Biometrical Journal **52** (2010) 61, zzz–zzz / DOI: 10.1002/bimj.200100000

Optimal individualized dosing strategies: A pharmacologic approach to developing dynamic treatment regimens for continuous-valued treatments

Benjamin Rich [1], **Erica E. M. Moodie** * [2], and **David A. Stephens** [3]

[1] Department of Decision Sciences, HEC Montreal, 3000 chemin de la Côte-Sainte-Catherine, Montreal, QC Canada, H3T 2A7

[2] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, QC Canada H3A 1A2

[3] Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, QC, Canada H3A 2K6

Received zzz, revised zzz, accepted zzz

There have been considerable advances in the methodology for estimating dynamic treatment regimens, and for the design of sequential trials which can be used to collect unconfounded data to inform such regimens. However relatively little attention has been paid to how such methodology could be used to advance understanding of optimal treatment strategies in a continuous dose setting, even though it is often the case that considerable patient heterogeneity in drug response along with a narrow therapeutic window may necessitate the tailoring of dosing over time. Such is the case with warfarin, a common oral anticoagulant. We propose novel, realistic simulation models based on pharmacokinetic-pharmacodynamic properties of the drug that can be used to evaluate potentially optimal dosing strategies. Our results suggest that this methodology can lead to a dosing strategy which performs well both within and across populations with different pharmacokinetic characteristics, and may assist in the design of randomized trials by narrowing the list of potential dosing strategies to those which are most promising.

Key words: Adaptive individualized dosing; Anticoagulation therapy; Continuous doses; Dynamic treatment regimens; G-estimation.

words)

Supporting Information for this article is available from the author or on the WWW under http://dx.doi.org/10.1022/bimj.XXXXXXX.

1 Introduction

Finding the right dose of a drug is typically a question of balancing the drug's desired action or *efficacy* with its side effects, safety, and tolerability. Whether or not a standard dose can be recommended for an entire target population depends on the degree of inter- and intra-individual variability in both the systemic exposure resulting from a given dose (pharmacokinetics, PK), and the effects – both good and bad – resulting from a given exposure (pharmacodynamics, PD), as well as the size of the therapeutic window within which the metabolic uptake of the drug is desired to lie. A dynamic treatment regimen (DTR) results when doses need to be individualized, with the adjustments being made in response to an evolving patient profile. In this setting, estimation of an optimal treatment strategy is desirable.

Warfarin is a highly effective anticoagulant that works to decrease the risk of thrombosis (clotting) by depleting the body's active vitamin K. Following a (single) dose of warfarin, anticoagulant effects are

*Corresponding author: e-mail: erica.moodie@mcgill.ca, Phone: +1-514-398-5520, Fax: +1-514-398-4503

typically observed in one day; the duration of the effect of a single dose is two to five days. The impact of warfarin varies considerably between and within individuals, as diet can replenish vitamin K, warfarin interacts with a variety of other medications, and there are known genetic polymorphisms which affect warfarin's metabolism and potency leading to an increased risk of bleeding or thrombosis in some patients (The International Warfarin Pharmacogenetics Consortium, 2009).

Anticoagulation therapy requires careful monitoring of the *international normalized ratio* (INR), a measure of the time it takes for blood to clot, which must be kept in a narrow target range, typically between 2.0 and 3.0 for most indications (Hirsh *et al.*, 2001). Important genetic and environmental factors that influence warfarin PK and PD result in considerable response heterogeneity, such that the appropriate dose to achieve an INR in the target range can vary by more than five-fold between individuals (Michaud *et al.*, 2008). Warfarin is frequently prescribed in the United States and inappropriate dosing is a major cause of emergency hospitalizations resulting from adverse drug events (Wysowski *et al.*, 2007; Budnitz *et al.*, 2011).

Because of the complexity of dosing warfarin (or other coumarin derivatives), several computer algorithms have been developed to assist physicians in making dosing decisions (Cromme *et al.*, 2010; Nielsen *et al.*, 2014, e.g.). For example, Figure 1 shows the INR and dose data from a random sample of three patients, obtained from a London (UK) anticoagulation clinic, for the first six months following warfarin initiation. As the data show, there are multiple adjustments over time, some large and others not, corresponding to departures of the INR from the target range. These data were extracted from the DAWN AC anticoagulation management system (http://www.4s-dawn.com/products/anticoagulation/dawnac/), which provides computer-assisted dosing based on the INR, allowing easy recuperation of INR and dose information. However it has been shown that computer-assisted dosing can fail to recommend a dose or medical staff may choose to alter the recommended dose anywhere from 6-20% of the time (Poller *et al.*, 2008, 2009).

That multiple seemingly largely heuristic algorithms exist, and that many physicians continue to rely on clinical judgement, suggests that the optimal dosing strategy has not yet been found. Furthermore, determining the optimal adaptive treatment strategy often requires very large samples. It is therefore recommended that the advantage of a proposed adaptive strategy – however identified – is ascertained by a confirmatory randomized trial (Lei *et al.*, 2012). In such circumstances, particularly when the PK and PD characteristics of a drug are well understood, it can be advantageous to use simulations to suggest an adaptive strategy to be tested in a confirmatory trial.

A dose adjustment strategy for warfarin is an example of a DTR. Such regimens have received much attention in the statistical literature recently; for example, see Murphy (2003), Robins (2004), Moodie *et al.* (2007), Dawid and Didelez (2010), Chakraborty and Moodie (2013) among others. However, the literature on dynamic regimen methodology has focused almost entirely on binary or discrete treatments, save for a few exceptions (Murphy, 2003; Henderson *et al.*, 2010; Cotton and Heagerty, 2011; Joffe *et al.*, 2012). Applications with continuous doses, including simulation studies, remain sparse. Simulation of data conforming to a particular treatment contrast model is most readily performed using the approach described by Murphy (2003); we provide an illustration in the Supporting Information. However, such simulation protocols are overly simplistic, and are unable to effectively capture the complexities that are encountered in treatment with a drug like warfarin, where the true model for an outcome that captures, for example, the deviations of INR from an optimal value over an extended period of time as a function of covariates is unknown and possibly highly non-linear.

Recently, Fusaro *et al.* (2013) pointed out the usefulness of simulation in designing clinical trials for warfarin, taking advantage of the availability of realistic PK/PD models for INR prediction. In their work, the dose adjustment strategies compared were pre-specified, not estimated from data. Here we propose that optimal DTR estimation can be effectively combined with PK/PD simulation to generate candidate dynamic dosing strategies for warfarin treatment that could subsequently be validated in randomized clinical trials. While the focus in this work is on warfarin, the methods and approach used are general and

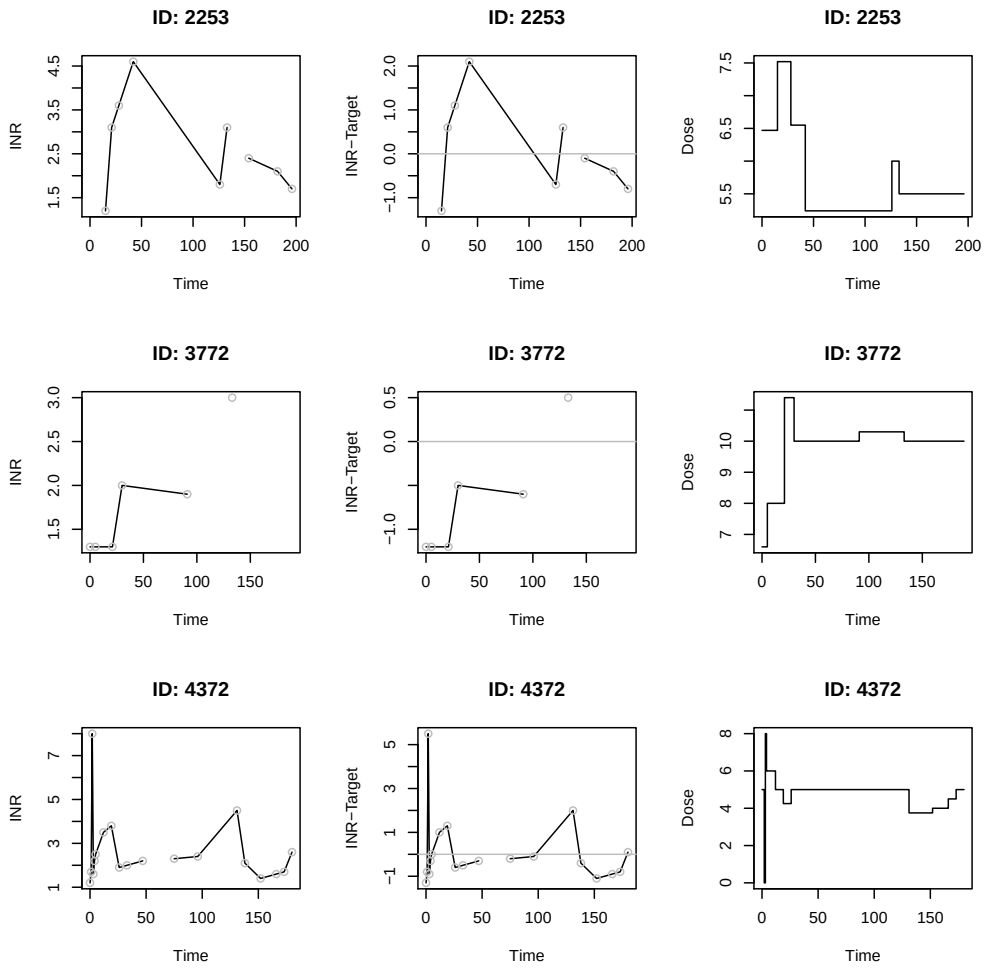


Figure 1 INR (raw and centered by the mid-point of the target INR range), as well as dosing for the first six months following warfarin initiation from a sample of patients in a London anticoagulation clinic.

could be applied to other drugs. Some examples might include thiopurines for the treatment of inflammatory bowel disease (Teml *et al.*, 2007), Theophylline (D'Argenio and Khakmahd, 1983), or other drugs requiring therapeutic monitoring and for which a realistic PK/PD model is known or could be developed.

The paper is structured as follows. In Section 2 we provide a brief review of the framework proposed by Robins (2004) for optimal dynamic treatment regimens using structural nested mean models (SNMMs) and g-estimation; throughout we use the abbreviation SNMM to refer specifically to models for optimal DTRs. We begin with the simpler case of binary treatments before focussing on methodological issues specific to the continuous-scale treatment setting. A key component of the model is the so-called *blip function*, which in the continuous treatment case generally ought to be non-monotone; a convenient choice, and the one we adopt, is a quadratic function. In Section 3, we propose a sophisticated approach to generating realistic data (i.e. mimicking those seen in Figure 1) using a PK/PD model under the plausible setting in which a therapeutic agent's biologic properties are sufficiently well understood. We demonstrate how such

a simulation model could be used to evaluate several dynamic regimens, which could then be tested in clinical practice or in a randomized trial setting. We conclude with a discussion in Section 4.

2 A framework for optimal dynamic regimens with continuous treatments

Robins (2004) gives a more general and complete treatment of optimal DTR estimation with SNMMs, estimated by g-estimation. A number of other approaches have been proposed and compared (Murphy, 2003; Moodie *et al.*, 2007; Rosthøj *et al.*, 2006; Henderson *et al.*, 2010). G-estimation offers several attractive properties, including robustness against some forms of model misspecification (Robins, 2004) and the potential for model checking using residual-based model diagnostics (Rich *et al.*, 2010).

2.1 Notation

We begin by describing the complete vectors of observations for a randomly sampled individual from an implicit underlying population, denoted by O, and suppose that the observed data consist of n i.i.d. such vectors, $O_1, \ldots, O_n$. We use capital letters to denote random variables or vectors, and lower case letters to denote realizations or constants. Thus, $O = (L_0, A_1, L_1, \ldots, A_K, L_K)$ where K is the number of treatment intervals, L_0 denotes a vector of baseline or pre-treatment covariates, $A_j, j = 1, \ldots, K$ denotes the treatment observed at interval j, and $L_j, j = 1, \ldots, K$ denotes a time-varying covariate (possibly multidimensional) associated to interval j. L_j is assumed to be an intermediate response, observed after treatment A_j and prior to treatment A_{j+1}, which it may inform. Indeed, it is assumed that the treatment A_{j+1} is chosen adaptively according to previous treatments and responses, in particular L_j. We define $Y = g(O)$ to be the final outcome of interest, where $g(\cdot)$ is a suitably defined *reward function*, meaning that the larger the value of Y, the better the outcome.

Let $\mathcal{A}_j, j = 1, \ldots, K$ denote the set of possible treatments at interval j, and similarly let $\mathcal{L}_j$ denote the set of the possible values for $L_j, j = 0, \ldots, K$. A dynamic treatment regimen $d = (d_1, \ldots, d_K)$ is a sequence of functions $d_j : \mathcal{H}_j \to \mathcal{A}_j$ where

$$\mathcal{H}_j = \mathcal{L}_0 \times \mathcal{A}_1 \times \mathcal{L}_1 \times \cdots \times \mathcal{A}_{j-1} \times \mathcal{L}_{j-1}$$

that map "histories" to treatments. We assume the existence of *counterfactual* or *potential outcomes* $Y(d)$ under every possible dynamic regimen d. In the DTR context, the potential outcome $Y(d)$ represents the outcome that would have been observed in the event that the treatments had been assigned according to regimen d, rather than the observed regime. An optimal DTR is one that maximizes the value of the counterfactual outcome in expectation. Under regularity conditions (Robins, 2004; Moodie, 2009), the optimal dynamic treatment regimen is uniquely defined and we write

$$d^{\text{opt}} = \arg\max_d E[Y(d)].$$

The outcome Y could be something such as a quality of life score, the ratio of healthy years gained per $1,000 spent on treatment, or, as in our warfarin example, a measure of the deviations from the therapeutic target. It is this optimal dynamic treatment regimen that one wishes to estimate. We now turn to the estimation of an optimal dosing strategy via g-estimation.

2.2 Structural nested mean model and g-estimation for a binary exposure

We consider inference in a semiparametric setting: Consider a dynamic treatment regimen indexed by a history $h_j \in \mathcal{H}_j$ and a treatment $a \in \mathcal{A}_j$ and denoted $d^{h_j,a,\text{opt}}$, defined as follows. For the intervals $m \leq j$, the regimen is static, i.e., fixed *a priori* and the same for all individuals. Note that h_j contains a sequence of treatments for intervals $m < j$, and these are precisely the treatments assigned by $d^{h_j,a,\text{opt}}$ at the corresponding intervals. At interval j the treatment assigned by $d^{h_j,a,\text{opt}}$ is a. For intervals $m > j$, the

regimen is dynamic, and optimal in the sense that the treatments assigned according to $d^{h_j,a,\text{opt}}$ and d^{opt} are identical for the same history.

A SNMM for optimal dynamic treatment regimen estimation is a parametrization of the *blip function*, defined here for the j^{th} interval as

$$\gamma_j(a, h_j) = E[Y(d^{h_j,a,\text{opt}}) - Y(d^{h_j,0,\text{opt}})|H_j = h_j, A_j = a]$$

where $a \in \mathcal{A}_j$ denotes a possible treatment at interval j, $0 \in \mathcal{A}_j$ denotes a reference treatment at interval j, $h_j \in \mathcal{H}_j$ denotes a possible history at interval j, $Y(\cdot)$ denotes the counterfactual outcome under the specified regimen as defined above. Two key features of the blip: (1) the blip function evaluated at $a = 0$ is zero (by definition); (2) the blip function determines the optimal DTR through $d^{\text{opt}}(h_j) = \arg\max_a \gamma_j(a, h_j)$. If a blip function is a function of a and h_j parameterized by a parameter $\psi_j \in \mathbb{R}^{p_j}$, an optimal dynamic treatment regimen structural nested mean model (SNMM) results. In the case of a binary treatment, a common choice would be a linear blip model of the form

$$\gamma_j(a, h_j; \psi_j) = (\psi_j^\top x_j)a$$

for a covariate vector $x_j = x_j(h_j)$. Note that for this model the value of the blip function evaluated at $a = 0$ is zero as required. Since the blip function determines the optimal DTR, the goal is to estimate $\psi = (\psi_1^\top, \ldots, \psi_K^\top)^\top$. This can be achieved by recursive g-estimation (Robins, 2004). In this procedure, we consider

$$G_j(\psi_j, \underline{\psi}_{j+1}) = Y - \gamma_j(A_j, H_j, \psi_j) + \sum_{m=j+1}^{K} \Big[\max_{a' \in \mathcal{A}_m} \{\gamma_m(a', H_m; \psi_m)\} - \gamma_m(A_m, H_m; \psi_m) \Big]$$

where $\underline{\psi}_{j+1} = (\psi_{j+1}^\top, \ldots, \psi_K^\top)^\top$. It then follows (Robins, 2004) that

$$E[G_j(\psi_j, \underline{\psi}_{j+1})|H_j, A_j] = E[Y(d^{H_j,0,\text{opt}})|H_j, A_j].$$

A g-estimating equation at interval j is given by

$$0 = \sum_{i=1}^{n} X_{ij} \big\{ A_{ij} - \pi_j(H_{ij}; \hat{\alpha}_j) \big\} \big\{ G_{ij}(\psi_j, \hat{\underline{\psi}}_{j+1}) - \eta_j(H_{ij}; \hat{\varsigma}_j(\psi_j, \hat{\underline{\psi}}_{j+1})) \big\} \tag{1}$$

where $\pi_j(h_j; \alpha_j)$ is an estimator of $E[A_j|H_j = h_j]$ which equals $\Pr[A_j = 1|H_j = h_j]$ in the binary treatment case, $\eta_j(h_j; \varsigma_j)$ is an estimator of $E[Y(d^{H_j,0,\text{opt}})|H_j] = E[G_j(\psi_j, \underline{\psi}_{j+1})|H_j]$, and $\hat{\varsigma}_j(\psi_j, \underline{\psi}_{j+1})$ is an estimator of ς_j, determined up to a parameter. We refer to $\pi_j(h_j; \alpha_j)$ as the *treatment model*, and to $\eta_j(h_j; \varsigma_j)$ as the *expected counterfactual* (EC) model. The g-estimating equations (1) are solved for ψ_j in reverse-time order, i.e., for $j = K, \ldots, 1$, to obtain the estimator $\hat{\psi}_j$. It can be shown (Robins, 2004) that under standard identifiability assumptions of consistency and exchangeability and regularity assumptions, the resulting estimator is doubly robust, that is, consistent and asymptotically normal under correct specification of either $\pi_j(h_j; \alpha_j)$ or $\eta_j(h_j; \varsigma_j)$. In addition, under regularity conditions, asymptotic standard errors can be estimated by the sandwich estimator. For further details see, for example, Robins (2004), Moodie *et al.* (2007), Moodie (2009), Moodie and Richardson (2010) and Rich *et al.* (2010).

2.3 Structural nested mean model and g-estimation for a continuous dose

With continuous doses, the need to balance efficacy and tolerability requires more complex modelling than is required for binary doses. If a is a continuous variable such as a particular dose of warfarin, then without further constraints, the value of a that maximizes a blip function that is linear in a, i.e., the optimal

treatment, will be either $+\infty$ or $-\infty$, depending on the sign of $\psi_j^\top x_j$. Thus, the choice of model needs to be reconsidered when treatments are on a continuous scale. First, one must consider the sensible finite range over which the treatments can realistically be assigned, and bound the treatments during estimation. If a represents a dose of a drug (in milligrams, say), then a is restricted to a range $[0, A_{\max}]$ where $A_{\max}$ is the highest conceivable dose. Second, if it is believed that in some circumstances the optimal dose does not correspond to either extreme of this acceptable range, but rather to some intermediate value, then the functional form of the blip function should not be linear (nor, indeed, monotone) in a, but should rather allow for a maximum at an interior point of the interval (Moodie and Richardson, 2010). One suitable choice is a quadratic function, namely to assume that the blip function is quadratic in a, and write

$$\gamma_j(a, h_j; \psi_j) = \begin{pmatrix} \psi_j^{(1)\top} & \psi_j^{(2)\top} \end{pmatrix} \begin{pmatrix} a x_j^{(1)} \\ a^2 x_j^{(2)} \end{pmatrix}$$

where the parameter has been partitioned as $\psi_j = (\psi_j^{(1)\top}, \psi_j^{(2)\top})^\top$ and $x_j^{(1)}$ and $x_j^{(2)}$ are two vectors of model covariates, i.e., functions of the history h_j. Robins (2004) considered a similar blip function model in an example, while Murphy (2003) and Rosthøj *et al.* (2006) also suggest the use of quadratic models, though in the context of modeling regret functions rather than blip functions (regret functions and blip functions are related by $\mu_j(a, h_j) = \max_{a'} \gamma_j(a', h_j) - \gamma(a, h_j)$; see Moodie *et al.* (2007)). We might expect $\psi_j^{(2)\top} x_j^{(2)} < 0$ for all histories h_j so that the quadratic indeed has a maximum and not a minimum. This constraint could be incorporated into the model, or the model could be left unconstrained in which case it should be verified that the constraint is indeed satisfied at the estimated value of $\psi_j^{(2)}$. While more complex forms could be considered, using a blip function that is quadratic in a has the advantage from an implementation standpoint of being straightforward to maximize in closed form (a different unimodal function is still easy to maximize numerically). The maximum occurs either at one of the two endpoints of the acceptable range for a, or at the value $a = -\frac{1}{2}(\psi_j^{(1)\top} x_j^{(1)})/(\psi_j^{(2)\top} x_j^{(2)})$.

G-estimation proceeds with the estimating equations taking the form

$$0 = \sum_{i=1}^{n} w_j(H_{ij}) \begin{pmatrix} X_{ij}^{(1)} \{A_{ij} - \pi_j^{(1)}(H_{ij}; \hat{\alpha}_j^{(1)})\} \\ X_{ij}^{(2)} \{A_{ij}^2 - \pi_j^{(2)}(H_{ij}; \hat{\alpha}_j^{(2)})\} \end{pmatrix} \left\{ G_{ij}(\psi_j, \underline{\hat{\psi}}_{j+1}) - \eta_j(H_{ij}; \hat{\varsigma}_j(\psi_j, \underline{\hat{\psi}}_{j+1})) \right\}$$

for $j = 1, \ldots, K$, where $\pi_j^{(1)}(H_j; \alpha_j^{(1)})$ is a model for $E[A_j|H_j]$ as before, and now $\pi_j^{(2)}(H_j; \alpha_j^{(2)})$ is a model for $E[A_j^2|H_j]$. Using a quadratic blip function leads to an attractive property: if we assume a generalized linear model for $E[A_j|H_j]$, then the GLM fit gives both fitted values $\pi_j^{(1)}(H_{ij}; \hat{\alpha}_j^{(1)})$ as well as estimates of the conditional variance $\widehat{\mathrm{Var}}(A_j|H_j; \hat{\alpha}_j)$ (the conditional variance could also be modelled separately as in Joffe *et al.* (2012)), from which we can derive $\pi_j^{(2)}(H_{ij}; \hat{\alpha}_j^{(2)}) = \widehat{\mathrm{Var}}(A_j|H_j; \hat{\alpha}_j) + \{\pi_j^{(1)}(H_{ij}; \hat{\alpha}_j^{(1)})\}^2$. This is similar in spirit to the approach used by Joffe *et al.* (2012), although their treatment model has an additional Bernoulli component for the probability of receiving a zero dose (i.e., their treatment model is zero-inflated).

3 A realistic biological simulation model for continuous dose effects

The simulation approach of Murphy (2003) (for details, please see the Supporting Information) allows one to generating data from a known SNMM but these data do not necessarily resemble what one would typically expect to find in a drug dosing study. In this section we develop a data generating strategy that will allow the evaluation of g-estimation as an optimal dose finding strategy in the realistic setting in which treatments are measured on a continuous scale, and the true model for the outcome as a function of dose and patient history is complex. This is accomplished using a more elaborate simulation protocol and a

PK/PD model that leads to data that conform better to our understanding of the underlying pharmacological principles, at the expense of not knowing the true SNMM.

In pharmaceutical sciences, it is a standard paradigm to assume that the effect of a drug is completely mediated through the true systemic exposure to the drug, i.e., the drug concentration at the site of action. This leads to the decomposition of the dose-effect relationship into two distinct components (and associated models): (1) the dose-concentration relationship; and (2) the concentration-effect relationship. *Pharmacokinetics* concerns dose-concentration relationships, or how the systemic exposure to a drug changes over time as a result of: (a) the administration of the drug; as well as (b) the processes of absorption, distribution, metabolism and excretion. *Pharmacodynamics* concerns concentration-effect (or exposure-effect) relationships.

In this section, we demonstrate the feasibility of utilizing the SNMM framework for continuous-scale treatments described in Section 2 to estimate adaptive dosing strategies assuming that the data are generated from an indirect pharmacodynamic model with random effects on the main structural PK and PD parameters. An indirect PD model was chosen for the simulation because it creates a delay between administration of dose and observed response. This delayed response makes the estimation of an adaptive dosing strategy more challenging. In particular, it implies that a *myopic* strategy that assigns at each interval the dose that is most likely to lead to the targeted response at the end of that interval may not be optimal as it fails to account for residual effects that carry over from one interval to the next. This indirect PD model has been used previously to characterize the PK and PD of warfarin (Dayneka *et al.*, 1993; Blesius *et al.*, 2006).

3.1 Data generating mechanism

To generate the simulation data, we require a treatment model by which doses are adaptively assigned, and a PK/PD model to generate the observed responses. The population PK/PD model is a two-level hierarchical model. The individual-level model, conditional on the individual-specific structural PK/PD parameters, is a standard indirect PD model (Dayneka *et al.*, 1993) that describes, through a system of differential equations, the time-courses of drug concentration in the central compartment (e.g. blood plasma) and the observed response (e.g. prothrombin time or INR), for a particular sequence of doses. The population-level model describes variability between individuals in the population in terms of random effects in the structural PK/PD parameters.

In the simulation, we assume the hypothetical drug to be taken orally once per day, for 21 days. Treatments are modified every 3 days, thus on days 1, 4, 7, 10, 13, 16 and 19. The first six days consist of a *loading* phase which serves to establish steady state conditions. Observations recorded during this phase are considered pre-baseline; this phase is viewed as part of the history at the first treatment interval which begins on day 7. Subsequently, five treatment intervals were considered for dose optimization, the first beginning on day 7 and each interval lasting three days. Thus, in the notation that follows, we assume L_0 is the response on day 7 and A_1 to be the dose assigned on the same day. Because of the loading phase leading up to the start of the first treatment interval, the notation used here differs slightly from that above in that, for notational convenience, we also define L_{-1} and A_0, corresponding to the response observed and dose assigned on day 4 respectively (in the more generic notation presented above this would all have been considered part of the baseline covariate L_0). For simplicity, we assume that for each individual the doses are exactly 24 hours apart, with no deviation in the time of administration. We also assume that a response measurement is taken immediately prior to each dose modification. All individuals are treated and observed for the full 21 days (no dropout or missing values).

Individual-level model

Let $\theta_{i1} = CL_i$, $\theta_{i2} = V_i$ and $\theta_{i3} = k_{a,i}$ be individual-specific pharmacokinetic parameters, where k_a is the absorption rate constant, $k_e = CL/V$ the elimination rate constant, CL the clearance, and V the central volume of distribution, i.e. the blood volume. Further, let $\theta_{i4} = IC_{50,i}$, $\theta_5 = \Upsilon$, $\theta_6 = S_0$ and $\theta_7 = k_d$

be pharmacodynamic parameters. IC_{50} is the concentration of the drug that results in 50% inhibition of clotting factors (prothrombin complexes), assumed to be individual-specific. Υ is a sigmoidity factor, controlling the relationship between drug concentration and clotting factor inhibition. S_0 is the rate of prothrombin production in the absence of the drug and k_d is the degradation rate constant, i.e. the rate at which prothrombin complexes degrade in the body.

The time-courses of the amount of drug in the depot compartment – the gut – $C_0^*(t)$, and the central compartment – the body's circulatory system – $C^*(t)$ are described by the system of differential equations:

$$\frac{\partial C_0^*(t)}{\partial t} = -k_a \times C_0^*(t)$$
$$\frac{\partial C^*(t)}{\partial t} = k_a \times C_0^*(t) - k_e \times C^*(t).$$

This is known in the pharmacokinetics literature as the one-compartment oral model with first-order absorption and first-order elimination (Gibaldi and Perrier, 1982; Gabrielsson and Weiner, 2007). The drug concentration in the central compartment at time t is $C(t) = C^*(t)/V$.

The administration of a dose D causes an instantaneous (i.e., discontinuous) jump in $C_0^*(t)$ at the time of administration by the amount FD where F is the bioavailable fraction, i.e. the amount of the drug that was absorbed, taken here to be 1. For a single dose D administered at time $t = 0$, the system has a closed form solution (Gabrielsson and Weiner, 2007) given by

$$C(t) = \frac{FDk_a}{V(k_a - k_e)}\Big\{ \exp(-k_e t) - \exp(-k_a t)\Big\}.$$

When more than one dose is administered, each new dose results in an instantaneous jump in $C_0^*(t)$ at the time of administration. In the simulation, a dose is taken every 24 hours for 21 days. In order to reach steady-state conditions more rapidly, the central compartment drug concentration of each individual is initialized to 2 at $t = 0$.

The PD component of the model consists of a sigmoid E_{max}-type inhibitory model (Holford, 1986)

$$I\big(C(t)\big) = \frac{1}{1 + \big(C(t)/IC_{50}\big)^{\Upsilon}},$$

and an indirect response model

$$\frac{\partial R(t)}{\partial t} = S_0 \times I\big(C(t)\big) - k_d \times R(t),$$

where in our motivating example, $R(t)$ corresponds to the level of clotting factors in the blood. Initial conditions are chosen so that $R(0) = S_0/k_d$, i.e., the rate of degradation matches the rate of synthesis and the system is in equilibrium. As the concentration of the drug increases, synthesis is inhibited and the rate of synthesis decreases. The response begins to decrease as degradation overtakes synthesis, however the response lags behind the concentration somewhat because it takes time for a new equilibrium state to be established.

This PD equation is solved simultaneously with the PK equations to produce the time-course of the therapeutic response. The observed response L_j, i.e. the deviation of the INR from the target value of 2.5, at time t_j is a transformation of $R(t_j)$ given by

$$L_j = \frac{R_0^{\text{pop}}}{R(t_j)} - 2.5$$

where the constant R_0^{pop} is the (fixed) initial response $R(0)$ (i.e., under no treatment) in the population, taken here to be 120. For simplicity, measurement error is assumed to be negligible. The target value for

L_j is 0. The outcome is defined by the reward function

$$Y = -\sum_{j=1}^{K} |L_j|,$$

i.e., the negative sum of absolute deviations of the responses from their target value. Thus, achieving responses close to the target value at all intervals implies a high value of Y, while deviations from the target value (either positive or negative) are penalized.

Population-level model
Associated with each individual is a vector $\theta_i = (CL, V, k_a, IC_{50}, \Upsilon, S_0, k_d)_i$ of individual-specific PK/PD model parameters. Of these, Υ, S_0 and k_d are assumed to be constant within the population. The remaining parameters are drawn from independent log-normal distributions. Specifically, for $m = 1, \ldots, 4$ we set

$$\theta_{im} = \theta_m^{\text{pop}} \times \exp\left(\eta_{im}\right)$$

where $\eta_{im} \sim N(0, \omega_m^2)$ are random effects, assumed independent, and θ_m^{pop} is the geometric mean or "typical value" of the parameter. The typical values for each parameter were taken from Blesius *et al.* (2006), except for k_a, which was not reported by those authors, and IC_{50}. The IC_{50} value reported by Blesius *et al.* (2006) did not produce suitable INR values in our model, so we used a value in line with those reported by Holford (1986). For k_a we used trial and error to obtain a value such that that the maximum concentration occurred between 2 and 6 hours after the dose as reported by Holford (1986). Thus, the parameter values used were: CL = 0.3, V = 15, k_a = 1.2, IC_{50} = 1.8 Υ = 2, S_0 = 4, k_d = 0.033. For all parameters that vary between individuals we set $\omega_m = 0.2$ so that about 95% of the population have parameter values that are between 68% and 148% of the typical value. Systems of differential equations were solved numerically using the `R` package `deSolve`.

Treatment model
The treatments (or actions) do not correspond to doses directly, but determine the doses through a transformation. On day 1 (pre-baseline), all individuals receive the same dose of 15 mg. On days 4, 7, 10, 13, 16 and 19 ($j = -1, 0, 1, 2, 3, 4, 5$ respectively), the dose is given by:

$$D_{ij} = 15 \times \exp(A_{ij})$$

where the treatments A_{ij} are generated from simple linear regression models

$$A_{ij} = \alpha_0 + \alpha_1 L_{i(j-1)} + \alpha_2 A_{i(j-1)} + \epsilon_{ij}$$

where ϵ_{ij} are independent, normally distributed with mean 0 and standard deviation 0.2, and $A_{i(-2)}$ is taken to be 0 for all $i = 1, \ldots, n$. The random noise term, ϵ_{ij}, is introduced to acknowledge the realistic scenario in which different physicians make different dosing decisions in the face of the same patient covariates, for reasons such as training, past experience, etc. Note that, due to the typical values for PK/PD parameters used in the model, the reference dose of 15 mg/day (corresponding to $A_{ij} = 0$), is somewhat higher than the typical maintenance doses for warfarin observed in practice, which tend to be closer to 5 mg/day. The value of the parameter vector α used in the simulation is $\alpha_0 = 0$, $\alpha_1 = -0.6$ and $\alpha_2 = 0.8$.

3.2 Data, models, and additional scenarios

We generated data sets of sample size $n = 2000$ and sample size $n = 1000$ from the PK/PD model described above. Profiles of three randomly generated individuals are shown in Figure 2.

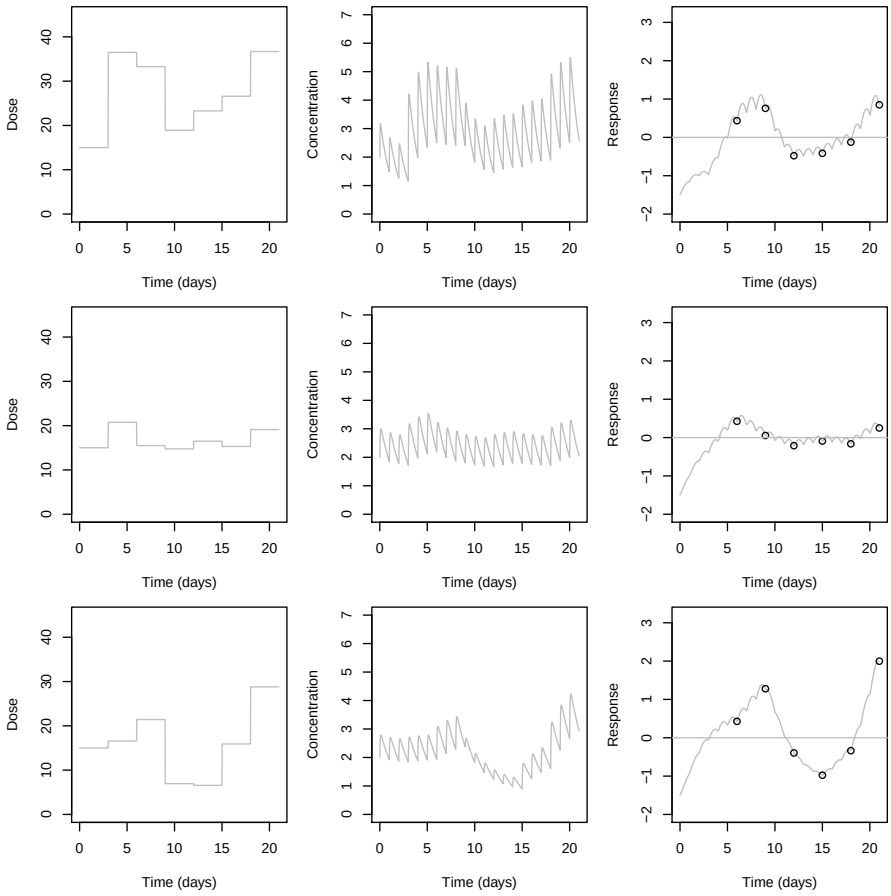


Figure 2 Profiles of three individuals randomly generated from the PK/PD model. Each row corresponds to a different individual. The first column shows the profile of the doses received, the second shows the (latent) drug concentration and the third shows the response profile. Complete time courses are shown, but only the circular points in the response profiles are available for estimation.

Multiple SNMMs were then fit to each simulated data set. Quadratic blip functions as described in Section 2 were used. The models differed only in the linear part of the blip functions, and in each case the quadratic part consisted of a single term (i.e., a constant). The first model included most recent treatment and response as additive terms in the linear part of the blip functions. A second model additionally included a linear interaction term between the most recent treatment and response. A third model included the last two responses (as well as the most recent treatment). The models considered are summarized in Table 1.

Table 1 Models considered in Warfarin dosing simulation.

Model	$\gamma_j(a, h_j)$ for $j = 2, \ldots, 5$
SNMM 1	$a(\psi_{j0} + l_{j-1}\psi_{j1} + a_{j-1}\psi_{j2}) + a^2(\psi_{j3})$
SNMM 2	$a(\psi_{j0} + l_{j-1}\psi_{j1} + a_{j-1}\psi_{j2} + l_{j-1}a_{j-1}\psi_{j3}) + a^2(\psi_{j4})$
SNMM 3	$a(\psi_{j0} + l_{j-1}\psi_{j1} + a_{j-1}\psi_{j2} + l_{j-2}\psi_{j3}) + a^2(\psi_{j4})$

In all cases, flexible cubic spline models were used for the expected counterfactual models. At each interval, the most recent treatment and two most recent responses were included in the EC model in an additive fashion using cubic splines with 3 degrees of freedom—this was based on an evaluation of residual plots (Rich *et al.*, 2010) to produce adequate model fit. Some examples of these residual plots can be found in the online Supporting Information. In all cases, the treatment model was a correctly specified linear regression model.

To test the stability of the results and their sensitivity to parameter settings, the simulation was repeated with modified parameter settings. In a second scenario, we considered the case where doses were adjusted daily rather than every three days, and in a third, we consider a population of "slow metabolizers", i.e., with reduced ability to clear the drug, was considered. To this end, the typical value of the clearance was reduced by 30%, from 0.3 L/h to 0.21 L/h. Optimal dynamic treatment regimens in one population do not carry over to a different population, but may still perform well as feasible and almost optimal regimens, especially if there is some similarity or overlap between the two populations. In the reduced clearance population, individuals would require lower doses on average to achieve the same responses and, due to the increased drug half-life, the residual effects of each dose will have longer range. Nonetheless, the rule used to assign doses based on history may be quite similar for the two populations.

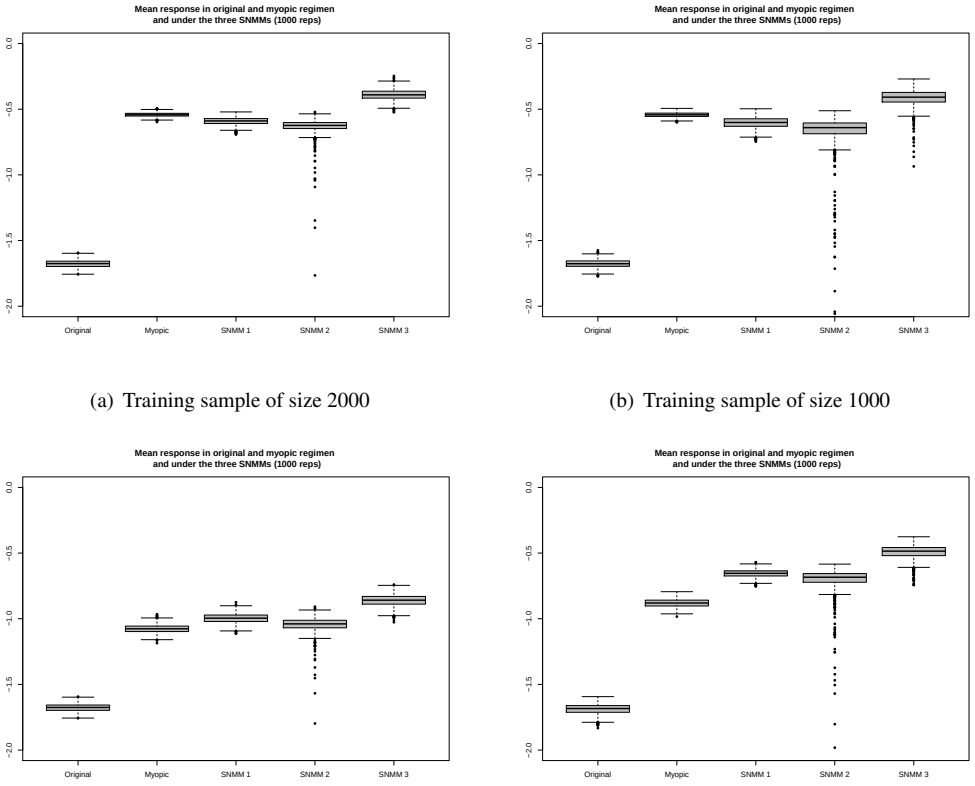
3.3 Evaluation

The relative performance of the DTRs estimated by the different models were evaluated and compared to a simple myopic dynamic regime. For each of the 1000 simulations, a training set was used to estimate parameters of each of the three SNMMs, thus providing estimated dosing regimens. A simple myopic dynamic regimen was obtained from the same data by fitting linear models regressing the response at each interval on the most recent treatment and response. The myopic regimen assigns at each interval the treatment for which the predicted response is zero. Note that unlike the SNMM approach, the myopic strategy does not consider the terminal outcome, only the next value of L_j. These regimens were then each used to "treat" new individuals (test set).

The performance of the different regimens can be compared according to different criteria. One approach is to compare the mean of the outcome Y in a test set across regimens, as in Figure 3. In this case, the SNMM regimens demonstrate an advantage over the myopic regimen. Whether or not a myopic regimen will perform well in a given situation will depend on the characteristics of the data generating mechanism, in particular on the magnitude of the direct effect of treatment on responses more distant than the current interval. In this particular simulation, there is relatively little carry-over effect of treatments past the end of each three-day interval which implies that a myopic strategy can perform reasonably well. The probability that SNMM 3 provides a better outcome, for an individual, than the myopic regime is 0.69 and 0.73 when regimes were estimated with training set sample sizes of $n = 1000$ and 2000, respectively.

A second, more 'individual-level' criterion for comparing the performance of the different regimens is to determine for each individual in the test set the regimen that leads to the highest response. These proportions are presented in Figure 4. In this winner-takes-all approach, SNMM 3, which incorporates more past information into the dosing decision, clearly performs best. When the optimal dosing strategy is estimated using a smaller training sample, results are remarkably consistent: when estimating the strategy using 500 individuals, SNMM 3 produces the best outcome for 77% of the test set; using a sample size of 250 to estimate the optimal strategy found SNMM 3 led to the best outcome for 90% of the test set (for further details, see Figures 2–4 in the Supporting Information).

If the duration of the treatment intervals is reduced, results change markedly. For instance, the results in Figures 3(c) and 4(c) were obtained from a similar simulation set-up, but with the time between dose adjustments reduced from three days to one day in the test set. In this scenario, the effect of each dose extends well past the end of the dose interval and so the SNMM regimens, which account for the overall course of treatment and response, perform much better. In both cases, there appears to be no benefit to



(a) Training sample of size 2000

(b) Training sample of size 1000

(c) Test set adjusting doses daily

(d) Test set has a slower drug clearance

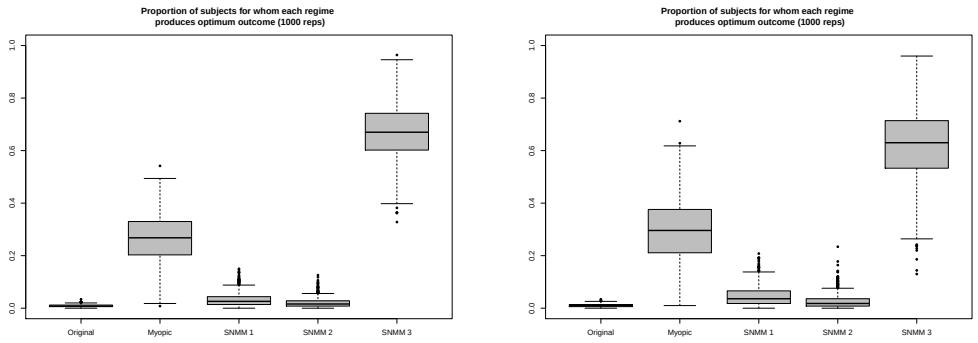
Figure 3 Distribution of the mean outcome Y obtained from the application of four different dynamic regimens to 500 new individuals (the test set) across 1000 simulations: (a) and (b) are setting in which the training sample and test set have identical characteristics; in (c), the doses are changed daily in the test set rather than every three days as in the training sample; in (d) the test set comes from a population with slower clearance rate, i.e. the PD characteristics of the training sample and test set differ.

including an interaction term in the blip model (SNMM 2), while the incorporation of an additional lagged response (SNMM 3) does confer improved outcomes.

We also evaluated the performance of each of these dynamic treatment regimens when used to treat a population different from that from which the data used for estimation were generated, namely the population of slow metabolizers. As discussed, these regimens may not be optimal in the different population, but nonetheless it is of interest to evaluate their performance in terms of external validity. The results are presented in Figures 3(d) and 4(d). It is observed that, while individuals from this population have worse outcomes on average, the dynamic treatment regimens still appear to provide useful treatment strategies that effectively lead to better outcomes.

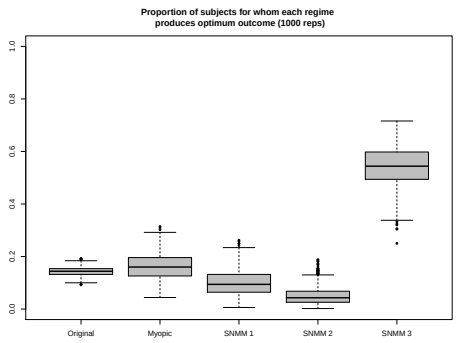
4 Discussion

In this paper, we provided the first realistic demonstration of the use of dynamic treatment regimen methodology, specifically G-estimation, to suggest individualized dosing strategies in the context of a continuous

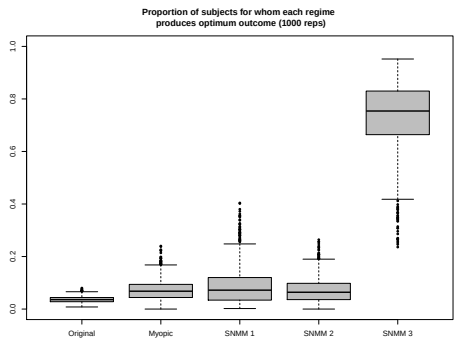


(a) Training sample of size 2000

(b) Training sample of size 1000



(c) Test set adjusting doses daily



(d) Test set has a slower drug clearance

Figure 4 Distribution of the proportion of individuals in the test set for whom a particular regime yields the best outcome across 1000 simulations: (a) and (b) are setting in which the training sample and test set have identical characteristics; in (c), the doses are changed daily in the test set rather than every three days as in the training sample; in (d) the test set comes from a population with slower clearance rate, i.e. the PD characteristics of the training sample and test set differ.

treatment. G-estimation, like other semi-parametric approaches to optimal dynamic treatment regimens, was developed for the setting where the conditional distribution of the responses given past treatments, responses and other covariates is complex and difficult to specify with a parametric model. We have correspondingly developed a framework for simulating data that mimic complex biological processes, as described by PK/PD models. This is, to our knowledge, the first time PK/PD information has been incorporated into the design of simulations for dynamic treatment regimens. It is an approach that may prove invaluable in the design of randomized trials for individualized dosing strategies, by allowing researchers to narrow the field of candidate regimens.

While our warfarin example did not include unknown or unmeasured factors that impact on the dose-response relationship, such as dietary or environmental factors that affect a drug's PK/PD at the individual level; omitting such factors should have no impact on bias provided that any unmeasured covariates are not confounders, i.e., they do not feature in the dosing decisions used to generate the observed data. It should be noted, however, that such covariates could easily be incorporated into the data generating models to enhance the realism of the simulation. Genetic variants, for example, could be simulated and incorporated

into the PK/PD model, although their usefulness has been found to be limited in practice (Furie, 2013, e.g.). If we are interested in pharmacogenetically guided dosing strategies, these patient-level characteristics could be included in the blip function as well. Local environmental changes, such as diet, could also be input into the data generating models but we would not consider them for dose adjustment as that level of tailoring is not realistic in clinical care.

Parametric approaches to optimal dosing have been considered Funatogawa and Funatogawa (2012), however model mis-specification was not considered. While the PK/PD characteristics of warfarin have been well-understood for many years, researchers have yet to use these nonlinear equations to suggest a dosing strategy that is optimal or that accounts for delayed effects of treatment or can incorporate interactions with other individual-level characteristics including laboratory measurements or the use of other medications. Fitting the PK/PD models is relatively straightforward; however determining the optimal dosing strategy is a large and complex optimization problem which does not have a closed-form solution due to its non-myopic nature. Given the complex and non-linear dependence of common outcomes (such as time in therapeutic range) on doses and other factors, we surmise that the utility of such approaches are extremely limited in practice.

Much remains to be explored, and considerable work is needed before g-estimation can be adequately applied to data such as those from the London anticoagulation clinic. Particular challenges include defining the outcome when visits are missed (since, for example, sums of deviations from the target INR or proportion of time in target range are no longer meaningful with different lengths of follow-up); irregular number and timing of visits between individuals; computational and theoretical aspects of the estimation under the assumption that parameters are shared across treatment intervals; and the possibility that important confounders may not be recorded in the electronic medical records. Thus, an important next step will be to consider how to make our simulations, which are realistic in terms of the complexity of the treatment effects, less 'idealized' in terms of the many ways that data acquisition and collection occurs in clinical settings, including irregular timing of follow-up visits, missed visits, and poor compliance with a dosing schedule.

Acknowledgements We are grateful to Frances Akor, Michael Laffan, and Sanjay Patel for providing the data shown in Figure 1. Drs. Moodie and Stephens are supported by Discovery Grants from the the Natural Sciences and Engineering Research Council of Canada (NSERC). The majority of this work was undertaken while Dr. Rich was undertaking doctoral studies, which were supported by a fellowship from NSERC.

Conflict of Interest

The authors have declared no conflict of interest.

References

Blesius, A., Chabaud, S., Cucherat, M., Mismetti, P., Boissel, J.P., and Nony, P. (2006) Compliance-guided therapy: A new insight into the potential role of clinical pharmacologists. *Clinical Pharmacokinetics*, **45** (1), 95–104.

Budnitz, D.S., Lovegrove, M.C., Shehab, N., and Richards, C.L. (2011) Emergency hospitalizations for adverse drug events in older Americans. *The New England Journal of Medicine*, **365** (21), 2002–2012.

Chakraborty, B. and Moodie, E.E.M. (2013) *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*, Springer, New York.

Cotton, C.A. and Heagerty, P.J. (2011) A data augmentation method for estimating the causal effect of adherence to treatment regimens targeting control of an intermediate measure. *Statistics in Biosciences*, **3** (1), 28–44.

Cromme, L., Völler, H., Gäbler, F., Salzwedel, A., and Taborski, U. (2010) Computer-aided dosage in oral anticoagulation therapy using phenprocoumon: Problems and approaches. *Hämostaseologie*, **30** (4), 183–189.

D'Argenio, D.Z. and Khakmahd, K. (1983) Adaptive control of theophylline therapy: Importance of blood sampling times. *Journal of Pharmacokinetics and Biopharmaceutics*, **11** (5), 547–559.

Dawid, A.P. and Didelez, V. (2010) Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Statistics Surveys*, **4** (October), 184–231.

Dayneka, N.L., Garg, V., and Jusko, W.J. (1993) Comparison of four basic models of indirect pharmacodynamic responses. *Journal of Pharmacokinetics and Biopharmaceutics*, **21** (4), 457–478.

Funatogawa, I. and Funatogawa, T. (2012) Dose-response relationship from longitudinal data with response-dependent dose modification using likelihood methods. *Thrombosis Research*, **54**, 494–506.

Furie, B. (2013) Do pharmacogenetics have a role in the dosing of vitamin K antagonists? *New England Journal of Medicine*, **369**, 2345–2346.

Fusaro, V.A., Patil, P., Chi, C.l., Contant, C.F., and Tonellato, P.J. (2013) A systems approach to designing effective clinical trials using simulations. *Circulation*, **127** (4), 517–526.

Gabrielsson, J. and Weiner, D. (2007) *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications, Fourth Edition*, Swedish Pharmaceutical Press, Stockholm.

Gibaldi, M. and Perrier, D. (1982) *Pharmacokinetics*, Marcel Dekker, New York.

Henderson, R., Ansell, P., and Alshibani, D. (2010) Regret-regression for optimal dynamic treatment regimes. *Biometrics*, **66** (4), 1192–1201.

Hirsh, J., Dalen, J.E., Anderson, D.R., Poller, L., Bussey, H., Ansell, J., and Deykin, D. (2001) Oral anticoagulants: Mechanism of action, clinical effectiveness, and optimal therapeutic range. *Chest*, **119** (1 Suppl), 8S–21S.

Holford, N.H.G. (1986) Clinical pharmacokinetics and pharmacodynamics of warfarin: Understanding the dose-effect relationship. *Clinical Pharmacokinetics*, **11** (6), 483–504.

Joffe, M.M., Yang, W.P., and Feldman, H. (2012) G-estimation and artificial censoring: Problems, challenges, and applications. *Biometrics*, **68** (1), 275–286.

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., and Murphy, S. (2012) A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology*, **8**, 21–48.

Michaud, V., Vanier, M.C., Brouillette, D., Roy, D., Verret, L., Noel, N., Taillon, I., O'Hara, G., Gossard, D., Champagne, M., Goodman, K., Renaud, Y., Brown, A., Phillips, M., Ajami, A.M., and Turgeon, J. (2008) Combination of phenotype assessments and CYP2C9–VKORC1 polymorphisms in the determination of warfarin dose requirements in heavily medicated patients. *Clinical Pharmacology and Therapeutics*, **83** (5), 740–748.

Moodie, E.E.M. (2009) A note on the variance of doubly-robust G-estimators. *Biometrika*, **96** (4), 998–1004.

Moodie, E.E.M. and Richardson, T.S. (2010) Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics*, **37** (1), 126–146.

Moodie, E.E.M., Richardson, T.S., and Stephens, D.A. (2007) Demystifying optimal dynamic treatment regimes. *Biometrics*, **63** (2), 447–455.

Murphy, S.A. (2003) Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65** (2), 331–355.

Nielsen, P.B., Lundbye-Christensen, S., Rasmussen, L.H., and Larsen, T.B. (2014) Improvement of anticoagulant treatment using a dynamic decision support algorithm: A Danish cohort study. *Thrombosis Research*, **133**, 375–379.

Poller, L., Keown, M., Ibrahim, S., Lowe, G., Moia, M., Turpie, A., Roberts, C., van den Besselaar, A., van der Meer, F., Tripodi, A., Palareti, G., Shiach, C., Bryan, S., Samama, M., Burgess-Wilson, M., Heagerty, A., Maccallum, P., Wright, D., and Jespersen, J. (2008) An international multicenter randomized study of computer-assisted oral anticoagulant dosage vs. medical staff dosage. *Journal of Thrombosis and Haemostasis*, **6**, 935–943.

Poller, L., Keown, M., Ibrahim, S., Lowe, G., Moia, M., Turpie, A., Roberts, C., van den Besselaar, A., van der Meer, F., Tripodi, A., Palareti, G., Shiach, C., Bryan, S., Samama, M., Burgess-Wilson, M., Heagerty, A., Maccallum, P., Wright, D., Jespersen, J., and on Anticoagulation (EAA), T.E.A. (2009) A multicentre randomised assessment of the DAWN AC computer-assisted oral anticoagulant dosage program. *Thrombosis and Haemostasis*, **101**, 487–494.

Rich, B., Moodie, E.E.M., Stephens, D.A., and Platt, R.W. (2010) Model checking with residuals for g-estimation of optimal dynamic treatment regimes. *The International Journal of Biostatistics*, **6** (2), Article 12.

Robins, J.M. (2004) Optimal structural nested models for optimal sequential decisions, in *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data* (eds D.Y. Lin and P.J. Heagerty), Springer.

Rosthøj, S., Fullwood, C., Henderson, R., and Stewart, S. (2006) Estimation of optimal dynamic anticoagulation regimes from observational data: A regret-based approach. *Statistics in Medicine*, **25**, 4197–4215.

Teml, A., Schaeffeler, E., Herrlinger, K.R., Klotz, U., and Schwab, M. (2007) Thiopurine treatment in inflammatory bowel disease. *Clinical Pharmacokinetics*, **46** (3), 187–208.

The International Warfarin Pharmacogenetics Consortium (2009) Estimation of the warfarin dose with clinical and pharmacogenetic data. *The New England Journal of Medicine*, **360** (8), 753–764.

Wysowski, D.K., Nourjah, P., and Swartz, L. (2007) Bleeding complications with warfarin use: A prevalent adverse effect resulting in regulatory action. *Archives of Internal Medicine*, **167** (13), 1414–9.

Supporting Information to "Optimal individualized dosing strategies: A pharmacologic approach to developing dynamic treatment regimens for continuous-valued treatments" by Benjamin Rich, Erica E. M. Moodie and David A. Stephens

1 A simplistic simulation for continuous dose effects

Here, we demonstrate the usual approach to simulating data to evaluate DTR estimating approaches in the case where the blip model for a continuous dose has been correctly specified, that is, where the data generating mechanism conforms to a known SNMM with continuous treatments, and study the small and large sample performance of g-estimation.

We consider a simulated example in which there are three treatment intervals. The treatments (doses) are given on a continuous scale. Each treatment depends linearly on the previous observation, but not on the previous treatment. That is, treatments may be viewed dose adjustments relative to the previously given dose; for the first interval, the adjustment is relative to a standard dose.

Data with continuous-scale treatments are generated according to an SNMM as follows. First, a baseline observation is drawn from a normal distribution: $L_0 \sim N(0, 4)$. Then, for treatment intervals $j = 1, 2, 3$, the treatments and observations are generated according to:

$$A_j \sim N(-0.231 L_{j-1}, 0.04),$$
$$L_j \sim N(0.2 L_{j-1} - 0.3 A_j, 0.25).$$

The blip functions at each interval are specified as:

$$\gamma_j(a, h_j) = a(\psi_{j0} + \psi_{j1} l_{j-1}) + a^2 \psi_{j2},$$

and finally the outcome is generated according to

$$Y \sim N(5 + 0.2 L_0 - \mu_1(A_1, H_1) - \mu_2(A_2, H_2) - \mu_3(A_3, H_3), 0.25)$$

where $\mu_j(a, h_j) = \max_{a'} \gamma_j(a', h_j) - \gamma(a, h_j)$, known as the regret function.

The parameters ψ_{j0}, ψ_{j1} and ψ_{j2} are varied to create different simulation scenarios. We consider sample sizes $n = 500$, 1,000, 2,000 and 10,000 generated under the scenario $\psi_{10} = \psi_{20} = \psi_{30} = 0.25$, $\psi_{j1} = \psi_{j1} = \psi_{j1} = -0.12$ and $\psi_{j2} = \psi_{j2} = \psi_{j2} = -0.5$.

To study the properties of the g-estimation procedure for continuous treatments described in Section 2 of the main text, we simulated 1000 data sets of varying sample sizes according to the data generating mechanism just described. The results are displayed in Table 1. We observe that bias is low, and coverage of the Wald-type confidence intervals based on the asymptotic variance sandwich estimator is close to the nominal level, for sample sizes 2000 and up, whereas for sample size 500, there is substantial bias in the estimators of the parameters relating to the first two treatment intervals, suggesting that for this sample size asymptotic convergence of the estimators has not been reached. Additional simulations performed with different parameter settings produced similar results (not shown).

The simulations described in this section are useful for studying properties of g-estimators (or other approaches to estimating dynamic treatment regimen), owing to the ease with which they are understood and implemented and the significant advantage of knowing the true regimen parameters. However it is difficult to conceive of many scenarios that would conform to such simplistic models. These models are unlikely to be useful in providing guidance in any real treatment setting.

Table 1: Simulation results for g-estimation with continuous treatment and quadratic blip function for the scenario $\psi_{10} = \psi_{20} = \psi_{30} = 0.25$, $\psi_{j1} = \psi_{j1} = \psi_{j1} = -0.12$ and $\psi_{j2} = \psi_{j2} = \psi_{j2} = -0.5$. Based on 1000 replicates. A '*' in the last column indicates that the coverage is significantly different from 95%.

Sample size	Truth	MC mean	MC SE	Mean est. SE	Absolute bias	Percent bias	Coverage
n=500	$\psi_{10} = 0.25$	0.080	13.387	0.324	-0.170	-68.2	93.8
	$\psi_{11} = -0.12$	-0.475	21.667	0.604	-0.355	296.2	97.0 *
	$\psi_{12} = -0.50$	-0.717	9.847	1.183	-0.217	43.4	96.2
	$\psi_{20} = 0.25$	0.226	2.777	0.164	-0.024	-9.6	92.2 *
	$\psi_{21} = -0.12$	-0.172	4.201	0.330	-0.052	43.2	93.5 *
	$\psi_{22} = -0.50$	-0.586	9.092	0.548	-0.086	17.2	93.9
	$\psi_{30} = 0.25$	0.247	0.117	0.111	-0.003	-1.0	93.8
	$\psi_{31} = -0.12$	-0.122	0.289	0.273	-0.002	1.5	93.4 *
	$\psi_{32} = -0.50$	-0.498	0.411	0.391	0.002	-0.3	93.3 *
n=1000	$\psi_{10} = 0.25$	0.241	0.567	0.097	-0.009	-3.6	93.1 *
	$\psi_{11} = -0.12$	-0.125	0.428	0.172	-0.005	4.6	96.2 *
	$\psi_{12} = -0.50$	-0.491	0.714	0.347	0.009	-1.9	95.6
	$\psi_{20} = 0.25$	0.234	0.734	0.086	-0.016	-6.4	91.6 *
	$\psi_{21} = -0.12$	-0.110	0.498	0.183	0.010	-8.7	93.7
	$\psi_{22} = -0.50$	-0.489	0.491	0.302	0.011	-2.2	94.0
	$\psi_{30} = 0.25$	0.249	0.080	0.079	-0.001	-0.3	94.9
	$\psi_{31} = -0.12$	-0.120	0.203	0.195	0.000	0.0	93.6 *
	$\psi_{32} = -0.50$	-0.497	0.291	0.279	0.003	-0.6	93.6 *
n=2000	$\psi_{10} = 0.25$	0.248	0.078	0.058	-0.002	-0.8	94.2
	$\psi_{11} = -0.12$	-0.122	0.102	0.100	-0.002	1.9	95.7
	$\psi_{12} = -0.50$	-0.500	0.209	0.205	0.000	0.0	95.5
	$\psi_{20} = 0.25$	0.249	0.089	0.057	-0.001	-0.6	94.2
	$\psi_{21} = -0.12$	-0.122	0.127	0.121	-0.002	1.8	94.6
	$\psi_{22} = -0.50$	-0.501	0.223	0.200	-0.001	0.1	94.5
	$\psi_{30} = 0.25$	0.251	0.056	0.056	0.001	0.3	95.0
	$\psi_{31} = -0.12$	-0.123	0.140	0.139	-0.003	2.5	94.6
	$\psi_{32} = -0.50$	-0.504	0.202	0.198	-0.004	0.7	94.5
n=10,000	$\psi_{10} = 0.25$	0.249	0.026	0.025	-0.001	-0.4	93.4 *
	$\psi_{11} = -0.12$	-0.122	0.042	0.043	-0.002	1.5	95.4
	$\psi_{12} = -0.50$	-0.502	0.088	0.089	-0.002	0.5	95.2
	$\psi_{20} = 0.25$	0.250	0.026	0.025	0.000	0.0	94.6
	$\psi_{21} = -0.12$	-0.121	0.044	0.044	-0.001	0.6	94.8
	$\psi_{22} = -0.50$	-0.500	0.091	0.090	0.000	0.0	94.6
	$\psi_{30} = 0.25$	0.250	0.025	0.026	0.000	0.0	95.0
	$\psi_{31} = -0.12$	-0.120	0.045	0.045	0.000	0.0	95.1
	$\psi_{32} = -0.50$	-0.499	0.091	0.091	0.001	-0.2	94.8

2 Additional results for the realistic simulation for continuous dose effects

Figure 1 consists of residual plots for the PK/PD simulation, constructed using the method described by (4). These plots are provided as an example of the much larger number of plots considered. Cubic splines were utilized to obtain adequate fit of the EC model (bottom row of panels).

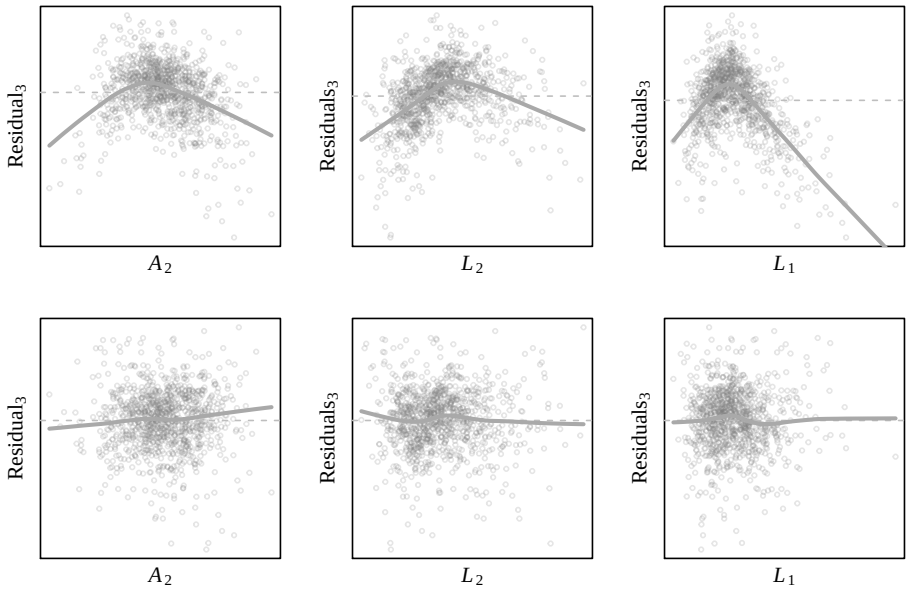
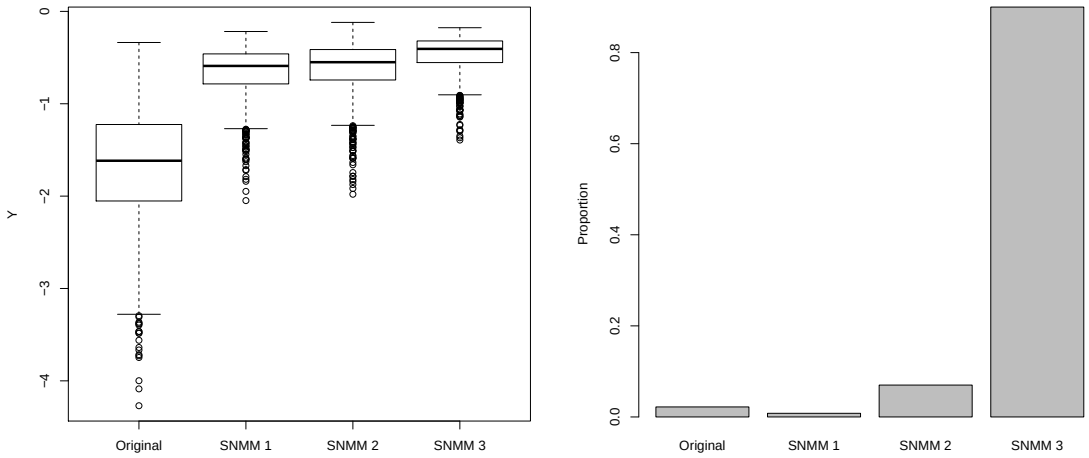


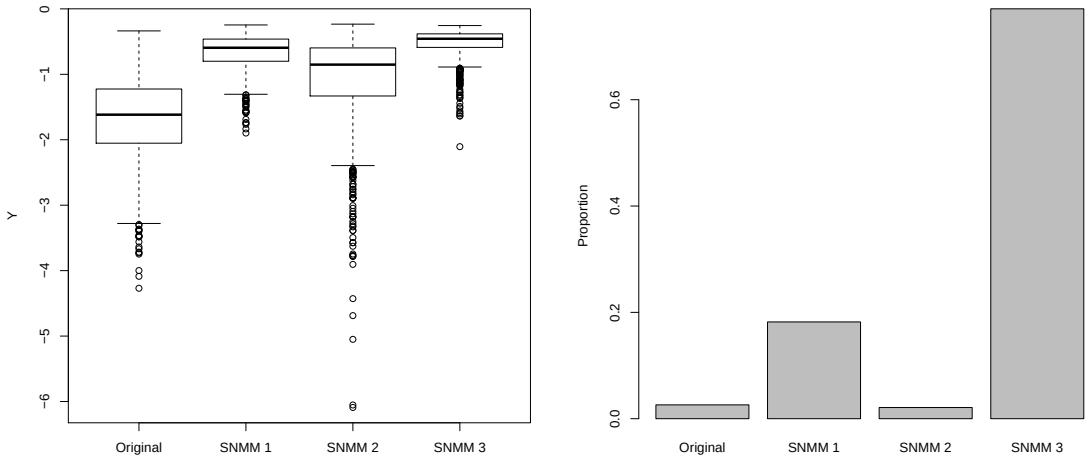
Figure 1: Residual plots from the fits of two different EC models to a data set of size $n = 1,000$ simulated from the PK/PD model. The residual at interval 3 is plotted against: most recent treatment A_2 (first column); most recent response L_2 (second column); next most recent response L_1 (third column). In the top row, the corresponding terms are omitted from the EC model while in the bottom row all terms are included (as cubic splines).

Figures 2 and 3 show the distribution of mean outcomes under the three SNMMs considered, as well as the distribution of the proportion of new individuals in the test sets for which each treatment regimen produced the highest outcome using sample sizes of 250 and 500, respectively, to estimate the optimal individualized dosing strategy. Figure 4 plots the outcomes in a single test set of 1,000 new individuals under the optimal regimen given by SNMM 3 estimated using a sample of size 250 as compared to the same regimen estimated using a sample of size 2,000.



(a) Distribution of outcomes Y (b) Regimen yielding the best outcome by individual

Figure 2: Results from the application of three different dynamic regimens estimated with a sample size of 250 (training set) to 1,000 new individuals (the test set).



(a) Distribution of outcomes Y (b) Regimen yielding the best outcome by individual

Figure 3: Results from the application of three different dynamic regimens estimated with a sample size of 500 (training set) to 1,000 new individuals (the test set).

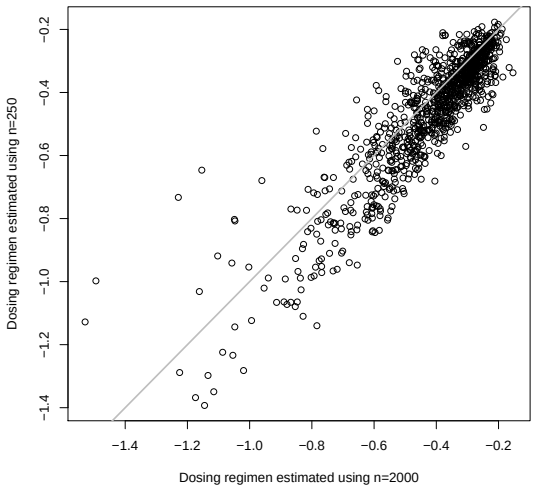


Figure 4: Outcomes in 1,000 new individuals (the test set) under the optimal regimen given by SNMM 3 estimated using a sample of size 250 (y-axis) as compared to optimal regimen given by SNMM 3 estimated using a sample of size 2000 (x-axis). The grey line indicates where $y = x$.

References

[1] Erica E. M. Moodie. A note on the variance of doubly-robust G-estimators. *Biometrika*, 96(4):998–1004, 2009.

[2] Erica E. M. Moodie and Thomas S. Richardson. Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics*, 37(1):126–146, 2010.

[3] Erica E. M. Moodie, Thomas S. Richardson, and David A. Stephens. Demystifying optimal dynamic treatment regimes. *Biometrics*, 63(2):447–455, 2007.

[4] Benjamin Rich, Erica E. M. Moodie, David A. Stephens, and Robert W. Platt. Model checking with residuals for g-estimation of optimal dynamic treatment regimes. *The International Journal of Biostatistics*, 6(2):Article 12, 2010.

[5] James M. Robins. Optimal structural nested models for optimal sequential decisions. In D Y Lin and P J Heagerty, editors, *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*. Springer, 2004.