underbrace{X_{1,1}(t)}_{=t} = t \\ &\vdots \end{aligned}$$

and hence all terms in the exponential map equals  $t$  for the scaling generator. Thus, it follows that the transformed coordinate is given by

$$e^{\epsilon X_{1,1}} t = \sum_{j=0}^{\infty} \frac{\epsilon^j}{j!} \underbrace{X_{1,1}^j(t)}_{=t} = t \sum_{j=0}^{\infty} \frac{\epsilon^j}{j!} = \underbrace{te^\epsilon}_{=e^\epsilon}.$$

The  $R$ -coordinate is unaffected by the action of the symmetry implying that

$$e^{\epsilon X_{1,1}} R = R$$

and hence the *scaling symmetry*  $\Gamma_{2,3}$  is given by

$$\Gamma_{1,1}(\epsilon) : (t, R) \mapsto (te^\epsilon, R).$$

To calculate the symmetry generated by  $X_{1,2} = t^\gamma \partial_R$ , the method based on the canonical coordinates will be used. Note that for this generator, the tangents are given by  $\xi_{1,2} = 0$  and  $\eta_{1,2} = t^\gamma$  respectively and the canonical coordinates  $(s, r)$  satisfy the following equations

$$\begin{aligned} \xi_{1,2}(t, R) \frac{\partial r}{\partial t} + \eta_{1,2}(t, R) \frac{\partial r}{\partial R} &= 0, \\ \xi_{1,2}(t, R) \frac{\partial s}{\partial t} + \eta_{1,2}(t, R) \frac{\partial s}{\partial R} &= 1. \end{aligned}$$

By plugging in the tangents  $\xi_{1,2}$  and  $\eta_{1,2}$ , these equations become

$$\begin{aligned} t^\gamma \frac{\partial r}{\partial R} &= 0, \\ t^\gamma \frac{\partial s}{\partial R} &= 1. \end{aligned}$$

By the same arguments as for the previous model, since the generator  $X_{1,2}$  does not transform points in the  $t$ -direction the second canonical coordinate is set to  $r = t$ . Consequently, the transformed coordinate for  $t$  is given by

$$\hat{t}(\epsilon) = \hat{r}(\epsilon) = r = t.$$

The first canonical coordinate is given by

$$s = \frac{R}{t^\gamma}$$

and solving this equation for  $R$  as well as re-writing the equation in terms of the canonical coordinates yields

$$R = st^\gamma = sr^\gamma.$$

Lastly, the transformed coordinate for  $R$  is given by

$$\hat{R}(\epsilon) = \hat{s}(\epsilon) \hat{r}(\epsilon)^\gamma = (s + \epsilon) r^\gamma = \underbrace{sr^\gamma}_{=R} + \epsilon r^\gamma = R + \epsilon t^\gamma.$$

In summary, the infinitesimal generator of the Lie group generated by  $X_{2,2} = t^\gamma \partial_R$  generates the following symmetry

$$\Gamma_{1,2}(\epsilon) : (t, R) \mapsto (t, R + \epsilon t^\gamma).$$

### 1.3.3 The IM-II

Denote the tangents of the IM-II by  $\xi_2(t, R)$  and  $\eta_2(t, R)$  respectively. The reaction term for this model is given by

$$\omega_2(t, R) = \frac{\alpha}{A} R^2 e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)})$$

and the linearised symmetry condition is given by

$$\frac{\partial \eta_2}{\partial t} + \left( \frac{\partial \eta_2}{\partial R} - \frac{\partial \xi_2}{\partial t} \right) \omega_2 - \frac{\partial \eta_2}{\partial t} \omega_2^2 = \xi_3 \frac{\partial \omega_2}{\partial t} + \eta_3 \frac{\partial \omega_2}{\partial R}.$$

Thus, the task at hand is to calculate the three terms  $\omega_2^2$ ,  $\partial \omega_2 / \partial t$  and  $\partial \omega_2 / \partial R$  in order to plug these expressions into the linearised symmetry condition above. For the quadratic term, it follows that the middle factor satisfies

$$(\exp(e^{-\alpha(t-\tau)}))^2 = \exp(e^{-\alpha(t-\tau)}) \exp(e^{-\alpha(t-\tau)}) = \exp(2e^{-\alpha(t-\tau)})$$

and hence the quadratic term is given by

$$\omega_2^2 = \frac{\alpha^2}{A^2} R^4 e^{-2\alpha(t-\tau)} \exp(2e^{-\alpha(t-\tau)}).$$

For the time derivative, it follows that

$$\begin{aligned} \frac{\partial \omega_2}{\partial t} &= \frac{\partial}{\partial t} \left[ \frac{\alpha}{A} e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) R^2 \right] \\ &= \frac{\alpha R^2}{A} [-\alpha e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) + e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) - \alpha e^{-\alpha(t-\tau)}] \\ &= -\frac{\alpha^2 R^2}{A} [e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) + e^{-2\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)})] \end{aligned}$$

holds. For the derivative with respect to  $R$ , we have

$$\frac{\partial \omega_2}{\partial R} = 2 \frac{\alpha}{A} e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) R.$$

Now, plugging in these expressions into the linearised symmetry condition leads to the following equation

$$\begin{aligned} \frac{\partial \eta_2}{\partial t} + \left( \frac{\partial \eta_2}{\partial R} - \frac{\partial \xi_2}{\partial t} \right) \left[ \frac{\alpha}{A} R^2 e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) \right] - \frac{\partial \xi_2}{\partial R} \left[ \frac{\alpha^2}{A^2} R^4 e^{-2\alpha(t-\tau)} \exp(2e^{-\alpha(t-\tau)}) \right] \\ = \xi_2 \left( -\frac{\alpha^2 R^2}{A} [e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) + e^{-2\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)})] \right) \\ + \eta_2 \left( 2 \frac{\alpha}{A} e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) R \right). \end{aligned}$$



Since the monomials  $\{R^0, R^1, R^2, R^4\}$  are *linearly independent*, it follows that the following four equations must hold

$$R^0 : \quad \frac{\partial \eta_2}{\partial t} = 0, \quad (18)$$

$$R^1 : \quad 2\eta_2 \frac{\alpha}{A} e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) = 0, \quad (19)$$

$$R^2 : \quad \frac{\alpha}{A} e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) \left[ \left( \frac{\partial \eta_2}{\partial R} - \frac{\partial \xi_2}{\partial t} \right) + \alpha \xi_2 (1 + e^{-\alpha(t-\tau)}) \right] = 0, \quad (20)$$

$$R^4 : \quad \left[ \frac{\alpha^2}{A^2} e^{-2\alpha(t-\tau)} \exp(2e^{-\alpha(t-\tau)}) \right] \frac{\partial \xi_2}{\partial R} = 0. \quad (21)$$

The tangent in the  $R$ -direction  $\eta_2$  must satisfy  $\eta_2 = 0$  according to (18) and (19). Moreover, the tangent in the  $t$ -direction must be a univariate function of  $t$ , i.e.  $\xi_2(t, R) = \xi_2(t)$ , as it follows that  $\partial \xi_2 / \partial R = 0$  according to (21). Lastly, tangent  $\xi_2$  in the  $t$ -direction must satisfy the following ODE

$$\frac{d\xi_2}{dt} = \alpha \xi_2 (1 + e^{-\alpha(t-\tau)})$$

according to (20). Now, by defining the function  $f(t) = \alpha (1 + e^{-\alpha(t-\tau)})$  its primitive function is

$$F(t) = \alpha t - e^{-\alpha(t-\tau)}.$$

Thus, by multiplying the above ODE with the integrating factor “ $\exp(-F(t))$ ”, the resulting equation is

$$\underbrace{\exp(-(\alpha t - e^{-\alpha(t-\tau)})) \frac{d\xi_2}{dt} - \underbrace{\exp(-(\alpha t - e^{-\alpha(t-\tau)})) \alpha (1 + e^{-\alpha(t-\tau)})}_{= -\frac{\partial \exp(-F(t))}{\partial t}} \xi_2}_{= \frac{\partial}{\partial t} (\exp(-F(t)) \xi_2) \left\{ = \exp(-F(t)) \frac{d\xi_2}{dt} + \frac{d \exp(-F(t))}{dt} \xi_2 \right\}} = 0$$

which can be re-written by using the product rule as follows

$$\frac{\partial}{\partial t} (\exp(-(\alpha t - e^{-\alpha(t-\tau)})) \xi_2) = 0.$$

Finally, by integrating with respect to  $t$  the tangent is given by

$$\xi_2(t) = C_1 \exp((\alpha t - e^{-\alpha(t-\tau)})) = C_1 e^{\alpha t} \exp(-e^{-\alpha(t-\tau)})$$

for an arbitrary integration constant  $C_1$ . Hence, the generator for the IM-II is given by

$$X_2 = e^{\alpha t} \exp(-e^{-\alpha(t-\tau)}) \partial_t.$$

The generator is non-trivial as

$$\begin{aligned} \overline{Q}(X_2) &= \eta_2 - \omega_2 \xi_2 = 0 - \left[ \frac{\alpha}{A} R^2 e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) \right] [e^{\alpha t} \exp(-e^{-\alpha(t-\tau)})] \\ &= \frac{\alpha}{A} R^2 e^{\alpha \tau} \neq 0. \end{aligned}$$

To calculate the symmetry  $\Gamma_2$  generated by the infinitesimal generator of the Lie group  $X_2$ , the approach based on the canonical coordinates is used. For the generator  $X_2$ , the tangents are given by

$$(\xi_2, \eta_2) = (e^{\alpha t} \exp(-e^{-\alpha(t-\tau)}), 0).$$

The canonical coordinates  $(s, r)$  satisfy the following equations

$$\begin{aligned}\xi_2(t, R) \frac{\partial r}{\partial t} + \eta_2(t, R) \frac{\partial r}{\partial R} &= 0, \\ \xi_2(t, R) \frac{\partial s}{\partial t} + \eta_2(t, R) \frac{\partial s}{\partial R} &= 1,\end{aligned}$$

which by plugging in the tangents  $(\xi_2, \eta_2)$  results in the following equations

$$\begin{aligned}e^{\alpha t} \exp(-e^{-\alpha(t-\tau)}) \frac{\partial r}{\partial t} &= 0, \\ e^{\alpha t} \exp(-e^{-\alpha(t-\tau)}) \frac{\partial s}{\partial t} &= 1.\end{aligned}$$

Since the generator  $X_2$  is unidirectional in the sense that it only moves points in the  $t$ -direction and as  $\partial r / \partial t = 0$  we can set  $r = R$ . From the second equation, it follows that

$$\begin{aligned}s &= \int e^{-\alpha t} \exp(e^{-\alpha(t-\tau)}) dt \\ &= -\frac{1}{\alpha e^{\alpha\tau}} \int \underbrace{(-\alpha e^{-\alpha t})}_{\frac{de^{-\alpha t}}{dt}} \underbrace{(e^{\alpha\tau} \exp(e^{\alpha\tau} e^{-\alpha t}))}_{\text{Outer derivative}} dt \\ \{\text{Chain rule}\} \implies &= -\frac{1}{\alpha e^{\alpha\tau}} \int \frac{d}{dt} (\exp(e^{\alpha\tau} e^{-\alpha t})) dt \\ \implies s &= -\frac{e^{-\alpha\tau} \exp(e^{-\alpha(t-\tau)})}{\alpha}.\end{aligned}$$

In order to be able to take the logarithm of the above expression, we must work with the absolute value as taking the logarithm is only allowed on positive expressions. Therefore, the value of  $t$  is calculated in the following steps

$$\begin{aligned}s &= -\frac{e^{-\alpha\tau} \exp(e^{-\alpha(t-\tau)})}{\alpha} \\ \implies -\alpha e^{\alpha\tau} s &= \exp(e^{-\alpha(t-\tau)}) \\ \implies |\alpha e^{\alpha\tau} s| &= \exp(e^{-\alpha(t-\tau)}) \\ \implies -\alpha(t-\tau) &= \ln(\ln(|\alpha e^{\alpha\tau} s|)) \\ \implies t &= \tau - \frac{\ln(\ln(|\alpha e^{\alpha\tau} s|))}{\alpha}.\end{aligned}$$

Now, using this last equation, the transformed coordinate  $\hat{t}$  is given by

$$\begin{aligned}\hat{t} &= \tau - \frac{\ln(\ln(|\alpha e^{\alpha\tau} \hat{s}|))}{\alpha} \\ &= \tau - \frac{\ln(\ln(|\alpha e^{\alpha\tau}(s + \epsilon)|))}{\alpha} \\ &= \tau - \frac{\ln(\ln(|\alpha e^{\alpha\tau}\epsilon - \exp(e^{-\alpha(t-\tau)})|))}{\alpha}.\end{aligned}$$

Consequently, the infinitesimal generator of the Lie group denoted by  $X_2$  generates the following symmetry

$$\Gamma_2(\epsilon) : (t, R) \mapsto \left( \tau - \frac{\ln(\ln(|\alpha e^{\alpha\tau}\epsilon - \exp(e^{-\alpha(t-\tau)})|))}{\alpha}, R \right).$$

Note that the value  $\epsilon = 0$  indeed corresponds to the trivial symmetry as  $\Gamma_2(0) : (t, R) \mapsto (t, R)$  which is a required property of any symmetry.

## 2 Summary of the calculated symmetries

In the article, the symmetries of the IM-I is omitted as its fit is much worse than the other two candidate models. Although, the emphasis is on the comparison between the PLM and the IM-II the symmetries of the IM-I have also been calculated. A summary of all the calculated symmetries is presented below (Tab 2).

A commonality between the three models is that all non-trivial generators are uni-directional. Both the PLM and the IM-I have  $R$ -directional symmetries, i.e.  $\Gamma_{1,2}$  and  $\Gamma_{3,2}$  whereas the IM-II merely has a  $t$ -directional symmetry in  $\Gamma_2$ . The action of the non-trivial symmetries of the PLM (Fig 1), the IM-II (Fig 2) and the IM-I (Fig 3) respectively is illustrated below.

## 3 Processing of the data and the curve fitting

In order to conduct the curve fitting, the data is processed in two steps (Fig 4). The data in this case is a time series containing the age of the subjects denoted by  $t$  and the number of incidences of Myeloma cancer which is a type of blood cancer as a function of the age denoted by  $R(t)$  [1, 5]. As the number of incidences of Myeloma is zero for most ages below the age of 25 years, the first 25 data points are removed in the time series. In addition, the fitting to the data is conducted on the logarithm of the incidences, i.e. the models are fitted to a modified time series containing ages  $t$  versus the logarithm of the incidences  $\ln(R(t))$ . Provided the fact that it is the logarithm of the incidences that are reported, it is necessary to fit the logarithmic version of the models.

The logarithmic version of the PLM given by

Table 2: **Summary of the symmetries.** The models, their generators and the non-trivial symmetries are presented from left to right.

<i>Model</i>	Generators	Non-trivial Symmetries
PLM	$X_{1,0} = \left(\frac{t^{\gamma+1}}{\gamma}\right) \partial_t + Rt^\gamma \partial_R$ $X_{1,1} = t \partial_t$ $X_{1,2} = t^\gamma \partial_R$	$\Gamma_{1,1}(\epsilon) : (t, R) \mapsto (te^\epsilon, R)$ $\Gamma_{1,2}(\epsilon) : (t, R) \mapsto (t, R + \epsilon t^\gamma)$
IM-II	$X_2 = e^{\alpha t} \exp(-e^{-\alpha(t-\tau)}) \partial_t$	$\Gamma_2(\epsilon) : (t, R) \mapsto \left( \tau - \frac{\ln(\ln( \alpha e^{\alpha\tau}\epsilon - \exp(e^{-\alpha(t-\tau)}) ))}{\alpha}, R \right)$
IM-I	$X_{3,0} = \frac{1}{\alpha} e^{\alpha t} \partial_t + R e^{\alpha t} \partial_R$ $X_{3,1} = \partial_t$ $X_{3,2} = e^{\alpha t} \partial_R$	$\Gamma_{3,1}(\epsilon) : (t, R) \mapsto (t + \epsilon, R)$ $\Gamma_{3,2} : (t, R) \mapsto (t, R + \epsilon e^{\alpha t})$

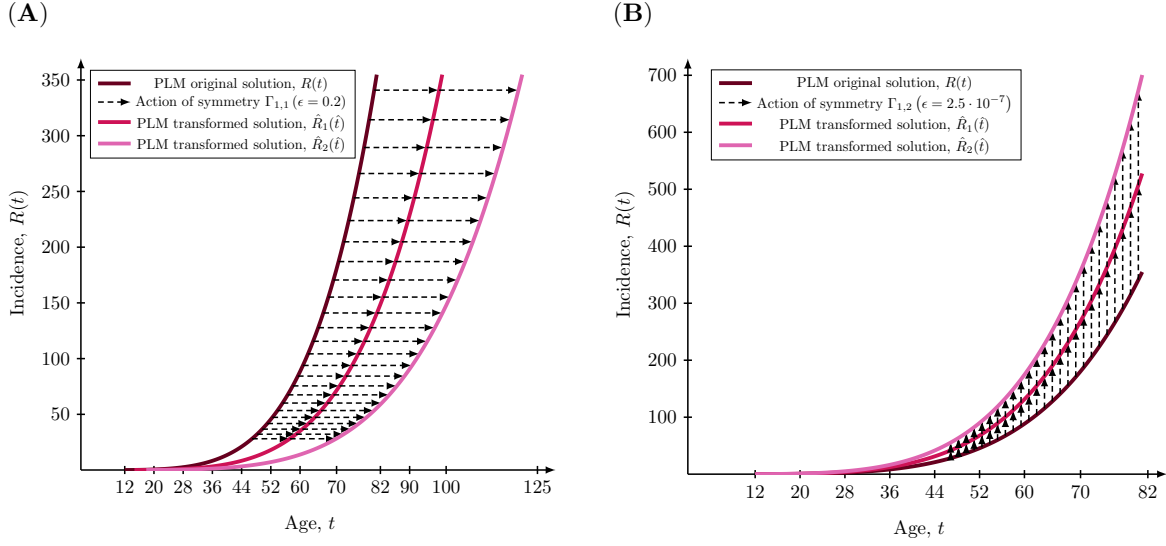


Figure 1: *Action of the non-trivial symmetries of the PLM.* The original curve which is generated from the PLM with parameters  $(A, \gamma) = (7.742, 4.528)$  is transformed twice with the two non-trivial symmetries of the PLM. **(A)** The  $t$ -directional scaling symmetry  $\Gamma_{1,1}$  with transformation parameter  $\epsilon = 0.2$ . **(B)** The  $R$ -directional symmetry  $\Gamma_{1,2}$  with transformation parameter  $\epsilon = 2.5 \cdot 10^{-7}$ .

$$R(t) = At^\gamma$$

is given by

$$\ln(R(t)) = \ln(A) + \gamma \ln(t).$$

The logarithmic version of the IM-II given by

$$R(t) = \frac{A}{\exp(e^{-\alpha(t-\tau)}) - 1}$$

is given by

$$\ln(R(t)) = \ln(A) - \ln(\exp(e^{-\alpha(t-\tau)}) - 1).$$

Again, it is the logarithmic versions of the models that are fitted to the processed data (Fig 4C).

The fit of each model is evaluated based on the *adjusted  $R^2$  value* denoted by  $R_{\text{adj}}^2$ . If  $n$  corresponds to the number of data points in the time series and  $k$  corresponds to the number of parameters in each model then this value is defined as follows

$$R_{\text{adj}}^2 = 1 - \left[ \frac{(1 - R^2)(n - 1)}{n - k - 1} \right]$$

where

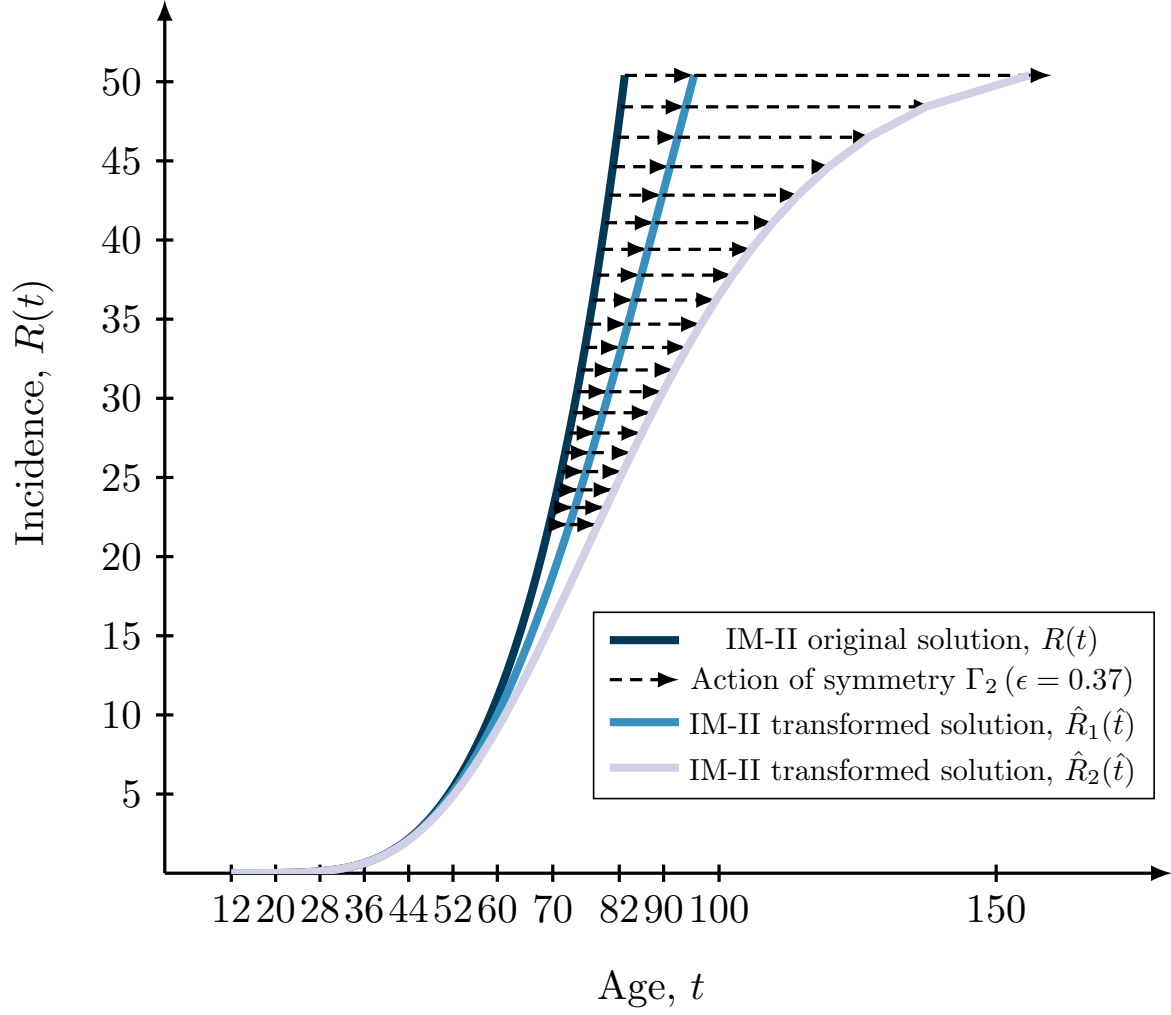


Figure 2: *Action of the non-trivial symmetry of the IM-II.* The original curve which is generated from the IM-II with parameters  $(\ln(A), \tau, \alpha) = (3.4822, 67.1564, 0.044)$  is transformed twice with the non-trivial symmetry of the IM-II. More precisely, the original curve is transformed twice with the  $t$ -directional symmetry  $\Gamma_2$  with transformation parameter  $\epsilon = 0.37$ .

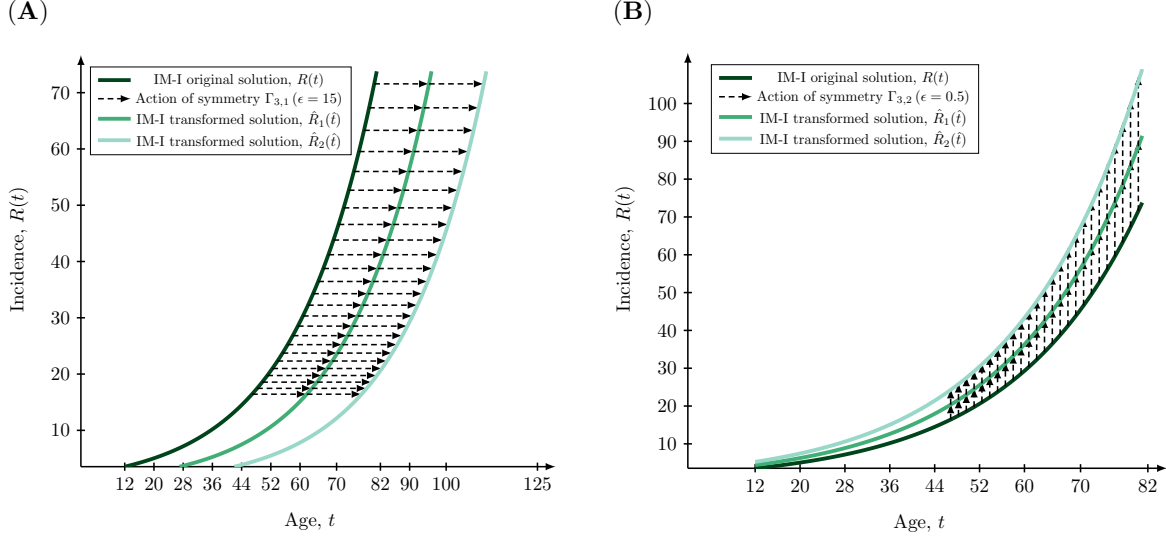


Figure 3: *Action of the non-trivial symmetries of the IM-I.* The original curve which is generated from the IM-I with parameters  $(A, \alpha) = (2.090, 0.044)$  is transformed twice with the two non-trivial symmetries of the PLM. (A) The  $t$ -directional translation symmetry  $\Gamma_{3,1}$  with transformation parameter  $\epsilon = 15$ . (B) The  $R$ -directional symmetry  $\Gamma_{3,2}$  with transformation parameter  $\epsilon = 0.5$ .

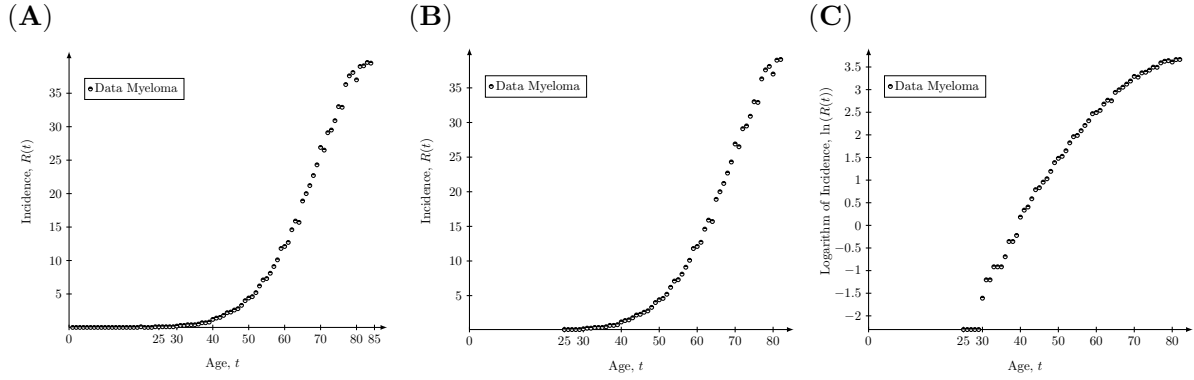


Figure 4: *Processing of the data.* The processing of the original data of the incidences of Myeloma in (A) is conducted in two steps. (B) Remove all zero incidences corresponding to all ages below 25 years. More precisely, this corresponds to the removal of the first 25 data points in the original time series. (C) Apply the logarithm to the incidences so that  $\ln(R(t))$  is reported as a function of the age  $t$ .

$$R^2 = 1 - \frac{\sum_{i=1}^n \left( \ln(R_i) - \ln(\tilde{R}(t_i)) \right)^2}{\sum_{i=1}^n \left( \ln(R_i) - \bar{R} \right)^2} = 1 - \frac{\text{Variation between model and data}}{\text{Internal variation in data}}.$$

In the above expression,  $R_i$  corresponds to the data points,  $\tilde{R}(t_i)$  corresponds to the output of the model and  $\bar{R}$  is the mean value of the *logarithmised data* (not the original data). The  $R^2_{\text{adj}}$  penalises the number of parameters, but as  $\alpha$  is kept fixed in the fitting procedure the number of parameters is  $k = 2$  for both the PLM and the IM-II. Otherwise, the  $R^2_{\text{adj}}$  shares the following two properties with the  $R^2$ :

1. A perfect fit corresponds to  $R^2_{\text{adj}} = 1$ ,
2. A really bad fit corresponds to  $R^2_{\text{adj}} < 0$  which says that the model fits the data worse than a straight line.

## 4 Analytical validation of the $t$ -directional generators of the PLM and the IM-II

The validation of the  $t$ -directional generators of the PLM and the IM-II is conducted using the *linearised symmetry condition*. More precisely, by plugging in the tangents of these generators into (7) on page 5 it is necessary that the resulting left hand side equals the resulting right hand side. Using this methodology, it is possible to validate the scaling generator of the PLM given by

$$X_{1,1} = t\partial_t$$

and the generator of the IM-II given by

$$X_2 = \left( e^{\alpha t} \exp(-e^{-\alpha(t-\tau)}) \right) \partial_t.$$

These two generators have two properties in common. Firstly, the tangent in the  $t$ -direction is univariate as  $\xi(t, R) = \xi(t)$  and the tangent in the  $R$ -direction is zero implying that  $\eta(t, R) = 0$ . Consequently, the linearised symmetry condition in (7) on page 5 decomposes to the following

$$-\frac{d\xi(t)}{dt} \omega(t, R) = \xi(t) \frac{\partial \omega(t, R)}{\partial t} \quad (22)$$

where  $\omega(t, R)$  is the reaction term of the ODE corresponding to a certain model. The reaction term of the PLM is given by

$$\omega(t, R) = \gamma \frac{R}{t}$$

and the reaction term of the IM-II is given by

$$\omega(t, R) = \frac{\alpha}{A} R^2 e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}).$$

Thus, using these reactions terms and the corresponding generators the validation procedure is conducted by controlling that (22) is satisfied.



For the scaling generator  $X_{1,1}$  of the PLM, the LHS of (22) equals the following

$$-\frac{d\xi(t)}{dt}\omega(t, R) = -\underbrace{\frac{d}{dt}(t)}_{=1}\gamma\frac{R}{t} = -\gamma\frac{R}{t}$$

and the RHS of (22) is given by the following calculations

$$\xi(t)\frac{\partial\omega(t, R)}{\partial t} = t\frac{\partial}{\partial t}\left(\gamma\frac{R}{t}\right) = -t\gamma\frac{R}{t^2} = -\gamma\frac{R}{t}.$$

Hence, the scaling generator  $X_{1,1}$  of the PLM satisfies the linearised symmetry condition in (22).

□

In the case of the generator  $X_2$  of the IM-II, the calculations are slightly longer but the methodology is the same. Starting with the LHS of (22), the first factor involving the time derivative is given by the following calculations

$$\begin{aligned}\frac{d\xi}{dt} &= \frac{d}{dt}\left(e^{\alpha t}\exp\left(-e^{-\alpha(t-\tau)}\right)\right) \\ &= \alpha e^{\alpha t}\exp\left(-e^{-\alpha(t-\tau)}\right) + e^{\alpha t}\exp\left(-e^{-\alpha(t-\tau)}\right)\alpha e^{-\alpha(t-\tau)} \\ &= \alpha e^{\alpha t}\exp\left(-e^{-\alpha(t-\tau)}\right)\left(1 + e^{-\alpha(t-\tau)}\right).\end{aligned}$$

Thus, it follows that the LHS of (22) in the case of the generator  $X_2$  for the IM-II is given by

$$\begin{aligned}-\frac{d\xi(t)}{dt}\omega(t, R) &= -\alpha e^{\alpha t}\exp\left(-e^{-\alpha(t-\tau)}\right)\left(1 + e^{-\alpha(t-\tau)}\right)\frac{\alpha}{A}R^2e^{-\alpha(t-\tau)}\exp\left(e^{-\alpha(t-\tau)}\right) \\ &= -\frac{\alpha^2 R^2}{A}e^{\alpha\tau}\left(1 + e^{-\alpha(t-\tau)}\right).\end{aligned}$$

Similarly, the second factor involving the time derivative in the RHS of (22) for the generator  $X_2$  of the IM-II is given by the following calculations

$$\begin{aligned}\frac{\partial\omega(t, R)}{\partial t} &= \frac{\partial}{\partial t}\left(\frac{\alpha}{A}R^2e^{-\alpha(t-\tau)}\exp\left(e^{-\alpha(t-\tau)}\right)\right) \\ &= \frac{\alpha}{A}R^2\left(-\alpha e^{-\alpha(t-\tau)}\exp\left(e^{-\alpha(t-\tau)}\right) + e^{-\alpha(t-\tau)}\exp\left(e^{-\alpha(t-\tau)}\right) - \alpha e^{-\alpha(t-\tau)}\right) \\ &= -\frac{\alpha^2 R^2}{A}e^{-\alpha(t-\tau)}\exp\left(e^{-\alpha(t-\tau)}\right)\left(1 + e^{-\alpha(t-\tau)}\right)\end{aligned}$$

and thus the RHS is given by

$$\begin{aligned}\frac{\partial \omega(t, R)}{\partial t} \xi(t) &= -\frac{\alpha^2 R^2}{A} e^{-\alpha(t-\tau)} \exp(e^{-\alpha(t-\tau)}) (1 + e^{-\alpha(t-\tau)}) (e^{\alpha t} \exp(-e^{-\alpha(t-\tau)})) \\ &= -\frac{\alpha^2 R^2}{A} e^{\alpha \tau} (1 + e^{-\alpha(t-\tau)}).\end{aligned}$$

Consequently, the generator  $X_2$  of the IM-II satisfies (22) since the LHS equals the RHS.

□

The symmetries corresponding to these generators, i.e.  $\Gamma_{1,1}$  and  $\Gamma_2$  respectively, were obtained by using the exponential map in (6) on page 5. Subsequently, these operators are validated numerically by using the defining property of a symmetry, namely that it maps solution curves to other solutions curves.

## 5 Numerical validation of the $t$ -directional symmetries of the PLM and the IM-II

To test the validity of the derived symmetries numerically, a curve fitting test based on the defining property of symmetries was conducted. A symmetry is defined by the fact that it maps one solution curve to another solution curve, and the test was implemented on the  $t$ -directional symmetries for the PLM as well as the IM-II. More specifically, the numerical validation was conducted in order to verify that the PLM given by

$$R(t) = At^\gamma$$

has a scaling-symmetry given by

$$\Gamma_{1,1}(\epsilon) : (t, R) \mapsto (te^\epsilon, R) \tag{23}$$

and that the IM-II given by

$$R(t) = \frac{A}{\exp(e^{-\alpha(t-\tau)}) - 1}$$

has a symmetry given by

$$\Gamma_2(\epsilon) : (t, R) \mapsto \left( \tau - \frac{\ln(\ln(|\alpha\epsilon - \exp(e^{-\alpha(t-\tau)})|))}{\alpha}, R \right).$$

In this section, the principles of the validation will be presented in detail for the PLM but the same steps are analogous for the IM-II.

For the PLM, the parameter  $\gamma$  is kept fixed as it determines the family of curves while the parameter  $A$  is estimated as it specifies the specific curve. Now, the validation is started

by obtaining a specific curve  $R(t) = A_0 t^\gamma$  for some fixed value  $A_0$  and then a time series is generated from this curve by gathering  $m \in \mathbb{N}_+$  realisations of this curve in a vector as follows

$$\mathbf{t}_0 = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{pmatrix} \quad \text{and} \quad \mathbf{R}_0 = \begin{pmatrix} A_0 t_1^\gamma \\ A_0 t_2^\gamma \\ \vdots \\ A_0 t_m^\gamma \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{pmatrix}.$$

Next, the scaling symmetry  $\Gamma_{1,1}(\epsilon)$  in (23) is applied to the time series given by  $\mathbf{t}_0$  and  $\mathbf{R}_0$  which yields a new time series determined by the transformation parameter  $\epsilon$ . Denote this transformed time series by  $\hat{\mathbf{t}}_0(\epsilon)$  and  $\hat{\mathbf{R}}_0(\epsilon)$  respectively and this time series looks as follows

$$\hat{\mathbf{t}}_0(\epsilon) = \begin{pmatrix} \hat{t}_1(\epsilon) \\ \hat{t}_2(\epsilon) \\ \vdots \\ \hat{t}_m(\epsilon) \end{pmatrix} = \begin{pmatrix} t_1 e^\epsilon \\ t_2 e^\epsilon \\ \vdots \\ t_m e^\epsilon \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{R}}_0(\epsilon) = \begin{pmatrix} \hat{R}_1(\epsilon) \\ \hat{R}_2(\epsilon) \\ \vdots \\ \hat{R}_m(\epsilon) \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{pmatrix}.$$

Since symmetries preserve solution curves it follows that this latter transformed time series should also be a realisation of another power law curve specified by a parameter  $\hat{A}(\epsilon)$ . In other words, if the fitted power law to the transformed time series is given by  $R(t) = \hat{A}(\epsilon) t^\gamma$  then it is possible to generate  $\hat{\mathbf{t}}_0(\epsilon)$  and  $\hat{\mathbf{R}}_0(\epsilon)$  from this curve. This means that the residuals between the fitted curve  $R(t) = \hat{A}(\epsilon) t^\gamma$  and the transformed time series given by  $\hat{\mathbf{t}}_0(\epsilon)$  and  $\hat{\mathbf{R}}_0(\epsilon)$  should all be zero as the fit is perfect. Consequently, the fit of the model to the transformed data denoted by  $R_{\text{adj}}^2(\epsilon)$  should be one for all transformation parameters  $\epsilon$ . Thus, by transforming a time series using any symmetry where the time series is a realisation of a curve of the same symmetry, then it should be possible to perfectly fit another curve within the same family of curves to the transformed time series. The same exact principles hold for the IM-II, but for this model the parameters  $\alpha$  and  $\tau$  are kept fixed as they determine the family of curves while the parameter  $A$  is estimated as it determines the specific curve.

The numerical validation of the  $t$ -directional symmetries of the PLM and the IM-II show that they are indeed symmetries. The fit of the models to the transformed solutions is  $R_{\text{adj}}^2(\epsilon) \approx 1$  (Fig 5) in both cases for a range of transformation parameters  $\epsilon$ . This illustrates that the scaling symmetry  $\Gamma_{1,1}$  in the case of the PLM and the symmetry  $\Gamma_2$  in case of the IM-II maps solution curves to other solution curves which is the defining property of a symmetry.

## 6 Although two models have very good fits to the same time series data it is essential to perform statistical tests

In addition to the fitting of the models to the Myeloma data, we also fitted the models to colon cancer data collected from [1, 5]. This data was processed in the same way as the Myeloma data with the modification that all ages below 12 years were removed in this data

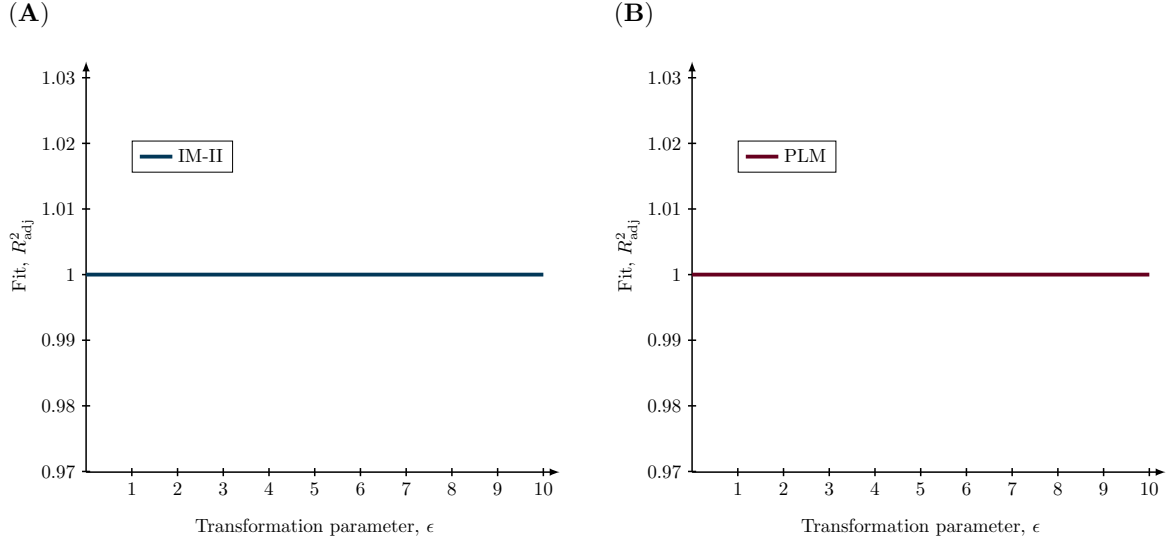


Figure 5: *Numerical validation of the symmetries.* The validation of the  $t$ -directional symmetries of the PLM and the IM-II is done in two steps. Firstly, each curve is transformed with a transformation parameter  $\epsilon$ . Secondly, the fit of the model to the transformed data denoted by  $R_{\text{adj}}^2(\epsilon)$  is reported. In both cases, it is clear that  $R_{\text{adj}}^2 \approx 1$  for all tested value of the transformation parameter  $\epsilon$ . The two tested symmetries are (A)  $\Gamma_2$  for the IM-II and (B)  $\Gamma_{1,1}$  for the PLM. The family of curves which these symmetries act on are defined by the parameters (A)  $(\tau, \alpha) = (67.1564, 0.044)$  in the case of the IM-II and (B)  $\gamma = 5.5600$  in the case of the PLM. The fit have been tested for 100 transformation parameters in the range  $\epsilon \in [0, 10]$ .

set as oppose to removing all ages below 25 years in the Myeloma data set. At first, the model selection problem with the colon cancer data looks similar to the one with Myeloma cancer but reversed, as the fits of both models are really good but the PLM has a slightly better fit with  $R_{\text{adj}} = 0.9953$  than the IM-II with  $R_{\text{adj}} = 0.9921$  (Fig 6). However, when we performed Vuong's non-nested likelihood ratio test [6] on this time series, it turns out that these models are in fact distinguishable where the PLM fits the data better than the IM-II (Tab 3). Also, we performed a variance test in  $R$  (Tab 4) where the conclusion is that the two models are indistinguishable in both the case of the Myeloma cancer *and* the case of the colon cancer data. Thus, the variance test suggests that both are indistinguishable in contrast to Vuong's non-nested likelihood ratio test. Based on these two test we decided to work with the Myeloma data as both statistical tests suggest that the models are indistinguishable.

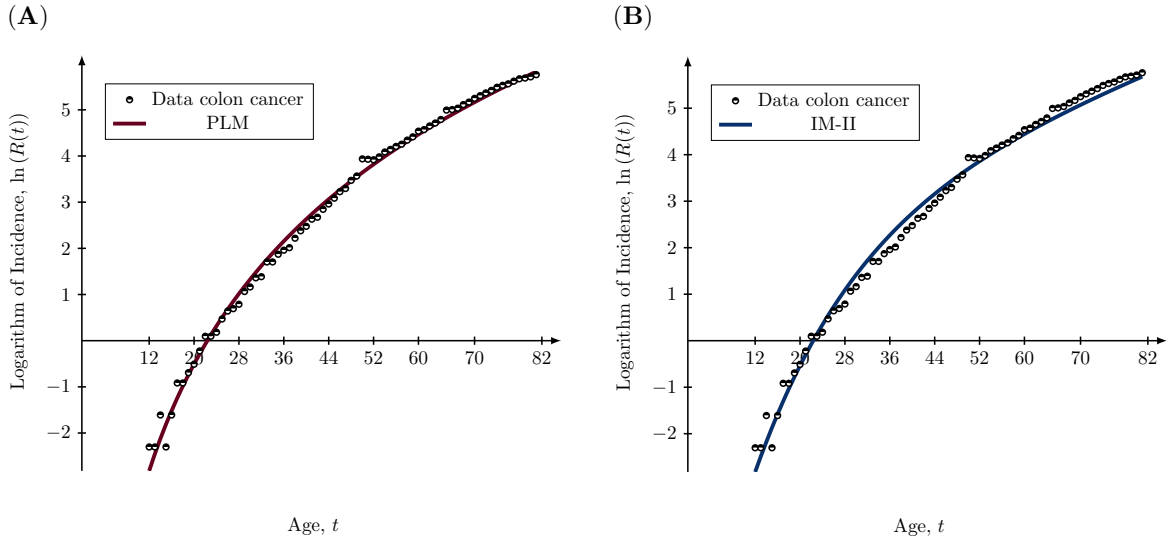


Figure 6: *Both the PLM and the IM-II fit colon cancer data well.* Two models are fitted to the processed time series consisting of the logarithms of the incidences of colon cancer for ages in the range (12, 83) years which is illustrated in two cases. **(A)** The PLM with optimal parameters  $(\ln(A), \gamma) = (-14.0715, 4.5280)$  and fit  $R_{\text{adj}} = 0.9953$ . **(B)** The IM-II with optimal parameters  $(\ln(A), \alpha, \tau) = (4.8731, 0.044, 58.3782)$  and fit  $R_{\text{adj}} = 0.9921$ .

Table 3: *The fit of the two candidate models to the colon cancer data.* The columns from left to right are: the name of the model, the function describing the solution curve, the optimal fit  $R_{\text{adj}}^2$  and the optimal parameters. In the symmetry based methodology for model selection, it is the parameter  $A$  that is fitted to the transformed time series for both the PLM and the IM-II. The parameters  $\gamma$  in the former case and the parameter  $\tau$  in the latter case are fixed to their optimal values displayed in this table. The parameter  $\alpha$  in the IM-II is held constant meaning that it is not estimated and its value is set to  $\alpha = 0.044$  [1]. The IM-II contains a “double exponential” for which we have chosen the notation “ $\exp(e^x)$ ”. To statistically test whether the fits of the two models are equal, we have performed Vuong’s non-nested likelihood ratio test [6] where a  $p$ -value that is larger than 0.05 means that the null hypothesis  $H_0$  is likely to be true.

<i>Model</i>	<i>Curve</i>	<i>Optimal fit, <math>R_{\text{adj}}^2</math></i>	<i>Optimal parameters</i>	<i>Non-nested likelihood ratio test</i>  $H_0$ : Model fits are equal
Power law model (PLM)	$R(t) = At^\gamma$	0.9953	$\ln(A) \approx -14.0715$ $\gamma \approx 4.5280$	$H_1$ : the PLM fits better than the IM-II $p = 2.19 \cdot 10^{-5}$
Immunological model (IM-II)	$R(t) = \frac{A}{\exp(e^{-\alpha(t-\tau)}) - 1}$	0.9921	$\ln(A) \approx 4.8731$ $\tau \approx 58.3782$ $\alpha = 0.044$ (Fixed)	$H_1$ : the IM-II fits better than the PLM $p = 1$

Table 4: *The variance test indicate the both models are indistinguishable when fitted to two different data sets. A value of  $p < 0.05$  indicates that the null hypothesis  $H_0$  is likely to be true while a value of  $p > 0.05$  indicates that the alternative hypothesis  $H_1$  is true.*

<i>Data set</i>	<i>Statistical hypotheses</i> $H_0$ : the PLM and the IM-II are indistinguishable $H_1$ : the PLM and the IM-II are distinguishable
Myeloma cancer	$p = 0.00718$
Colon cancer	$p < 2 \cdot 10^{-16}$

## References

- [1] S. Palmer, L. Albergante, C. C. Blackburn, and T. J. Newman, “Thymic involution and rising disease incidence with age,” *Proceedings of the National Academy of Sciences*, vol. 115, pp. 1883–1888, 2018.
- [2] F. Ohlsson, J. Borgqvist, and M. Cvijovic, “Symmetry structures in dynamic models of biochemical systems,” *Journal of the Royal Society Interface*, vol. 17, p. 20200204, 2020.
- [3] P. E. Hydon and P. E. Hydon, *Symmetry methods for differential equations: a beginner’s guide*, vol. 22. Cambridge University Press, 2000.
- [4] B. Johannes, “The construction, analysis and validation of mechanistic mathematical models of protein kinetics in the context of replicative ageing in budding yeast,” 2020. URI:<http://hdl.handle.net/2077/64055>.
- [5] “Surveillance, epidemiology, and end results (seer) program. seer\*stat database: Incidence—seer 18 regs research data (2000–2016) (national cancer institute, dccps, surveillance research program, surveillance systems branch).” Accessed November 2, 2015. Available at <https://seer.cancer.gov/>.
- [6] Q. H. Vuong, “Likelihood ratio tests for model selection and non-nested hypotheses,” *Econometrica: Journal of the Econometric Society*, pp. 307–333, 1989.