# Structural Identifiability for Mathematical Pharmacology: Models of myelosuppression

Neil D Evans, S.Y. Amy Cheung and James W.T. Yates

*School of Engineering, University of Warwick, Coventry, CV4 7AL UK*

*Innovative Medicines and Early Development, AstraZeneca, Cambridge, UK*

Structural identifiability is an often overlooked, but essential, prerequisite to the experiment design stage. The application of structural identifiability analysis to models of myelosuppression is used to demonstrate the importance of its considerations. It is shown that, under certain assumptions, these models are structural identifiable and so drug and system specific parameters can truly be separated. Further it is shown via a meta-analysis of the literature that because of this the reported system parameter estimates for the "Friberg" or "Uppsala" model are consistent in the literature.

## Introduction

The promises of model informed drug discovery and development [1] are the reduced time lines and number of experiments required. This is because mathematical models can incorporate current learning about a compound or class of compounds. This can include previous learning about disease state, progression and drug induced toxicities. Learning is usually encapsulated in the form of model structures and parameters identified from previous modelling studies. Thus models can be used to integrate information from many sources. Predictions generated from these models inform the path of drug discovery and development because they focus on what is known and not known.

This is especially so for systems pharmacology models whereby different components of the model represent specific discrete aspects of pharmacokinetic-pharmacodynamic (PKPD) relationships and therefore separate out "drug specific" and "system specific" parameters. This mapping of parameters and their meaning allows learning from diverse sources to be integrated and predictions about an untested scenario to be made. For example, the pharmacodynamic profile of a compound previously untested in a particular disease can be generated by bringing together pharmacokinetic and potency information together with pharmacodynamic information for another compound that has been tested in that disease: pharmacokinetics and potency are "drug specific" while the disease aspects are "disease specific". Thus the aim of systems approaches is to learn and translate to new situations.

To do this with confidence we need to ensure that we are separating out drug versus system parameters. Otherwise the approach outlined in the preceding paragraph will fail. If a model is not identifiable: i.e. multiple parameter values describe the data, and are therefore dependent upon the set of parameter values chosen, very different predictions may be generated when integrating subsystem models from different sources.

*Structural identifiability* [2] is an often overlooked, but essential, prerequisite to the experiment design and parameter estimation steps early in the validation process for parametric *knowledge-based* models, i.e. where the structure is based on assumptions or knowledge of the real system. The property of identifiability arises in the validation process because experiments for data collection give rise to an input-output structure for the model, which defines how external inputs (perturbations)

arise in the model and what functions of the model variables can be directly measured. Structural identifiability considers the uniqueness of the unknown parameters with respect to this input-output structure, and is fundamental since estimates for unidentifiable parameters are effectively meaningless. Moreover, the presence of unidentifiable parameters can result in errors in predictions or inferences made from the model. Structural identifiability is widely applied in many application areas, such as pharmacokinetics, pharmacodynamics and disease modelling, and can increase the confidence of *in silico* modelling approaches to better utilise experimental data.

A lack of structural identifiability can (depending on the quality of numerical techniques and their implementations employed) manifest itself in a lack of sensitivity of model outputs to the parameters (and thus lack of sensitivity of fits). But similar insensitivity can be manifest from the choice of sampling times, measurement errors and model/system uncertainties, and this is sometimes referred to in the literature as *practical identifiability*.

It must be remembered that identifiability is with respect to the input-output behaviour of the model, which relates to what can be perturbed and measured. In addition it is generally not a trivial matter to ascertain the effect of parameters on the output except via a formal structural identifiability analysis. Sometimes intuition about identifiable parameters is proved correct, but other times almost surprising results arise. One of the motivations for this paper is to show that intuition should not replace a formal mathematical analysis that is a precursor to aspects of parameter estimation accuracy.

Numerous techniques exist for performing a structural identifiability analysis on linear (see [3] and the references therein) and nonlinear continuous-time parametric models (for example, the Taylor series approach [4], similarity transformation based approaches [5, 6], computational and differential algebra techniques [7, 8] and methods based on polynomial realisation [9]). Significant computational problems can arise in *structural* identifiability analyses because of the symbolic nature of a genuine structural analysis, even for relatively simple models. As a result it can often be extremely difficult to perform a structural identifiability analysis, and typically in order to overcome these problems some authors resort to a numerical analysis using a suitable sensitivity matrix. However, such an analysis is heavily dependent on notional values for the parameters (that are to be estimated), and involves applying a sampling rate to the output. The results are therefore affected by a number of factors that one would wish to understand the individual effect of – for example, is a model over-parameterised regardless of the number and timing of samples taken?

There exist a number of computational tools for performing a structural identifiability analysis (see [10] for a comparison of three key ones), one of which is based on the *Exact Arithmetic Rank* method and implemented in the mathematical computational system *Mathematica* [11]. The method uses the ideas introduced by Sedoglavic [12] to replace symbolic computations with random integer ones, and thus overcomes the significant computational problems of fully symbolic approaches. However, being based on a rank condition the method "only" returns local identifiability results and does not characterise the identifiability of combinations of unidentifiable parameters.

An important application of mathematical modelling is to the prediction of drug toxicity [13]. If dose and regimen specific toxicity can be predicted from non-clinical data before going into patients, or from current patient data in Phase 1 and Phase 2then, coupled with models of the effectiveness of the drug, the therapeutic index (TI) can be rapidly optimised *in silico*. But only, as we have argued above, if the various parts of the system have been well characterised.

For treatments of cancer, one important dose limiting toxicity (DLT) is myelosuppression. In this article, we demonstrate the application of structural identifiability analysis to a range of myelosuppression

models. For one of these models evidence has been published that the model can predict the potency of the drug against human bone marrow using data in the rat and *in vitro* [14]. This is suggestive that the model is correctly separating out drug-specific and system (species) specific parameters. We demonstrate that under specific scenarios the parameters in this model are uniquely identifiable. A meta-analysis of application of this model demonstrates that the system specific parameters estimated in patients are consistent across chemotherapies and patients with varying types of solid malignancies. Some modifications of this model have been proposed that provide improvements under specific circumstances. We investigate the structural identifiability of these models as well.

## Methods

### Mathematical models of myelosuppression

The first model, often referred to as the "Friberg" or "Uppsala" model [15] describes the time and drug exposure dependence of neutropenia in patients. It was first developed based upon data generated in patients after dosing with various chemotherapies. A central claim of the research was that parameters were consistent across data sets. In its most general form it is a five state model with separate rate constants for proliferation, maturation and lifespan in the periphery:

$$\begin{aligned}
\frac{\mathrm{d}\,Prol}{\mathrm{d}t} &= k_{prol} Prol\left(1-E_{drug}\right)\left(\frac{Circ_0}{Circ}\right)^{\gamma} - k_{tr} Prol \\
\frac{\mathrm{d}\,Transit_1}{\mathrm{d}t} &= k_{tr} Prol - k_{tr} Transit_1 \\
\frac{\mathrm{d}\,Transit_2}{\mathrm{d}t} &= k_{tr} Transit_1 - k_{tr} Transit_2 \\
\frac{\mathrm{d}\,Transit_3}{\mathrm{d}t} &= k_{tr} Transit_2 - k_{tr} Transit_3 \\
\frac{\mathrm{d}\,Circ}{\mathrm{d}t} &= k_{tr} Transit_3 - k_{circ} Circ
\end{aligned} \tag{1}$$

where the rate constant $k_{tr}$ is derived from the mean transit time (MTT):

$$k_{tr} = \frac{4}{MTT},$$

and $E_{drug}$ is usually a linear function of drug concentration.

Despite its success in terms of describing and predicting the regimen dependence of neutropenia (and thrombocytopenia) and also providing a translational bridge between the clinical and non-clinical studies [14], the model does have deficiencies. Specifically there are issues when considering combinations of treatments with acute dosing or the long term effects of chemotherapy. Since then several modifications have been proposed to overcome these challenges.

The first extension, by Bender et al [19], describes an apparent downward drift of the baseline by having a concentration dependent reduction in the baseline over time. This is the addition of a second process that operates on a time scale longer than that described by *MTT*. The first equation is subsequently modified as:

$$\frac{\mathrm{d}\,Prol}{\mathrm{d}t} = k_{prol} Prol\left(1-E_{drug}\right)\left(\frac{Circ_0(t)}{Circ}\right)^{\gamma} - k_{tr} Prol$$
$$Circ_0(t) = \left(Circ_0 - Circ_1\right)e^{-K_{deplete}C_{avg}t} + Circ_1 \qquad (2)$$

Therefore this model has two additional parameters to estimate: namely $Circ_0$ and $k_{deplete}$.

In Mangas-Sanjuan et al [23] the model is adapted to explain the schedule dependence of the effects of diflomotecan on neutrophil counts. To attain this the *Prol* compartment was divided up into 3 separate compartments representing cycling and non-cycling cells in the bone marrow. Therefore d*Prol*/d*t* is replaced with:

$$\frac{\mathrm{d}\,Prol}{\mathrm{d}t} = k_{prol} Prol\left(1-E_{drug}\right)\left(\frac{Circ_0}{Circ}\right)^{\gamma} - k_{tr} F_{prol} Prol - k_{cycle}\left(1-F_{prol}\right)Prol + k_{cycle} Qc_2$$
$$\frac{\mathrm{d}\,Qc_1}{\mathrm{d}t} = k_{cycle}\left(1-F_{prol}\right)Prol - k_{cycle} Qc_1 \qquad (3)$$
$$\frac{\mathrm{d}\,Qc_2}{\mathrm{d}t} = k_{cycle} Qc_1 - k_{cycle} Qc_2$$

In Quartino et al [24], the role of granulocyte-colony stimulating factor (GCSF) in the homeostasis of peripheral neutrophil counts is considered explicitly. In the model peripheral neutrophils utilise GCSF ($k_{ANC}$) and so provide a clearance route. GCSF will accumulate when absolute neutrophil count (ANC) reduces and consequently stimulate proliferation of progenitor cells. We consider the following case:

$$\frac{\mathrm{d}\,Prol}{\mathrm{d}t} = k_{prol} Prol\left(1-E_{drug}\right)\left(\frac{GCSF}{GCSF_0}\right)^{\gamma} - k'_{tr} Prol$$
$$\frac{\mathrm{d}\,Transit_1}{\mathrm{d}t} = k'_{tr} Prol - k'_{tr} Transit_1$$
$$\frac{\mathrm{d}\,Transit_2}{\mathrm{d}t} = k'_{tr} Transit_1 - k'_{tr} Transit_2$$
$$\frac{\mathrm{d}\,Transit_3}{\mathrm{d}t} = k'_{tr} Transit_2 - k'_{tr} Transit_3 \qquad (4)$$
$$\frac{\mathrm{d}\,Circ}{\mathrm{d}t} = k'_{tr} Transit_3 - k_{circ} Circ$$
$$\frac{\mathrm{d}\,GCSF}{\mathrm{d}t} = k_{in} - \left(k_e + k_{ANC} Circ\right)GCSF$$
$$k'_{tr} = k_{tr}\left(\frac{GCSF}{GCSF_0}\right)^{\beta}$$

Figure 1 displays schematic representations of these models. In each of the above models it is generally assumed that the underlying system (represented by the model) is in equilibrium (or at baseline) prior to treatment. This requirement is typically built into the models by ensuring that the constant baseline steady state exists and that the model variables start at these baseline levels.

For the Friberg [15] and Bender et al [19] models the baseline considerations result in the assumption of $k_{prol} = k_{circ} = k_{tr}$, resulting in a baseline initial condition of $Circ_0$ for all compartments. For the Mangas-Sanjuan et al [23] extension (as discussed in their paper) the assumption that $k_{prol} = F_{prol}\, k_{circ} = F_{prol}\, k_{tr}$ is

made, resulting in a baseline initial condition of $Circ_0$ for all of the transit compartments and *Circ*. The remaining initial conditions are:

$$Prol(0) = \frac{Circ_0}{F_{prol}}, Qc_1(0) = Qc_2(0) = \left(1 - F_{prol}\right)\frac{Circ_0}{F_{prol}} \tag{5}$$

In the final extension, by Quartino et al [24], it is similarly assumed that $k_{prol} = k_{circ} = k_{tr}$, and

$$k_{in} = \left(k_e + k_{ANC} Circ_0\right) GCSF_0 \tag{6}$$

with a baseline initial condition of $Circ_0$ for all compartments, except the *GCSF* one that has a baseline of $GCSF_0$.

### Structural Identifiability

For each of the models identifiability was analysed for the control (drug-free) and full models separately. The control case is only practical in situations where the underlying system is not in steady state at a constant baseline – since in the latter case the only identifiable parameter would be $Circ_0$ – and is included here for completeness. Similarly, the most general analysis for the full model is included in which the baseline considerations are not imposed.

Two parameter vectors are said to be *indistinguishable* with respect to a given experiment, if they give rise to identical outputs (the measured variables) for all admissible inputs, that is, they have identical input-output structures. Let $S(p)$ denote the set of all allowable parameter vectors that are indistinguishable from $p$. A model is defined to be *structurally globally identifiable* (**SGI**) if for generic $p$ the set $S(p) = \{p\}$. If, for generic $p$, $S(p) \neq \{p\}$ and is countable then the model is *structurally locally identifiable* (**SLI**). If $S(p)$ is uncountable for generic $p$ then the model is *structurally unidentifiable* (**SU**).

With respect to the previous definitions, an individual parameter $p_i$, is *globally identifiable* if, for generic $p$, all vectors $q$ in $S(p)$ satisfy $q_i = p_i$. It is *locally identifiable* if there are only a countable number of distinct values (including $p_i$) that any such $q_i$ can take. Otherwise the parameter $p_i$ is said to be *unidentifiable*, and there are an uncountable number of vectors $q$ in the set $S(p)$ such that $q_i \neq p_i$ and all of the $q_i$ are different.

The approach taken is based on that presented in [25], which is based on consideration of the input-output equation for each model. Based on the computational algebraic methods of Forsman [26] an input-output equation of the following form is generated:

$$A(y,u,p)y^{(n)} + B(y,u,p) = 0 \tag{5}$$

where $A$ and $B$ are multivariate polynomials in the input $u$, and its derivatives, and the output $y$, and its derivatives up to order $n - 1$, where $n$ is the number of variables in the model. The coefficients of these polynomials are rational functions of the parameters in the parameter vector $p$. Therefore, any parameter vector $\bar{p}$ that gives rise to the same output as $p$ for every admissible input $u$ satisfies the following equation:

$$A(y,u,p)B(y,u,\bar{p}) + B(y,u,p)A(y,u,\bar{p}) = 0\,. \tag{6}$$

Since this is a differential polynomial of order less than $n$ in $y$ we see that the monomial terms (products of $u$ and $y$, and their derivatives) are linearly independent since the initial conditions (for the state variables) are generic (parameters). Therefore identifiability is determined by equating the coefficients of the monomials, which are rational functions of $p$ and $\bar{p}$, to zero.

When the initial conditions for the model are not general but rather specific ones then care has to be taken with the above approach. Saccomani *et al.* [7] showed that when the initial conditions are specified it is possible that the output $y(t)$ can evolve in such a way as to invalidate the assumption of linearly independent monomials in (6). However, if the system is *accessible* [36] with respect to the input $u$ then the assumption is valid.

Once the identifiability of the model is assessed from the input-output equation (5) we conclude the analysis by assessing the implication that the initial conditions for $y$ and its derivatives up to order $n-1$, which are also rational functions of $p$ (including the general initial conditions for the model variables), must also be unique for the experimental conditions.

The drug concentration in the central compartment is treated as an input to the models considered in this paper in order to reduce the computational burden of determining the input-output equation (5). Since we wish to assess the identifiability of the combined pharmacodynamic and response models we assume that the pharmacokinetic model is known and represented by a controllable compartment model. This ensures that, at least theoretically, all input profiles can be achieved.

A challenge to the identifiability approach results from the unknown exponent $\gamma$; since we wish to determine identifiability of this parameter, which appears in an essentially non-algebraic way, we define the following "dummy" variables as needed:

Friberg *et al* (2002), Mangas-Sanjuan (2015) models $$z(t) = \left( \frac{Circ_0}{circ(t)} \right)^{\gamma}$$

Bender *et al* (2012) model $$z(t) = \left( \frac{Circ_0(t)}{circ(t)} \right)^{\gamma}$$

Quartino *et al* (2015) model $$z_1(t) = \left( \frac{GCSF(t)}{GCSF_0} \right)^{\gamma}, \; z_2(t) = \left( \frac{GCSF(t)}{GCSF_0} \right)^{\beta}.$$

For the latter model the feedback of $GCSF(t)$ to affect the transit rate coefficients leads to computational problems when $\beta \neq 0$. This illustrates the computational problems associated with fully symbolic approaches, and to overcome this problem we apply the *Mathematica* package **IdentifiabilityAnalysis** [11] for performing an identifiability analysis based on the *Exact Arithmetic Rank* (EAR) method.

The advantage of the approach in [25] is that when a model proves to be unidentifiable the approach provides globally identifiable combinations of the unidentifiable parameters. The disadvantage of the approach lies in it being entirely symbolic and requiring a Gröbner basis construction. **IdentifiabilityAnalysis** has the advantage of being computationally more feasible to apply, even for large scale systems (see the comparison paper by Raue *et al* [10]), but only provides the unidentifiable parameters and not any globally identifiable combinations of them.

### Meta-analysis of reported parameter values

A search was performed to find papers that had used the Friberg model to analyse longitudinal peripheral neutrophil counts. Population average (theta) and inter-individual variability (omega) were extracted from each publication. Publications were identified as those that cited the first publication of the Friberg paper [15] and reported a nonlinear mixed effect analysis of blood neutrophil count changes after anti-cancer treatments.

The consistency of the estimates was visualised using a Galbraith plot [27]. In this plot the reported parameter estimate divided by the reported standard error (SE) is plotted on the ordinate versus 1/SE. This plot served two purposes: Firstly, if estimates are consistent given their precision, points will lie on a straight line. Secondly, the estimates with greater precision will aggregate away from the origin and so the slope of the regression will be the mean of the parameter estimates weighted by their precision. Thus this process was applied to the population mean and inter-individual variability estimates from the identified modelling reports.

Where percent SE was reported, SE was calculated as SE=%SE × estimate, and where 95% confidence intervals were reported, SE was calculated assuming a large sample size and dividing the range of the confidence interval by 1.96.

## Results

### Structural identifiability analysis

| Model | Model with PKPD, no constraints | Model with PKPD, constrained for equilibrium | |
|---|---|---|---|
| | General initial conditions (ICs) | General ICs | Baseline ICs |
| **Friberg *et al* (2002) [15]** | **SU:** $k_{prol}$ $Circ_0$ | **SGI** | **SGI** |
| **Bender et al (2012) [19]** | **SU:** $k_{prol}$ $Circ_0$, $k_{prol}/Circ_1$ | **SU:** $Circ_0$ $Circ_1$ | **SGI** |
| **Mangas-Sanjuan et al (2015) [23]** | **SU:** $k_{prol}$ $Circ_0$ | **SGI** | **SGI** |
| **Quartino et al (2015) [24] (β = 0)** | **SU:** $k_{in}/GCSF_0$ | **SU:** $k_{in}/GCSF_0$ | **SU:** (only $GCSF_0$ unidentifiable) |
| **Quartino et al (2015) [24] (Full Model; EAR approach[a])** | **SU[a]:** $k_{in}/GCSF_0$ | **SU[a]:** $k_{in}/GCSF_0$ | **SU[a]:** (only $GCSF_0$ unidentifiable) |

**Table 1:** Key identifiability results for all models. All model parameters (including initial conditions) are *globally identifiable* unless a parameter combination is provided; in this case the individual parameters are *unidentifiable* but the combination given is *globally identifiable*. [a]EAR approach can only guarantee local results, so combinations here are *at least* locally identifiable.

A summary of the identifiability results is presented in Table 1. The accessibility of each model was tested to verify that specified initial conditions would not affect the results of the identifiability analyses.

#### *Friberg et al (2002)*

In the drug-free case $c_P(t) = 0$ for all $t$. The vector of unknown parameters is given by the following:

$$\boldsymbol{p} = \left(k_{prol}\ k_{tr}\ k_{circ}\ \gamma\ Prol(0)\ Transit_1(0) \ldots Transit_3(0)\ Circ(0)\ z(0)\right)^{\mathrm{T}},$$

where the baseline parameter, $circ_0$, is identifiable if and only if $z(0)$ is. Considering two cases:

- The model is **structurally unidentifiable** (**SU**). All parameters and initial conditions are *globally identifiable* **except** the following (which are unidentifiable): $k_{prol}$ and $circ_0$. The product of these

two parameters is globally identifiable and so if either of these is known *a priori* then the model is **structurally globally identifiable** (**SGI**).

- If we make the simplification that $k_{prol} = k_{tr}$ then the model is **SGI**.

In the presence of drug, and assuming linear PD effect ($E_{drug} = k_E\ c_P(t)$), the vector of unknown parameters is given by the following:

$$\boldsymbol{p} = \left(k_{prol}\ k_{tr}\ k_{circ}\ \gamma\ k_E\ Prol(0)\ Transit_1(0) \ldots Transit_3(0)\ Circ(0)\ z(0)\right)^{\mathrm{T}}.$$

Again considering two cases:

- The general version of the model is **SU**. All parameters and initial conditions are *globally identifiable* **except** the following (which are unidentifiable): $k_{prol}$ and $Circ_0$. The product of these two parameters is globally identifiable and so if either of these is known then the model is **SGI**.
- Baseline initial conditions: If we assume that $k_{prol} = k_{circ} = k_{tr}$ (to enable steady state conditions) then the model is **SGI**. This is also the case when assuming baseline initial conditions.

It is worth noting that being able to run control and drug experiments, and then appealing to a *parallel experiments* approach [28] does not change the identifiability result; from a *structural* (but almost certainly not from a *numerical* or *practical*) perspective one set of experiments is enough for identifiability when $k_{prol} = k_{tr}$.

### *Bender et al (2012)*

The drug-free version of this model is basically the Friberg model and so is **SGI** provided $k_{prol} = k_{tr}$, and **SU** otherwise.

In the presence of drug, and assuming linear PD effect ($E_{drug} = k_E\ c_P(t)$), the vector of unknown parameters is given by the following:

$$\boldsymbol{p} = \left(k_{prol}\ k_{tr}\ k_{circ}\ \gamma\ k_E\ k_{deplete}\ circ_1\ Prol(0) \ldots z(0)\right)^{\mathrm{T}}.$$

Again considering two cases:

- The general version of the model is **SU**. All parameters and initial conditions are *globally identifiable* **except** the following (which are unidentifiable): $k_{prol}$, $Circ_0$ and $Circ_1$. The product of $k_{prol}$ and $Circ_0$ and the ratio of $k_{prol}$ and $Circ_1$ are globally identifiable. Therefore if any of these parameters is known *a priori* then the model is **SGI**.
- Baseline initial conditions: If we assume that $k_{prol} = k_{circ} = k_{tr}$ then the model is still **SU**, with all parameters and initial conditions globally identifiable except $Circ_0$ and $Circ_1$. The product of $Circ_0$ and $Circ_1$ is *globally identifiable*. Assuming baseline initial conditions the model is **SGI**.

### *Mangas-Sanjuan et al (2015)*

In the drug-free case ($c_P(t) = 0$ for all $t$) the vector of unknown parameters is given by the following:

$$\boldsymbol{p} = \left(k_{prol}\ F_{prol}\ k_{tr}\ k_{circ}\ k_{cycle}\ \gamma\ Prol(0) \ldots z(0)\right)^{\mathrm{T}}.$$

Considering two cases:

- The model is **SU**. All parameters and initial conditions are *globally identifiable* **except** the following (which are unidentifiable): $k_{prol}$ and $circ_0$. The product of these two parameters is globally identifiable and so if either of these is known then the model is **SGI**.
- If we assume that $k_{prol} = F_{prol}\ k_{tr}$ then the model is **SGI**.

In the presence of drug, and assuming linear PD effect ($E_{drug} = k_E\ c_P(t)$), the vector of unknown parameters is given by the following:

$$\boldsymbol{p} = \left(k_{prol}\ F_{prol}\ k_{tr}\ k_{circ}\ k_{cycle}\ \gamma\ k_E\ Prol(0) \ldots z(0)\right)^{\mathrm{T}}.$$

We again consider two cases:

- The general version of the model is **SU**. All parameters and initial conditions are *globally identifiable* **except** $k_{prol}$ and $circ_0$, but their product is *globally identifiable* and so if either of these is known then the model is **SGI**.
- Baseline initial conditions: If we assume that $k_{prol} = F_{prol}\ k_{tr} = F_{prol}\ k_{circ}$ then the model is **SGI**. This is also the case when assuming baseline initial conditions.

*Quartino et al (2015)*

Due to computational problems it was only possible to perform a symbolic structural identifiability analysis for the special case with $\beta = 0$. The model was found to be unidentifiable for all cases previously considered. Therefore a further, natural assumption was made that $GCSF(0) = GCSF_0$.

In the drug-free case ($c_P(t) = 0$ for all $t$) the vector of unknown parameters is given by the following:

$$\boldsymbol{p} = \left(k_{prol}\ k_{tr}\ k_{circ}\ k_{in}\ k_e\ k_{ANC}\ \gamma\ \beta\ Prol(0) \ldots z_1(0)\ z_2(0)\right)^{\mathrm{T}}.$$

The model is **SU**. All parameters and initial conditions are at least *locally (globally* when $\beta = 0$) *identifiable* **except** $k_{in}$ and $GCSF_0$, which are *unidentifiable*. For the special case with $\beta = 0$ the ratio of these two parameters is globally identifiable (thus if either is known *a priori* then the model is **SGI**). Making the assumption that $k_{prol} = k_{circ} = k_{tr}$, does not affect the identifiability of the model until it is also assumed that $k_{in} = (k_e + k_{ANC}\ Circ_0)\ GCSF_0$, and then only $GCSF_0$ is unidentifiable. Thus if this were measured or otherwise known then the model would be **SGI**.

In the presence of drug, and assuming linear PD effect ($E_{drug} = k_{Effect}\ c_P(t)$), the vector of unknown parameters is given by the following:

$$\boldsymbol{p} = \left(k_{prol}\ k_{tr}\ k_{circ}\ k_{in}\ k_e\ k_{ANC}\ k_{Effect}\ \gamma\ \beta\ Prol(0) \ldots z_1(0)\ z_2(0)\right)^{\mathrm{T}}.$$

The same identifiability result applies in that the model is **SU** with all parameters and initial conditions at least *locally identifiable* **except** $k_{in}$ and $GCSF_0$, which are unidentifiable. For the special case with $\beta = 0$ the ratio of these two parameters is globally identifiable (thus if either is known *a priori* then the model is **SGI**). Making the assumption that $k_{prol} = k_{circ} = k_{tr}$, does not affect the identifiability of the model until it is also assumed that $k_{in} = (k_e + k_{ANC}\ Circ_0)\ GCSF_0$, and then only $GCSF_0$ is unidentifiable. From a structural identifiability analysis perspective this is an unusual result in that a single parameter is unidentifiable and there are no identifiable parameter combinations that involve it. If *GCSF* were measured (actually only measurement of the initial condition is necessary) or $GCSF_0$ otherwise known then the model would be **SGI**.

For the cases when $\beta \neq 0$, by using the identifiability of all parameters except $k_{in}$ and $GCSF_0$, consideration of the initial values of the measured variable $Circ(t)$ and its derivatives shows that the ratio $k_{in}/GCSF_0$ is at least locally identifiable. In fact, this suggests a straightforward reparameterisation of the model that is *at least* structurally locally identifiable, namely that we define $GCSF^* = GCSF/GCSF_0$ and replace $k_{in}$ by $k_{in}' = k_{in}/GCSF_0$. In this case intuition matches the results of the formal identifiability analysis.

### Meta-analysis of population PKPD parameters for Friberg *et al* model

Estimates for 18 data sets (9 publications reporting estimates for 11 different chemotherapies) were identified (Table 2). All publications reported estimates of inter-individual variability (IIV) on *MTT* and $Circ_0$ however only two reported IIV on γ and so this was not evaluated. Galbraith plots for each parameter (Figure 2) exhibit strong linear relationships demonstrating concordance across the published parameters. The resulting meta-analysis parameter values are shown in Table 3. Whilst one might expect baseline circulating neutrophils to be relatively similar, what is surprising is the consistency of the estimates of MTT. These drugs have very different molecular targets and so it would appear this model is inferring truly drug independent parameters.

## Conclusions

Structural identifiability is a model-based, data independent, concept that allows researchers to ascertain *a priori* whether a unique set of parameter estimates are to be expected. This assures a one-to-one mapping of biological behaviour onto parameter estimates via experimental observations.

In this paper, we have demonstrated the utility of applying structural identifiability analysis to four models of drug induced myelosuppression with the model of Friberg *et al* as a starting point. The implicit assumption of these models is that the drug is a perturbation on the system and does not directly impact on processes involved in the homeostasis of haematopoiesis. If this holds true these models should allow insight, conditional on the model assumptions, into the biology behind the drug induced toxicity. Any insight will be obscured if there are not unique parameter values that describe the observed behaviour. Furthermore, in this case consistency will not be achieved because different analyses will have chosen different optimal parameter estimates.

Promisingly, all four models are identifiable under certain simplifying assumptions. These simplifications, including $k_{prol} = k_{tr}$ were previously arrived at by considering the steady states of the models and some empirical observations of parameter estimation performance. Here we have demonstrated the impact these assumptions have on structural identifiability. Some issues still exist with current analysis techniques due to the complexity of mathematical expressions relating parameters to the input-output behaviour of the system. Surprisingly it seems that structures distal from the observation such as distinct proliferating and quiescent cells are identifiable. Inspecting this structure it appears to the authors that this structure could be appropriate for modelling the cell cycle specific effects of treatments as well. Explicitly modelling GCSF accumulation as the mechanism of feedback is also identifiable under the assumption that production and degradation of GCSF at steady state are equal.

With this knowledge in hand we have investigated how consistent parameter estimates reported in the literature are. For the Friberg model applied to data from patients receiving chemotherapy we observe consistent parameter values. This is an example of where systems pharmacology can achieve its promise – with a potency estimate derived from nonclinical studies [14] these human system parameters can be used to predict myelosuppression in the clinic. This is particularly useful for planning dose escalation studies for phase 1 studies prior to dosing patients.

There are of course other mathematical models of myelosuppression reported in the literature and consideration of the identifiability of these models should come next. Some of these consider multiple blood cell lineages and may be applicable to parameter estimation from clinical data [29; 30; 31]. Some

of these models are relatively complex to the point that parameter identification is unfeasible. However structural identifiability and its related concept of structural indistinguishability [32] should still be considered. This is because model validation for more complex models is usually simulation based: demonstrating the model describes experimental data. This does not prove the model is a unique explanation and so structural identifiability and indistinguishability analysis would allow the alternative explanations, in terms of model structure and parameter values, to be revealed. We have only considered a close family of models here but have shown that even with relatively complex model structures that are distal from the point of observation, such as separating out proliferating and non-proliferating cells, structural identifiability is assured. We also have not considered the impact of placing a nonlinear mixed effects structure on the model, however early attempts at a methodology for this statistical aspect are encouraging [33].

The symbolic analysis can be computationally intensive. In an attempt to avoid this researchers have published on other analysis approaches, for example [34; 35]. Unfortunately, these are numerical based approaches and so do not provide a general proof of structural identifiability in the way it is reported here.

We have demonstrated that structural identifiability is an important pre-requisite, not just for parameter estimation, but to ensure that system (disease and biology specific) parameters can truly be untangled from drug effect.

## References


[1] Marshall SF, Burghaus R, Cosson V, Cheung SYA, Chenel M, DellaPasqua O, Frey N, Hamrén B, Harnisch L, Ivanow F, Kerbusch T, Lippert J, Milligan PA, Rohou S, Staab A, Steimer JL, Tornøe C, Visser SA. Good Practices in Model-Informed Drug Discovery and Development: Practice, Application, and Documentation.*CPT Pharmacometrics Syst Pharmacol*. **5**:93-122, 2016.

[2] Bellman, R. and Åström, K.J. On structural identifiability. *Math. Biosci.* **7**:329–339, 1970.

[3] Walter, E. *Identifiability of State Space Models*. Springer-Verlag, 1982.

[4] Pohjanpalo, H. System identifiability based on the power series expansion of the solution *Math. Biosci.* **41**:21–33, 1978.

[5] Vajda, S. Godfrey, K.R. and Rabitz, H. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Math. Biosci.* **93**:217–248, 1989.

[6] Evans, N.D., Chapman, M.J., Chappell, M.J. and Godfrey, K.R. Identifiability of uncontrolled nonlinear rational systems. *Automatica* **38**:1799–1805, 2002.

[7] Saccomani. M.P., Audoly, S. and D'Angiò, L. Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica* **39**:619–632, 2003.

[8] Bearup, D.J., Evans, N.D. and Chappell, M.J. The input-output relationship approach to structural identifiability analysis. *Comput. Meth. Prog. Bio.* **109**:171 – 181, 2013.

[9] Němcová, J. Structural identifiability of polynomial and rational systems. *Math. Biosci.* **223**:83–96, 2010.

[10] Raue, A., Karlsson, J., Saccomani, M.P., Jirstrand, M. and Timmer, J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics* **30**:1440 – 1448, 2014.

[11] Karlsson, J., Jirstrand, M. An efficient method for structural identifiability analysis of large dynamic systems. In: *Proceedings of the 16th IFAC Symposium on System Identification*. 2012.

[12] Sedoglavic, A. A probabilistic algorithm to test local algebraic observability in polynomial time. *J. Symbolic Comput.*, **33**: 735–755, 2002.

[13] B. Ribba, N. H. Holford, P. Magni, I. Trocóniz, I. Gueorguieva, P. Girard, C. Sarr, M. Elishmereni, C. Kloft, and L. E. Friberg, "A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis.," *CPT pharmacometrics Syst. Pharmacol.* **3** : e113, 2014.

[14] L. E. Friberg, M. Sandström, and M. O. Karlsson, "Scaling the time-course of myelosuppression from rats to patients with a semi-physiological model," *Invest. New Drugs*, **28**: 744–753, 2010.

[15] L. E. Friberg, A. Henningsson, H. Maas, L. Nguyen, and M. O. Karlsson, "Model of chemotherapy-induced myelosuppression with parameter consistency across drugs," *J. Clin. Oncol.*, **20:** 4713–4721, 2002.

[16] F. Leger, W.J. Loos, R. Bugat, R.H.J. Mathijssen, M. Goffinet, J. Verwij, A. Sparreboom, and E. Chatelut. "Mechanism-based models for topotecan-induced neutropenia" *Clinical Pharmacology and Therapeutics*, **76**: 567-578, 2004

[17] C. Kloft, J. Wallin, A. Henningsson, E. Chatelut, and M.O. Karlsson, "Population Pharmacokinetic-Pharmacodynamic Model for Neutropenia with Patient Subgroup Identification: Comparison across Anticancer Drugs" *Clinical Cancer Research*, **12**: 5481-5490, 2006

[18] K.Ozawa, H. Minami, and H. Sato. "Population pharmacokinetic and pharmacodynamic analysis for time courses of docetaxel-induced neutropenia in Japanese cancer patients" *Cancer Science,* **98**:1985-1992, 2007

[19] B. C. Bender, F. Schaedeli-Stark, R. Koch, A. Joshi, Y. W. Chu, H. Rugo, I. E. Krop, S. Girish, L. E. Friberg, and M. Gupta, "A population pharmacokinetic/pharmacodynamic model of thrombocytopenia characterizing the effect of trastuzumab emtansine (T-DM1) on platelet counts in patients with HER2-positive metastatic breast cancer," *Cancer Chemotherapy and Pharmacology*, **70:** pp. 591–601, 2012.

[20] E. Soto, A. Staab, M. Freiwald, G. Munzert, H. Fritsch, C. Doge, I.F. Troconiz, "Prediction of Neutropenia-Related Effects of a New Combination Therapy With the Anticancer Drugs BI 2536 (a Plk1 Inhibitor) and Pemetrexed", *CPT*, **88**:660-667, 2010

[21] M. Joerger, A.D.R. Huitema, D.J. Richel, C. Dittrich, N. Pavlidis, E. Briasoulis *et al*. "Population Pharmacokinetics and Pharmacodynamics of Paclitaxel and Carboplatin in Ovarian Cancer Patients: A Study by the European Organization for Research and Treatment of Cancer-Pharmacology and Molecular Mechanisms Group and New Drug Development Group", *Clinical Cancer Research*, **13**:6410-6418, 2007

[22] A. Schmitt, L. Gladieff, C.M. Laffont, A. Evrard, J.C. Boyer, A. Lansiaux *et al.* "Factors for Hematopoietic Toxicity of Carboplatin: Refining the Targeting of Carboplatin Systemic Exposure". *Journal of Clinical Oncology*., **28**: 4568-4574, 2010.

[23] V. Mangas-Sanjuan, N. Buil-Bruna, M. J. Garrido, E. Soto, and I. F. Troconiz, "Semimechanistic Cell-Cycle Type-Based Pharmacokinetic/Pharmacodynamic Model of Chemotherapy-Induced Neutropenic Effects of Diflomotecan under Different Dosing Schedules," *J. Pharmacol. Exp. Ther.*, **354:**55–64, 2015.

[24] A. L. Quartino, M. O. Karlsson, H. Lindman, and E. Lena, "Characterization of Endogenous G-CSF and the Inverse Correlation to Chemotherapy-Induced Neutropenia in Patients with Breast Cancer Using Population Modeling," *Pharm Res* **31**:3390-3403, 2014.

[25] Evans, N.D., Moyse, H., Lowe, D., Briggs, D., Higgins, R., Mitchell, D., Zehnder D. and Chappell, M.J. Structural identifiability of surface binding reactions involving heterogeneous analyte: application to surface plasmon resonance experiments. *Automatica*, **49**:48–57, 2013.

[26] Forsman, K. *Constructive Commutative Algebra in Nonlinear Control Theory*. PhD thesis, Linköping Institute of Technology, 1991. Number 261.

[27] R. Galbraith, "Graphical display of Estimates having differing standard errors," *Technometrics*, **30:** 271–281, 1988.

[28] S. Y. A. Cheung, J. W. T. Yates, and L. Aarons, "The design and analysis of parallel experiments to produce structurally identifiable models," *J. Pharmacokinet. Pharmacodyn.*, **40:**93–100, 2013.

[29] J. C. Panetta, M. N. Kirstein, A. J. Gajjar, G. Nair, M. Fouladi, and C. F. Stewart, "A mechanistic mathematical model of temozolomide myelosuppression in children with high-grade gliomas," *Math. Biosci.*, **186**:29–41, 2003.

[30] L. Pujo-Menjouet, "Blood Cell Dynamics : Half of a Century of Modelling," *Math. Model. Nat. Phenom.*, **10**:182–205, 2015.

[31] M. Craig, A. R. Humphries, F. Nekka, J. Belair, J. Li, M. C. Mackey, "Neutrophil dynamics during concurrent chemotherapy and G-CSF administration: Mathematical modelling guides dose optimisation to minimise neutropenia," *Journal of Theoretical Biology. Vol 385 p. 77-89*

[32] N. D. Evans, M. J. Chappell, M. J. Chapman, and K. R. Godfrey, "Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems," *Automatica*, **40**: 1947–1953, 2004.

[33] D. L. I. Janzén, M. Jirstrand, M. J. Chappell, and N. D. Evans, "Three novel approaches to structural identifiability analysis in mixed-effects models," *Comput. Methods Programs Biomed.*, In press (DOI: 10.1016/j.cmpb.2016.04.024).

[34] V. Shivva, J. Korell, I.G. Tucker, and S.B. Duffull. "An Approach for Identifiability of Population Pharmacokinetic–Pharmacodynamic Models", *CPT: Pharmacometrics and Systems Pharmacology*, **2**: e49, 2013

[35] V.K. Siripuram, D.F.B. Wright, M.L. Barclay, S.B. Duffull, "Deterministic identifiability of population pharmacokinetic and pharmacokinetic–pharmacodynamic models", *Journal of Pharmacokinetics and Pharmacodynamics*, **44**:1-9, 2017

[36] E. D. Sontag, *"Mathematical Control Theory. Deterministic Finite Dimensional Systems"*, Texts in Applied Mathematics 6, Springer-Verlag, New York, 1998

Figure 1 Schematics of models: from top to bottom these are Friberg et al [14], Bender et al [19], Mangas-Sanjuan et al [23] and Quartino et al [24]

Figure 2 Galbraith plots for Friberg et al model parameters

Table 2 List of papers used in meta analysis

| DRUG | Reference (Year) | MTT | SE MTT | IIV MTT | SE IIV MTT | Gamma | SE gamma | IIV gamma | IIV gamma Sigma | Circ0 | Circ0 Sigma | IIV Circ0 | IIV Circ0 Sigma |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Docetaxel | 15 (2002) | 88.7 | 2.2175 | 16 | 3.84 | 0.161 | 0.005957 | | 0 | 5.05 | 0.09595 | 42 | 2.94 |
| Paclitaxel | 15 (2002) | 127 | 2.667 | 18 | 5.4 | 0.23 | 0.00644 | | 0 | 5.2 | 0.1872 | 35 | 3.85 |
| Etoposide | 15 (2002) | 135 | 4.995 | 14 | 3.22 | 0.174 | 0.011484 | | 0 | 5.45 | 0.39785 | 42 | 8.4 |
| DMDC | 15 (2002) | 123 | 1230 | 49 | 11.27 | 0.16 | 0.0208 | | 0 | 5.43 | 0.21177 | 39 | 6.24 |
| CPT-11 | 15 (2002) | 113 | 7.797 | 29 | 11.89 | 0.132 | 0.012936 | | 0 | 5.51 | 0.18734 | 29 | 5.51 |
| Vinflunine | 15 (2002) | 122 | 4.514 | 21 | 4.41 | 0.162 | 0.010854 | | 0 | 4.72 | 0.12744 | 41 | 7.38 |
| Diflometecan | 14 (2010 | 117.84 | 5.892 | 25.9 | 5.439 | 0.113 | 0.00904 | | 0 | 5.37 | 0.10203 | 40 | 7.6 |
| Indusalam | 14 (2010) | 165.6 | 1.656 | 23.5 | 2.115 | 0.152 | 0.00304 | | 0 | | 0 | | 0 |
| Topotecan | 16 (2006) | 133 | 6.122449 | 33 | 6.122449 | 0.119 | 0.003061 | | 0 | 5.11 | 0.183673 | 46 | 8.163265306 |
| Docetaxel | 17 (2006) | 83.9 | 1.0068 | 14 | 1.82 | 0.144 | 0.003744 | 15 | 3.3 | 5.3 | 0.2438 | 37 | 4.07 |
| Paclitaxel | 17 (2006) | 126 | 5.292 | 17 | 7.31 | 0.223 | 0.02899 | | 0 | 5.4 | 0.3888 | 35 | 8.05 |
| Etoposide | 17 (2006) | 140 | 5.04 | 17 | 3.57 | 0.172 | 0.010148 | | 0 | 5.26 | 0.32612 | 41 | 8.2 |
| Topotecan | 17 (2006) | 137 | 10.549 | 35 | 6.65 | 0.101 | 0.011716 | | 0 | 5.02 | 0.18072 | 43 | 8.6 |
| Docetaxel | 18 (2007) | 113 | 4.62 | 33.3 | 5 | 0.196 | 0.013 | | 0 | 5.19 | 1.51 | 104 | 62.9 |
| Pemetrexed | 16 (2004) | 107 | 2.9853 | 10.4 | 3.7648 | 0.1902 | 0.017137 | 38.7 | 20.8593 | 5.19 | 0.283893 | 32.9 | 9.9029 |
| Pemetrexed | 20 (2010) | 87.8 | 6.8484 | | 0 | 0.129 | 0.018202 | | 0 | 6.29 | 0.557923 | 31.9 | 8.5811 |
| Carboplatin | 21 (2007) | 141 | 5.217 | | 0 | 0.26 | 0.0195 | | 0 | | 0 | | 0 |
| Carboplatin | 22 (2010) | 150 | 8.163265 | 20.7 | 4.846939 | 0.146 | 0.010204 | | 0 | 4.54 | 0.255102 | 36.2 | 3.979591837 |

Table 3 Result of parameter meta-analysis

| **Parameter** | **Description** | **Units** | **Population mean** | **IIV** |
|---|---|---|---|---|
| MTT | Mean transit time from precursor cells to circulating | Hours | 109 | 19.0% |
| Gamma | Strength of feedback | - | 0.148 | - |
| $Circ_0$ | Baseline neutrophil count | $10^9$/l | 5.15 | 38.1% |