POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 0, No. 0, xxxx–xxxx 2015, pp. 1–17
ISSN 1059-1478 | EISSN 1937-5956 | 15 | 00 | 0001
DOI 10.1111/poms.12514

Designing Personalized Treatment: An Application to Anticoagulation Therapy

Rouba Ibrahim
UCL School of Management, University College London, London WC1E 6BT, UK, rouba.ibrahim@ucl.ac.uk

Beste Kucukyazici, Vedat Verter
Desautels Faculty of Management, McGill University, Montreal, Québec H3A 0G4, Canada, beste.kucukyazici@mcgill.ca, vedat.verter@mcgill.ca

Michel Gendreau
Mathematical and Industrial Engineering, Ecole Polytechnique Montreal, Montreal, Québec H3T 1J4, Canada, michel.gendreau@cirrelt.ca

Mark Bolstein
Montreal Jewish General Hospital, Montreal, Québec H3T 1E2, Canada, mark.blostein@mcgill.ca

In this study, we develop an analytical framework for personalizing the anticoagulation therapy of patients who are taking warfarin. Consistent with medical practice, our treatment design consists of two stages: (i) the initiation stage, modeled using a partially-observable Markov decision process, during which the physician learns through systematic belief updates about the unobservable patient sensitivity to warfarin, and (ii) the maintenance stage, modeled using a Markov decision process, during which the physician relies on his formed belief about patient sensitivity to determine the stable, patient-specific, warfarin dose to prescribe. We develop an expression for belief updates in the POMDP, establish the optimality of the myopic policy for the MDP, and derive conditions for the existence and uniqueness of a myopically optimal dose. We validate our models using a real-life patient data set gathered at the Hematology Clinic of the Jewish General Hospital in Montreal. The proposed analytical framework and case study enable us to develop useful clinical insights, for example, concerning the length of the initiation period and the importance of correctly assessing patient sensitivity.

Key words: personalized treatment; stroke prevention; treatment design; warfarin
History: Received: April 2013; Accepted: September 2015 by Sergei Savin, 2 revisions.

1. Introduction

The field of medicine has spent a considerable effort on the standardization of care during the past few decades. However, recognizing the drawbacks of this "one size fits all" approach, the more recent trend has been the development of *personalized* treatment schemes which incorporate patient-specific characteristics and are tailored to each patient's personal needs; for example, see Hamburg and Collins (2010) and Personalized Medicine Coalition (2015).

In this study, we focus on anticoagulation therapy for stroke prevention in atrial fibrillation (AF), and present an analytical framework for its personalization. In so doing, we are motivated by recent clinical evidence presented by the American Heart Association (2014) who stated that "most management strategies for (anticoagulation therapy) should be individualized," that "more research is needed for individualized approaches," and that determining "the optimal treatment (. . .) remains challenging."

Atrial Fibrillation and Warfarin. Warfarin is the most commonly prescribed anticoagulant drug in the world. However, despite warfarin's popularity, its usage is complicated by a narrow therapeutic window. Indeed, a high warfarin dose could have life-threatening consequences, such as excessive bleeding, while a low warfarin dose could be ineffective in preventing the formation of blood clots which, in turn, increases the risk of stroke; see Oldgren et al. (2011). Moreover, patient response to warfarin is affected by both demographic and clinical factors (age, ethnicity, gender, and diet), and is strongly influenced by several comorbid conditions. Warfarin is also known to interact with other medication that the patient may be taking;for example, see White (2010). Due to those complications, warfarin remains designated as one of the most unsafe drugs; see the American Stroke Association (2015) and the Institute for Safe Medication Practices (2012).

Recent medical research indicates that pharmacogenetic factors, that is, genetic differences between

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

patients, may significantly affect patient response to warfarin; see Botton et al. (2011). Nevertheless, genetic testing remains difficult, costly, and not widely used; for example, see Bhatt (2014), Ginsburg and Voora (2010), and Patrick et al. (2009). This evidence has prompted Holbrook et al. (2012) to specify in the 2012 ACCP guidelines that genetic testing is "not (considered) cost-effective by most drug policy experts" (p. e159S). Therefore, there is a need for a treatment design that enables systematic learning about that unobservable patient sensitivity without expensive genetic testing. In this study, we propose such a design.

Personalized Treatment Design for AF. Physicians typically use the International Normalized Ratio (INR) to quantify patient response to warfarin. The desired INR range for patients with AF is between 2 and 3. There are two stages involved (International Warfarin Pharmacogenetics Consortium 2009). In the initiation stage, the physician learns about patient sensitivity so as to prescribe an appropriate warfarin dose. In the maintenance stage, the physician assumes prior knowledge of his patient's sensitivity, with some acceptable degree of uncertainty, and prescribes a stable warfarin dose accordingly. In this study, we model both stages involved in anticoagulation therapy. In particular, we use a partially-observable Markov decision process (POMDP) to model the initiation stage of treatment and enable learning about the unobservable patient sensitivity, and a Markov decision process (MDP) to model the maintenance stage of treatment. As such, ours is the first study which proposes a systematic approach for the entire treatment process.

Contributions. We develop a closed-form expression for belief updates (about sensitivity) in the POMDP framework; we prove that the myopic policy is optimal for the MDP model; and we formulate sufficient conditions for the existence and uniqueness of the myopically optimal warfarin dose for the MDP model: This is the *stable* dose; see Botton et al. (2011). We also analyze the impact of this risk-minimizing dose on the time in therapeutic range (TTR). The TTR is defined as the proportion of visits that the INR remains within the target interval (2, 3) (Ansell et al. 2001). We prove that contrary to common medical belief, a risk-minimizing dose may not always lead to a desirable TTR.

Our work is grounded in the realities of current medical practice. The real-life patient data, which we use in calibrating the proposed methodology, are gathered at the Hematology Clinic of the Jewish General Hospital in the city of Montreal, Canada. We show that our proposed models fit the data well. Last but not least, we conduct a detailed numerical study and show, numerically, that the myopic policy is very close to optimal for the POMDP model. This suggests that it suffices for physicians to solve the single-stage belief updating problem, at each decision epoch, during the initiation stage. We also formulate several recommendations, for example, concerning the required length of the initiation period and how it is affected by the level of patient sensitivity, the effect of previous adverse events on subsequent treatment, and the importance of correctly assessing patient sensitivity.

Literature Review. In this study, we contribute to the emerging body of literature employing operational research techniques for the *personalization* of medical guidelines, which was designated as one of the main open research challenges in Denton et al. (2011). In a setting such as ours, where medical decisions need to be made sequentially in highly stochastic environments, Markov decision processes (MDP) are ideally suited; see Schaefer et al. (2004) for a comprehensive survey. More recent references include MDP applications to liver transplantation (Alagoz et al. 2007), to HIV (Shechter et al. 2008), ovarian hyperstimulation (He et al. 2010), the timing of biopsy decisions in mammography screening (Chhatwal et al. 2010), and the optimizing of diagnostic decisions after a mammography (Ayvaci et al. 2012). POMDPs are an extension of MDPs which relax the assumption that the state of the system is known. Recent references include POMDP applications to heart disease (Hauskrecht and Fraser 2000), Parkinson's disease (Goulionis and Vozikis 2009), colorectal cancer (Leshno et al. 2003), breast cancer (Maillart et al. 2008 and Ayer et al. 2012), and stroke (Coroian and Hauser 2015). The work of Coroian and Hauser is different from ours, in that the authors do not solve the problem of determining personalized dosing decisions, as we do in this study. Instead, they focus on selecting the learning technique in the POMDP framework which optimizes predictive accuracy. Botton et al. (2011), The International Warfarin Pharmacogenetics Consortium (2009), Millican et al. (2007), Witt et al. (2009), and references therein, use regression models to determine the stable warfarin dose as a function of several covariates. However, they do not address how to sequentially learn about an unobservable patient sensitivity.

The remainder of this study is organized as follows. In section 2, we describe our two-stage framework. In section 3, we describe our POMDP model, and in section 4, our MDP model. In section 5, we describe our data set. In section 6, we fit our POMDP model to data. In section 7, we describe our detailed numerical study. In section 8, we make concluding remarks. We include supportive material in an online supplement to this study.

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

2. General Framework: A Two-Stage Methodology

In Figure 1, we describe the conceptual framework of this study. We use a POMDP framework to model the initiation stage of treatment. At the onset of treatment, the physician has some initial belief about the patient's sensitivity. At each patient visit to the clinic, the physician measures the patient's INR. Based on this new INR measurement and the patient's treatment history, the physician updates his/her belief about the patient's sensitivity. Based on this updated belief, the physician aims to prescribe an optimal, risk-minimizing, warfarin dose. We quantify the risk associated with a given INR value by using a convex combination of bleeding and stroke relative risks. For the functional forms of the bleeding and stroke risks, we rely on the literature; see Hylek et al. (2003).

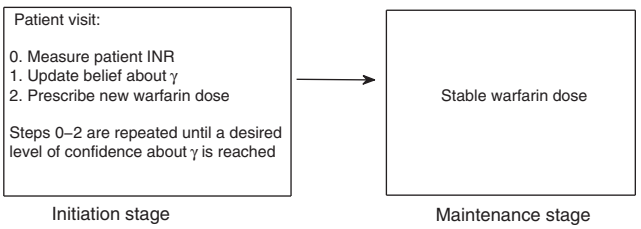
The belief-updating steps are repeated until the physician forms a belief about the patient's sensitivity which is deemed to be sufficiently accurate. We assume that the convergence criterion for the POMDP (the level of desired accuracy about the patient's sensitivity) is exogenously specified. We do so intentionally because of the inherent trade-off between the clinical benefits of enhanced learning about patient sensitivity on one hand, and the inconvenience of frequent visits to the clinic during the initiation stage on the other hand. Indeed, clinical visits are typically costly for patients; for example, see Hwang et al. (2011) and Attaya et al. (2012). Since we take the physician's perspective in this work, we do not consider additional costs incurred by the patient per visit to the anticoagulation clinic. However, we indirectly allow for those costs by assuming an exogenous stopping criterion for the POMDP.

We use an MDP framework to model the maintenance stage of treatment. Based on a preformed belief about the patient's sensitivity, which results from the initiation stage, the physician prescribes the optimal dosage. We prove that the myopic policy is optimal for that MDP, so that it suffices to solve the single-stage problem at each decision epoch.

In this study, we determine optimal dosing decisions in the initiation and maintenance stages separately. An alternative approach is to jointly solve the problems in both stages. In a joint solution, the physician makes optimal dosing decisions which minimize the total expected discounted risk, cumulative over both stages, up to and beyond the stopping criterion (for belief updating) in the initiation stage. We opt against such a joint solution approach for two main reasons. First, our two-stage solution is in line with the realities of medical practice where initiation and maintenance are typically treated independently because they involve different visit frequencies, risks, and clinical objectives. Second, commercially available software for solving POMDP's does not allow for such a joint solution where there is no belief updating beyond a certain, pre-specified, point (end of initiation stage).

In general, a joint solution should typically lead to a smaller cumulative expected risk compared to a separate solution. In the supplement, we numerically compare between the joint and separate solutions of the problem, and show that they do not differ greatly; as such, our two-stage solution approach is further justified.

Figure 1 Conceptual Framework



3. The Initiation Stage of Treatment

We formulate a discrete-time, finite-horizon, POMDP model to solve the initiation-stage problem. In this modeling framework, a single decision maker (the physician) minimizes the cumulative expected risk for the patient by prescribing appropriate warfarin dosages at successive decision epochs, that is, at clinical visits. A POMDP model is appropriate because the physician cannot directly observe patient sensitivity to warfarin, that is, the states of the process are partially hidden. In this section, we first present the POMDP model, then we describe both our dose–response model and the belief updating procedure.

3.1. POMDP Model Formulation

In order to behave optimally in a partially observable setting such as ours, it is necessary to use the history of the treatment to aid in the disambiguation of patient sensitivity. Based on previous dose prescriptions and corresponding INR measurements, the physician maintains and updates a probability distribution over the set of possible patient sensitivity values. This probability distribution represents the physician's current belief about patient sensitivity. Consecutive optimal dosages depend on sequentially updated beliefs. Our POMDP framework constitutes a systematic way of doing that. Here is some notation that we use.

Decision Epochs: $t = 1,2,\ldots,T$. Each decision epoch in the POMDP model corresponds to a patient visit to

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

the clinic, that is, $t = i$ corresponds to the ith visit, where $1 \leq i \leq T$. Consistent with medical practice, we assume that the times between successive visits are long enough to allow for doses to exert their full effects; for example, visits are scheduled every 2 or 3 days. We make no other assumptions about the times between successive decision epochs.

The time horizon, T, taken to be the time needed to adequately learn about patient sensitivity, is typically not known at the onset of treatment. Indeed, T depends both on the desired ensuing level of confidence about patient sensitivity and the specific characteristics of the patient at hand. In section 7, we use simulation to estimate the number of visits required to form different confidence levels (about sensitivity) for alternative patients.

State Space: S. We let $s_t \in S$ denote the state of the system at time t. In particular, we let s_t be the pair (INR_t, γ_t), where INR_t is the INR of the patient at t, and γ_t is the sensitivity of the patient at t. We assume that the physician measures INR_t but cannot observe γ_t. Instead, the physician must maintain and update a belief distribution about γ_t. We let γ_t be the slope of a linearly additive Gaussian dose–response model for the natural logarithm of INR_t; see section 3.2. In our model, we do not impose a finite state space. Instead, we assume that both INR_t and γ_t may take values in the set of non-negative real numbers. We take patient sensitivity to be the result of the patient's genetic information which does not vary with time. Thus, in our context, it is reasonable to make the assumption of a constant, time-invariant, γ_t; accordingly, we drop the time subscript, t, from its notation.

Action Space: A. We let $d_t \in A$ be the warfarin dose (in milligrams) prescribed by the physician at visiting epoch t. That is, $d_t \geq 0$.

POMDPs are computationally difficult to solve; see Braziunas (2003). Indeed, since the underlying states are not known with certainty, the decision maker has to base his decisions on belief states. Since a belief state assigns a given probability to every system state, there is a continuous number of possible belief states to encounter. Essentially, solving a POMDP amounts to solving a continuous state-space belief MDP.

To alleviate some of that computational burden, and to be consistent with medical practice, we discretize the action and state spaces, A and S, when numerically solving our POMDP in section 7. For simplicity, and to be consistent with our subsequent numerical results, we hereafter use notation which is consistent with having a discrete state space.

State-Transition Function: $P_S : S \times A \times S \to [0,1]$. We let $P_S(s,d,s')$ denote the probability that the next state is s' given that the current state is s and the current action taken is d. This probability is specified by our dose–response model; see section 3.2. (We assume time-homogeneous transition probabilities.) In a model with time-independent γ, we assume that transitions from a given state (INR_t, γ) to other states with different γ values are not possible, that is, $P_S((INR,\gamma),d,(INR',\gamma')) = 0$ if $\gamma \neq \gamma'$. In sections 6 and 7, we discretize states by replacing individual INR values by corresponding intervals.

Observation Space: O. We denote an observation made at time t by $o_t \in O$. Since we assume that the physician can only observe INR_t at t, we let $o_t = INR_t$.

Observation Probability Function: $P_O : S \times O \to [0,1]$. We let $P_O(s,o)$ denote the probability of making observation o, given that the current state is s. In our context, P_O has the following simplified form. Since we assume that the physician correctly measures the patient's INR at each visiting epoch, we must have that $P_O((INR,\gamma),INR') = 1$ if and only if $INR = INR'$ and $P_O((INR,\gamma),INR') = 0$ otherwise. Allowing for measurement errors in the INR, or for INR measurements outside scheduled clinical visits, for example, as would be the case with patient self-monitoring of their INR, are left as directions for future research.

Belief space: B. An element $b \in B$ is a probability vector over possible γ values. (To denote the belief at a particular state s, we will use the notation $b(s)$.) The policy component of a POMDP maps the physician's current belief state into an optimal dose. In section 3.3, we describe how the physician transitions between different belief states.

Risk function: $r : S \times A \to \mathbb{R}$, where $\mathbb{R}$ denotes the set of real numbers. This function gives the immediate risk, $r(s,d)$, associated with prescribing a dose d and ending up in state s; for a specification of this risk function in our case study, see section 5.

Our objective is to minimize the cumulative discounted expected risk over the horizon T. The solution of the POMDP can be carried out using dynamic programming techniques; see Bellman (1957). Let the value function V_j^* denote the optimal expected reward when there are j steps to go until T. Then, V_j^* satisfies the following Bellman's equation:

$$V_{j+1}^*(b) = \min_{d \in A} \left\{ R(b,d) + \theta \sum_{o \in O} p(b,d,o) V_j^*(\tau(b,d,o)) \right\}, \tag{1}$$

where θ is some discount factor (since the decision does not observe the underlying state, we do not sum over these). In Equation (1), $p(b,d,o)$ is the probability of making an observation o in the next epoch

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

given that the current action is d and the current belief is b, and $R(b,d)$ is the expected one-step risk given current belief b and action d. That is, $p(b,d,o)$ can be calculated using basic probability as follows:

$$p(b,d,o) = \sum_{s' \in S} P_O(s',o) P_S(s,d,s') b(s).$$

In Equation (1), τ is an update function that computes a new belief $b' = \tau(b,d,o)$ given the current belief b, the current dose decision d, and the future observation o; see section 3.3. Finally, in Equation (1), $R(b,d)$ can be calculated as:

$$R(b,d) = \sum_{s' \in S} \sum_{s \in S} r(s',d) P_S(s,d,s') b(s).$$

It was shown in Smallwood and Sondik (1973) that the value functions for POMDPs are finite, piecewise linear, and convex which greatly simplifies the solution of Equation (1). We will make use of that piecewise linear representation when solving our POMDP in section 7.

3.2. Dose–Response Model

We model the log-transformed INR using a linearly additive Gaussian model with several clinical and demographic fixed effects. In section 5, we fit this model to data.

Let $t_1^j < t_2^j < \ldots < t_{N^j}^j$ be the visiting epochs for patient j. Let N^j be the total number of visits for patient j. Let INR_i^j be the INR of patient j measured at visiting epoch t_i^j. Define $Y_i^j \equiv \ln(\text{INR}^j)$ where $\ln(\cdot)$ denotes the natural logarithm function. Let d_i^j be the dose prescription between t_{i-1}^j and t_i^j, resulting in INR_i^j. Let γ^j denote the sensitivity of patient j. Our model for the log-transformed INR is given by the following:

$$Y_i^j \equiv \ln(INR_i^j) = \sum_{k=1}^{K} v_k I_k^j + \gamma^j d_i^j + \epsilon_i^j = \beta + \gamma^j d_i^j + \epsilon_i^j, \quad (2)$$

where β is defined as $\beta \equiv \Sigma_{k=1}^K v_k I_k^j$ and $1 \le i \le N^j$. We let ϵ_i^j be independent and identically distributed (i.i.d.) normal random variables with mean 0 and variance σ_ϵ^2. We let I_k^j be an indicator random variable which equals 1 when fixed effect k is present for patient j. We define K as the total number of fixed effects. For example, if k corresponds to gender, then we let $I_k^j = 1$ when patient j is male, and 0 otherwise. We omit the superscript j when the specific patient index does not matter. In Table 1 of the supplement, we describe all fixed effects in Equation (2). The parameters v_k in Equation (2) and the variance parameter, σ_ϵ^2, need to be estimated from data.

Table 1 Partial Results for the Fixed Effects and Covariance Parameters Specified in Equation (2). Point estimates of model coefficients are shown with corresponding standard errors and *p*-values of tests for statistical significance. The numbers in parentheses indicate whether or not the effect is present (1) or not (0)

Category	Coefficient	SE	*p*-value
AODM (0)	0.914	0.11	<0.001
RH (0)	0.00175	0.025	0.95
HYP (0)	0.0384	0.025	0.13
MHV (0)	−0.339	0.083	<0.001
CVA (0)	−0.0696	0.027	0.011
ULC (0)	−0.01323	0.027	0.062
AGE IND (0)	−0.0325	0.023	0.15
Risk of fall medium (1)	0.0392	0.76	0.45
AC (0)	0.0309	0.031	0.33
Gender (Female)	0.0554	0.030	0.57
Warfarin dose	–	–	<0.001

In section 6, we use Equation (2) to compute state transition probabilities.

3.3. Belief Updates

The physician is uncertain about the value of γ, which he treats as random. Here, we assume that the physician's initial belief about γ is normally distributed with mean μ_0 and variance σ_0^2. In Bayesian parlance, this is the prior distribution. (We selected a normal prior for analytical convenience; however, we conducted numerical experiments using a lognormal prior instead and reached largely the same conclusions.) Let $Y_1, Y_2, \ldots, Y_n$ denote the history of log-transformed INR observations up to, and including, epoch n, and the parameters β, d_i for $1 \le i \le n$, and σ_ϵ^2 be defined as in Equation (2). Then, the following lemma holds.

LEMMA 1. *The updated belief distribution about γ at the nth decision epoch is normal with variance $\sigma_n^2 = (\frac{1}{\sigma_0^2} + \sum_{i=1}^n \frac{d_i^2}{\sigma_\epsilon^2})^{-1}$ and mean $\mu_n = \sigma_n^2(\sum_{i=1}^n d_i \frac{y_i - \beta}{\sigma_\epsilon^2} + \frac{\mu_0}{\sigma_0^2})$.*

PROOF. The proof readily follows from conditional multivariate Gaussian theory and the linearity assumption in Equation (2). Indeed, we can write the following vector equation:

$$(Y_1, \ldots, Y_n) = (d_1, d_2, \ldots, d_n) \cdot \gamma + (\beta, \beta, \ldots, \beta) + (\epsilon_1, \ldots, \epsilon_n),$$

where γ is normally distributed with mean μ_0 and variance σ_0^2, . The updated belief distribution about γ is the conditional distribution of γ given the observations $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$. This conditional distribution can be easily shown to be normal as well; applying equations (2.113)–(2.117) of Bishop (2006) yields the desired expressions for the mean

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

and variance of this conditional distribution, as stated in the lemma. □

It is insightful that the variance term in Lemma 1 decreases with increasing dosages. This indicates that at any decision epoch prior to T, a myopically suboptimal dose may be optimal for the POMDP since it may lead to faster learning about γ (reduction in the belief distribution's variance). In other words, the optimal solution for the POMDP balances, at every epoch, risk minimization with faster learning about γ. This is known as the "exploitation vs. exploration" trade-off in reinforcement learning.

4. The Maintenance Stage of Treatment

In the maintenance stage, the physician assumes prior knowledge of his patient's sensitivity, with some degree of uncertainty, and prescribes warfarin doses accordingly. We model the maintenance stage of treatment using a discrete-time, finite-horizon MDP framework. We begin by briefly describing our MDP model, and then derive some relevant analytical results.

4.1. MDP Model Formulation

State Space and decision epochs. We define the state of the MDP, at a given decision epoch, to be the observable INR. As such, the state of the process is fully observable which makes the MDP framework appropriate. The decision epochs correspond, as in the POMDP framework, to patient visiting epochs to the clinic.

State transition function. Transitions between system states are specified by the dose–response model in Equation (2) where, taking the physician's perspective, γ is now assumed to be normally distributed with a mean of μ_M and a variance of σ_M^2. That is, the log-transformed INR is assumed to be normally distributed with a mean equal to $\beta + d_i\mu_M$ and a variance equal to $d_i^2\sigma_M^2 + \sigma_\epsilon^2$. The distribution of γ at the onset of the maintenance stage results from belief updates during the initiation stage: μ_M and σ_M^2 are given by Lemma 1, and depend on the entire history of treatment during the initiation stage. We assume that the distribution of γ does not change during the maintenance stage, that is, consistently with medical practice, the physician's belief about patient sensitivity is invariant during that stage.

Action space, risk function, horizon, and objective. The action space and risk function for the MDP are as in section 3.1. The time horizon, T_M is set to 100 visits. Finally, the objective of the problem remains to determine appropriate dosages so as to minimize the total expected cumulative risk. Let H_j^* denote the optimal expected reward when there are j steps to go until T_M. Then, H_j^* satisfies the following Bellman's equation:

$$H_{j+1}^*(s) = \min_{d\in A}\left\{ r(s,d) + \theta \sum_{s'\in S} P_S(s,d,s')H_j^*(s') \right\}. \quad (3)$$

4.2 Optimality of the Myopic Policy

We now characterize the optimal policy for the MDP.

THEOREM 1. *The myopic, single-stage, policy is optimal for the MDP.*

PROOF. Our proof proceeds by induction on the number, n, of remaining epochs until T_M. It is not hard to see that the induction hypothesis holds initially: $n = 0$ corresponds to time T_M, at which point it is optimal to select the myopically optimal dose. To prove the inductive step, we assume that myopically optimal doses are optimal for the MDP for the final k decision epochs, that is, for all $n \le k$, where $k \le T_M$ is some nonnegative integer. We now show that the same must hold for $n = k + 1$. Let d_{T_M-k} be the selected dose when there are $n = k + 1$ decision epochs remaining. Since the distribution of γ is invariant during the maintenance stage, the future dynamics of the system and, in particular, the future risks are unaffected by d_{T_M-k}. Thus, given that future optimal doses are also myopically optimal, by the induction hypothesis, the same must also hold for d_{T_M-k}; indeed, there is no advantage in selecting a myopically subpotimal dose for d_{T_M-k}. This completes the inductive proof. □

The myopic, single-stage, problem is the same across all decisions epochs of the MDP. Thus, the solution to that problem, if it exists and is unique, is constant throughout the maintenance stage. This coincides with medical practice where it is common that a unique, stable, dose is prescribed to patients during the maintenance stage of treatment; for example, see Botton et al. (2011), The International Warfarin Pharmacogenetics Consortium (2009), Millican et al. (2007) and Witt et al. (2009). Next, we investigate sufficient conditions for the existence and uniqueness of that stable dose.

4.3. Stable Dose

Consistent with the literature, we assume that the stroke and bleeding risks are exponential functions of the INR; see Rosendaal et al. (1993). For exact functional forms, we rely on previous literature; for example, see Hylek et al. (2003).

We focus on *relative risks* of stroke and bleeding, as a function of the reported INR for the patient. The

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

relative risk for stroke (bleeding) is the ratio between the (estimated) probability of stroke for a given INR value, and the (estimated) probability of stroke for a baseline INR value lying in the desired interval (2, 3). Indeed, most current guidelines recommend an INR target of 2.5 (with the range of 2 and 3); see Oden et al. (2006).

Usually, estimates of probabilities of stroke and bleeding events are reported in the literature in units of incidence rates per 100 patient-years for a given INR level; for example, see Hylek et al. (2003), Hylek et al. (2007), and Patrick et al. (2009). More precisely, a cohort of patients is followed up for some period of time during warfarin treatment, usually a year, and the proportion of patients during that year who have experienced an adverse event, be it a stroke or a severe bleeding, for a given INR value, is reported.

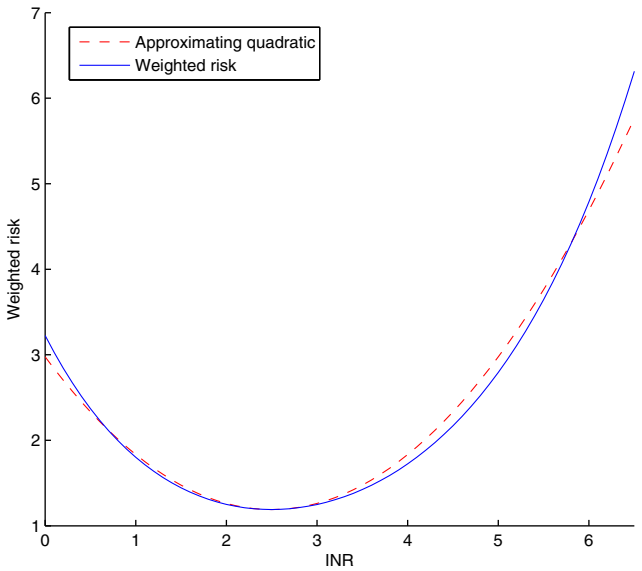
It is important to note that relative risks do not have time units, since they are ratios of probabilities, and that is why we assume that they are independent of time. We let the total risk associated with a given INR be equal to a convex combination of the corresponding stroke and bleeding risks. We choose the weights in this convex combination so that the risk is minimized for an INR value of 2.5. In the appendix (section A.2), we justify the additive nature of our risk criterion in solving the POMDP and MDP models.

In Figure 2, we plot the resulting risk function. For analytical tractability, we use an approximating quadratic function (dashed curve) as our risk function. Figure 2 shows that this is a good approximation for reasonable values of the INR. In the closed form, for a given INR value equal to x, where $x \geq 0$, the risk function $r(x)$ can then be given by

$$r(x) = a + b(x - c)^2; \tag{4}$$

in Figure 2, $a = 1.19$, $b = 2/7$, and $c = 2.5$.

Figure 2 Weighted Total Risk Function, Minimized at an INR of 2.5



The single-stage problem, to be solved at every decision epoch i of the MDP, is to minimize $E[r(e^{Y_i})]$ where $Y_i = \beta + \gamma d_i + \epsilon_i$ is normally distributed with a of mean $\beta + d_i\mu_M$ and a variance of $d_i^2\sigma_M^2 + \sigma_\epsilon^2$. That is, e^{Y_i} is lognormally distributed.

Given the quadratic form of the risk function in Equation (4), the myopically optimal dose is the one which minimizes $E[(e^{Y_i} - c)^2]$, that is, it minimizes $Var[e^{Y_i}] + (E[e^{Y_i}] - c)^2$. Hereafter, we drop dependence on i since the specific visiting epoch does not matter. By exploiting the properties of the lognormal distribution, our single-stage problem becomes:

$$\min_{d \geq 0} (e^{\sigma_M^2 d^2 + \sigma_\epsilon^2} - 1)(E[e^Y])^2 + (E[e^Y])^2 - 2cE[e^Y] + c^2,$$

where we can drop the constant term and simplify as follows:

$$\min_{d \geq 0} e^{\sigma_M^2 d^2 + \sigma_\epsilon^2}(E[e^Y])^2 - 2cE[e^Y]. \tag{5}$$

We can expand (5) using the fact that $E[e^Y] = e^{\beta + \mu_M d + 0.5(d^2\sigma_M^2 + \sigma_\epsilon^2)}$, and obtain:

$$\min_{d \geq 0} e^{2\beta + 2\mu_M d + 2\sigma_M^2 d^2 + 2\sigma_\epsilon^2} - 2ce^{\beta + \mu_M d + 0.5\sigma_M^2 d^2 + 0.5\sigma_\epsilon^2}.$$

In the following theorem, we investigate sufficient conditions guaranteeing the existence and uniqueness of a stable dose. Essentially, the lower bound in these conditions guarantees that the objective of our single-stage problem is strictly convex. We impose the upper bound to guarantee the existence of a local minimum, which must, by strict convexity, be both global and unique. We relegate the proof of the theorem to the appendix.

THEOREM 2. *A sufficient condition for the existence and uniqueness of the stable dose is the following:*

$$\max\{\frac{c}{2}, \frac{c\sigma_M^4}{3\sigma_M^4 + 4\sigma_M^2\mu_M^2 + \mu_M^4}, \frac{c\sigma_M^2}{\sigma_M^2 + \mu_M^2}\} < e^{\beta + 1.5\sigma_\epsilon^2} < c,$$

where μ_M and σ_M^2 are the physician's belief about the mean and variance of γ, respectively, at the beginning of the maintenance stage; β and σ_ϵ^2 are as in Equation (2).

4.4 An Alternative Criterion: Time in Therapeutic Range

The time in therapeutic range, TTR, is defined as the probability that the INR lies in the therapeutic range;

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

see Ansell et al. (2001). The higher the TTR, the more effective the treatment. In this section, we investigate how the TTR maximization and risk minimization criteria are different, in the maintenance stage, by computing the optimal dose under each criterion. This is important because those two criteria are often confused in practice.

In what follows, we assume an idealized scenario where patient sensitivity is known. (For numerical results comparing the two criteria when there is uncertainty about γ, see Table S2 in the supplement.) We denote patient sensitivity by γ_a. Under that assumption, we can derive a closed-form expression for the optimal dose under each criterion; see Lemma 2. The dynamics of the log-transformed INR are given by $Y = \beta + \gamma_a d + \epsilon$. The optimal dose under the TTR criterion maximizes the probability that e^Y lies in the therapeutic range; the optimal dose under the risk minimization criterion minimizes $E[(e^Y - c)^2]$.

LEMMA 2. *The stable dose which maximizes the TTR is:*

$$d^*_{TTR} = \frac{\ln(2) + \ln(3) - 2\beta}{2\gamma_a},$$

and the stable dose which minimizes the risk is:

$$d^*_R = \frac{1}{\gamma_a}(\ln(c) - \beta - 1.5\sigma^2_\epsilon),$$

for known patient sensitivity γ_a, and β and σ^2_ϵ as in Equation (2).

PROOF. Given the normality of Y, a closed-form expression for the TTR is given by

$$\begin{aligned}\text{TTR} &= P(Y \in (\ln(2), \ln(3))) \\ &= \Phi\left(\frac{\ln(3) - \beta - \gamma_a d}{\sigma_\epsilon}\right) - \Phi\left(\frac{\ln(2) - \beta - \gamma_a d}{\sigma_\epsilon}\right),\end{aligned} \tag{6}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a normal random variable with a mean equal to 0 and a variance equal to 1. Given Equation (6), we see that selecting d such that the origin is the midpoint of the interval $((\ln(2) - \beta - \gamma_a d)/\sigma_\epsilon, (\ln(3) - \beta - \gamma_a d)/\sigma_\epsilon)$ maximizes the TTR. This corresponds to an optimal value d^*_{TTR} as in the statement of the lemma.

Deriving the risk minimizing dose amounts to solving the problem

$$\min_{d \geq 0} e^{2\beta + 2\gamma_T d + 2\sigma^2_\epsilon} - 2ce^{\beta + \gamma_T d + 0.5\sigma^2_\epsilon}.$$

Upon differentiation of the objective function and setting the derivative to 0, we obtain that the unique minimizing dose is $d^*_R = (\ln(c) - \beta - 1.5\sigma^2_\epsilon)/\gamma_T$, as desired. □

Interestingly, the optimal dosages under the TTR maximization and risk minimization criteria can be quite different. For example, for a patient sensitivity equal to the average sensitivity in our data set, and for β and σ^2_ϵ as estimated from our data, we find that $d^*_{TTR} = 4$ mg (leading to a maximal value of the TTR equal to 47.5%) whereas $d^*_R = 2.34$ mg: such dosages typically lead to different patient responses in practice. In this section, we compared the TTR and risk criteria in the maintenance stage of the problem. In section 2 of the supplement, we include numerical results comparing the TTR and risk criteria in the initiation stage (by solving the POMDP under each criterion), and show that when there is uncertainty about γ, the difference in performance between those two criteria is generally not too great.

4.5. Adverse Events

The occurrence of a previous stroke is incorporated in the fixed effect term, β, in our dose–response model. Thus, in the event of a stroke, β should be updated. Consequently, the future dynamics of the INR and all subsequent risks (which are functions of the INR) are effected by that stroke event through the updated value of β. In section 7, we include a numerical study which quantifies the effect of ignoring a previous stroke on future dosages.

5. The Case Study

The data were gathered at the hematology clinic of the Jewish General Hospital in Montreal, Canada. They were collected over several years, ranging from January 3, 2000 to May 10, 2007. The data describe the anticoagulation treatment of 547 patients diagnosed with AF, and taking warfarin to reduce their risk of stroke. Each patient's treatment consists of successive visits to the clinic. At each visit, the patient's INR is measured through a blood test and, depending on this INR value, a new warfarin dose (in milligrams) is prescribed. In our data set, warfarin dose prescriptions are based solely on the physician's best judgment. As this is the common situation in practice, there is no systematic bias in the data resulting from the usage of a specific algorithm to determine warfarin dose recommendations.

5.1. Description of the Data Set

Our data set contains both demographic and clinical information for each patient. Demographic information consists of age and gender. Clinical information can be divided into two main categories. The first category describes the relevant past medical history of

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

the patient, that is, the presence of other diseases. The second category contains information relating to the bleeding risk of the patient. In Table 1 of the supplement, we specify all data variables.

We do not have any additional information about how patients manage their prescriptions. In the absence of such information, we assume that all doses are consistently taken by patients, as prescribed, throughout their treatments. The data set represents all patients at the hematology clinic who are receiving warfarin for stroke prevention, and who were active patients for given time period. As such, there is no sampling bias, and the data does not correspond to only one physician. In Figure 3, we plot INR measurements for three patients who received treatment at the clinic. Figure 3 illustrates the need to develop a more effective anticoagulation treatment. Indeed, Figure 3 shows that the INR for those patients consistently falls outside the desired target range, between 2 and 3. The patterns observed in Figure 3 are consistent across most patients.

The number of clinical visits depends on the patient at hand, and ranges from a single visit to a maximum of 29 visits. It is important to note that this maximum number of visits is due to the fact that our data collection is limited in time, and does not span the entire length of the patient's treatment. There are 17 patients who visited the clinic only once. We remove those patients from our data set since we cannot observe their INR responses to the prescribed dosages. For each patient, we do not know the last dose taken prior to the first recorded INR value (patients may have begun treatment earlier). Therefore, we remove all initial INR measurements from the data. There are also some missing values in our data set. For example, there are 10 patients with no gender listed. Additionally, there are missing INR values for six patients. Finally, different conventions for warfarin dose prescriptions are used, and some are hard to interpret. We remove all patients with missing or erroneous data, and are left with a total of 503 patients. For each patient, we compute the average daily warfarin dose prescribed at each visit, and use that as a covariate in our model.

5.2 Fitting to Data

We begin by investigating the validity of our normality assumption for the log-transformed INR, as described in section 3.2. In Figure 4, we present a QQ-plot for the residuals of the model in Equation (2). Figure 4 shows that the normality assumption is reasonable.

In Table 1, we present point estimates of some model parameters and statistical significance results, at the 95% confidence level, for selected fixed effects in Equation (2). The fixed effects that were found to be significant are: the warfarin dose, adult-onset-diabetes-mellitus (AODM), and previous cerebrovascular accident (CVA). The significance of those factors is consistent with the physician's intuition.

6. Estimates of POMDP Model Parameters

In this section, we estimate the parameters of the POMDP model, described in section 3, based on our

Figure 3 Plot of the INR for Three AF Patients Treated with Warfarin at the Clinic

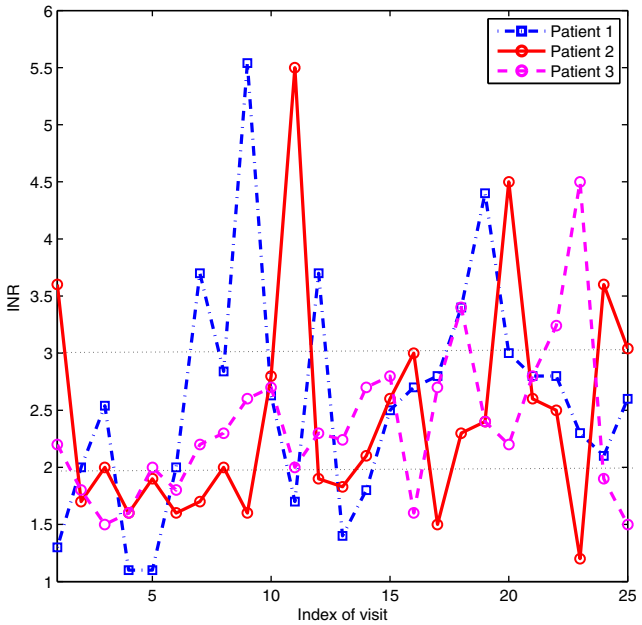
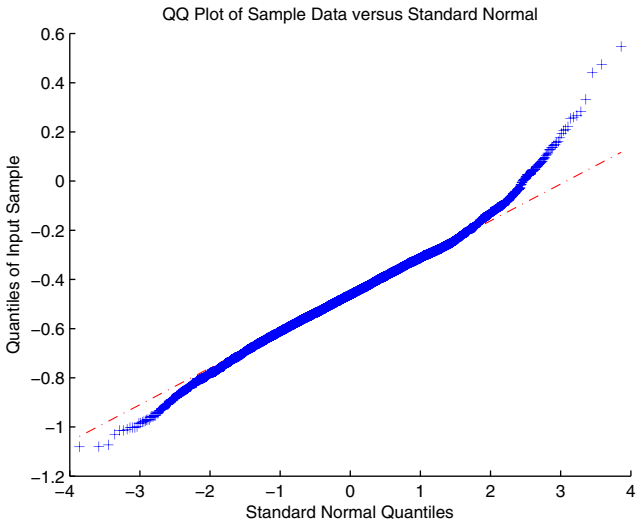


Figure 4 QQ-Plot of Residuals of the Dose–Response model



Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

data set. In section 7, we numerically solve the resulting POMDP and formulate several insights. Solving POMDPs is intractable in general, partly because the optimal policy may be infinitely large. To alleviate some of that computational burden, we discretize as follows.

State space. As explained in section 3, we define the state in our POMDP model, at a given decision epoch t, to be the pair (INR_t, γ). We fit our dose–response model in (3.2) to data, and estimate corresponding γ values for our patient population. We find that the average value of γ in our data set is equal to 0.08, its first quartile is equal to 0.03, and its third quartile equal to 0.13. Based on these estimates, we designate three levels of patient sensitivity: (1) Low ($\gamma = 0.05$), (2) Medium ($\gamma = 0.11$) and (3) High ($\gamma = 0.15$). Categorizing patients into three levels of sensitivity is often done by physicians in practice; see Moyer et al. (2009).

We discretize INR values into 5 different intervals. These intervals are: $I_1 = (0, 1.5]$, $I_2 = (1.5, 2]$, $I_3 = (2, 3]$, $I_4 = (3, 4]$, and $I_5 = [4, \infty)$. Discretizing the INR as such is consistent with medical practice; for example, see Jaffer and Bragg (2003). Since we have three values for γ and five intervals for the INR, the total number of states is $5 \times 3 = 15$ states.

Action space. We let the set of allowed warfarin doses coincide with the available concentrations of warfarin pills in practice; see Ansell et al. (2001). That is, we let d (in milligrams) be one of the following: 0, 2.5, 5, 7.5, and 10. A dose $d = 0$ corresponds to interrupting warfarin treatment until the next patient visit.

State-transition probabilities. For each dose d, we associate a 15×15 state-transition probability matrix, M_S^d. We let $M_S^d(s, s') = P_S(s, d, s')$ where, as explained in section 3, $P_S(s, d, s')$ denotes the probability that the next state is s' given that the current state is s and the current action taken is d. Since the transition from a given state to another state with a different patient sensitivity is not possible, M_S^d has a block diagonal matrix structure, where each block corresponds to a given patient's sensitivity level.

To compute the terms of each submatrix, we rely on the dynamics of our dose–response model in Equation (2). To illustrate, let the current dose be equal to d_0 and let γ_0 denote patient sensitivity. Then, ln (INR) is normally distributed with mean $\beta + \gamma_0 d_0$ and variance σ_ϵ^2, independently of the current INR level. Thus, the probability of transitioning into a certain INR interval can be computed based on the cumulative distribution function of a normal random variable. For example, the probability of transitioning into the interval (1.5,2] is given by $\Phi((\ln(2) - \beta - \gamma_0 d_0)/\sigma_\epsilon) - \Phi((\ln(1.5) - \beta - \gamma_0 d_0)/\sigma_\epsilon)$ where $\Phi(\cdot)$ is the cumulative distribution function of a (standard) normal random variable with mean 0 and variance 1.

Risk function. We use the risk function in Figure 2. In order to compute the immediate risk $r((I,\gamma),d)$ associated with transitioning into state (I,γ), given a current prescribed dose d, we assume that the new INR coincides with the midpoint of I. For example, the immediate risk associated with an INR transition into the interval (2,3) is $r(2.5) = 0$. We let $r(5)$ be the immediate risk associated with a transition into the interval $[4,\infty)$.

We let the time horizon T be long enough to allow convergence to a desired belief about patient sensitivity. The belief about patient sensitivity, in our discrete context, is a probability vector (p_L, p_M, p_H) of probabilities that γ is low, medium, or high, respectively. This probability vector is updated sequentially according to a simple application of Bayes' rule, given the observed INR values. The convergence criterion for our belief updating procedure is the final desired belief probability that γ is either low, medium, or high; we denote this convergence probability by p. In section 7, we vary the desired belief and find that an upper bound $T = 200$ guarantees convergence in each case. We let the discount factor $\theta = 0.8$.

Recognizing that $\theta = 0.8$ may be considered low in a clinical setting, we also conducted some of the experiments with a discount factor of 0.98. The results that we obtain are largely consistent. For the MDP, we proved that the myopic policy is optimal, so the discount factor has no effect at all. In the POMDP, we show, numerically, next that the myopic policy is roughly optimal, so that there is also little impact for the discount factor.

7. Numerical Study

In this section, we present numerical results corresponding to both the initiation and maintenance stages of treatment. For the numerical solution of the POMDP model, we rely on the widely used "pomdp-solve" software; see Cassandra (2014). Based on our results, we derive insights pertaining to the design of anticoagulation therapy in practice. We present additional numerical results and discuss related insights into the supplement.

7.1. The Initiation Stage

Near optimality of the myopic policy. We use simulation to compare the performances of the optimal solution for the POMDP and the myopically optimal solution. (We include detailed numerical results and a description of our simulation experiments in section S4.1 of the supplement.) We consider low, medium, and high sensitivity patients, alternative values for p (0.7, 0.9, and 0.995), and different initial

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

belief probability vectors. For each sensitivity level and each value of p, we report estimates of the average TTR, average risk value, and average number of visits required in the initiation stage to reach convergence.

Our results indicate that *the performance of the myopically optimal policy is close to that of the POMDP optimal policy*, given our discretization of the problem which is consistent with medical practice. Indeed, for all initial belief vectors and all patient sensitivities considered, the average TTR and risk values are close under those two policies. Thus, our study gives numerical support to solving the single-stage problem instead of the POMDP; this is practically useful, given the numerical complexity of the POMDP.

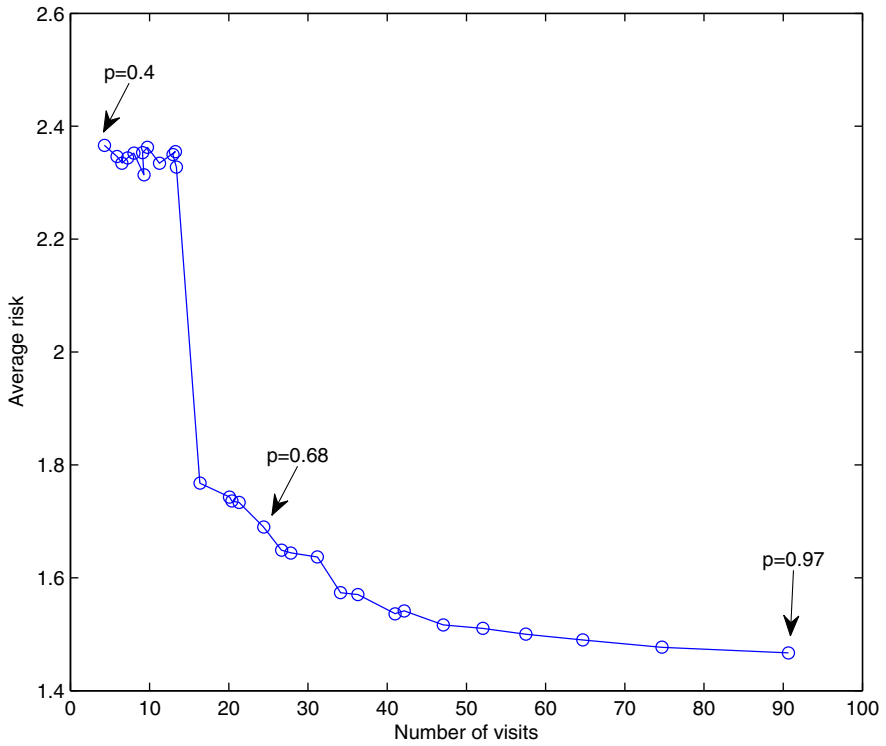
Length of the initiation stage. In Figures 5, 6 and 7, we report simulation estimates of the average risk as a function of the number of visits in the initiation stage, averaged over 10,000 independent simulation replications (we deliberately choose such a large number of replications to smoothe out the stochastic noise in the plots). In Figure 5, we consider a uniform initial belief vector and a medium sensitivity level. In Figure 6, we consider a uniform initial belief vector and a high sensitivity level. Finally, in Figure 7, we consider a high sensitivity level and an initial probability vector which assigns a probability equal to 0.6 to high sensitivity, and 0.2 to both medium or low sensitivities.

Figures 5, 6 and 7 show that even though smaller risk values are obtained for high values of p, as expected, the risks corresponding to smaller values of p need not be much larger. Therefore, *a relatively small number of visits for the initiation stage may be sufficient in practice*. For example, Figures 5 and 6 show that the greatest reduction in risk results from the initial 20 visits to the clinic. Beyond 20 visits (corresponding to $p = 0.54$ in Figure 5 and $p = 0.68$ in Figure 6), risk reduction is relatively marginal.

Such lengths for the initiation stage are reasonable particularly in light of recent technological advances which permit INR self-monitoring by patients. Indeed, self-monitoring reduces the number of required patient visits to the clinic, enables more frequent testing, and is recommended by both The National Institute for Health and Care Excellence (2014) and the United States Department of Health and Human Services (2014).

Effect of patient sensitivity on learning. Comparing Figures 5 and 6 shows that, for the same initial probability vector, *it usually takes longer to learn about a medium sensitivity, compared to a high sensitivity (and to a low sensitivity)*. Indeed, in both figures, we consider a uniform initial probability vector, but different sensitivity values (medium in Figure 5 and high in Figure 6). However, reaching $p = 0.4$ for a medium γ, that is, in Figure 4, requires around 15 visits, whereas it can be reached in around 5 visits for high γ, that is, in

Figure 5 Optimal Average Risk for a Uniform Initial Belief Vector and a Medium γ



Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

Figure 6 Optimal Average Risk for a Uniform Initial Belief Vector and a High γ

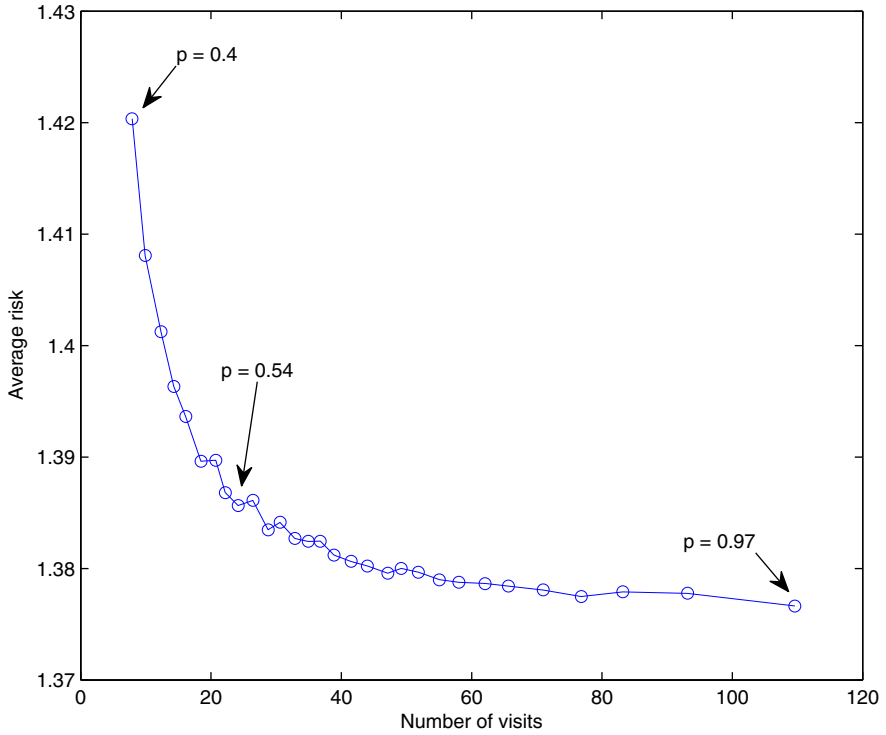


Figure 7 Optimal Average Risk for an Initial Belief Vector (low, med., high) = (0.2, 0.2, 0.6), and a high γ.

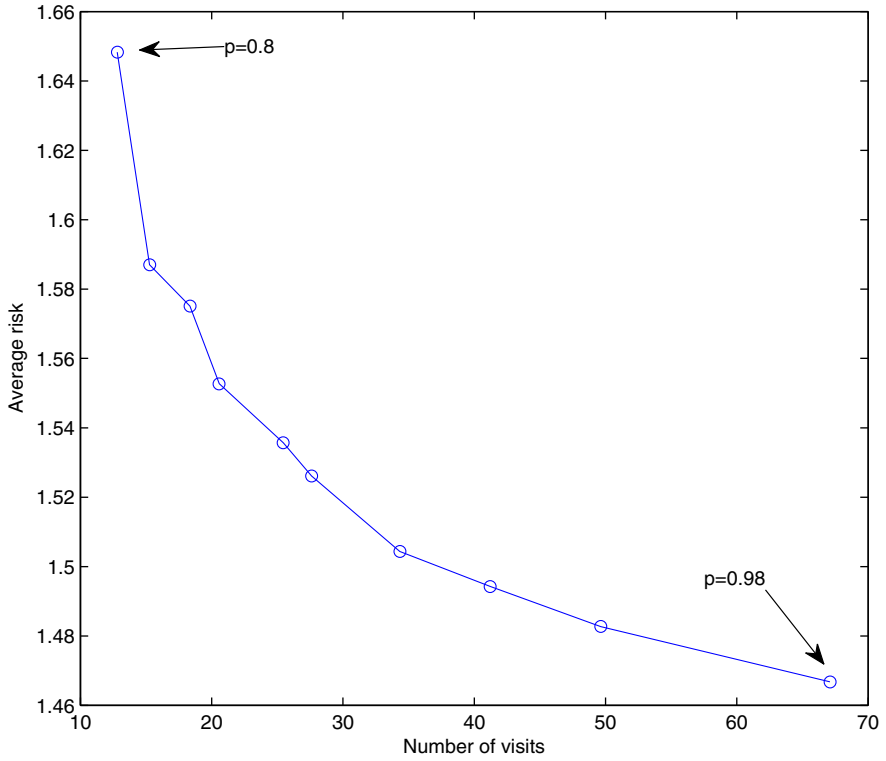


Figure 6. The intuition behind this observation is that patients with either low or high sensitivities are easier to identify due to their more extreme reactions to prescribed dosages. Also, Figures 5, 6, and 7 clearly show that *the required number of patient visits sharply increases as our desired final belief probability about γ increases*.

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

Effect of the initial belief vector. Comparing Figures 6 and 7 quantifies the advantage of selecting an adequate initial belief vector. Indeed, those figures assume the same (high) sensitivity but different initial belief vectors. Figure 7 shows that a high level of confidence, $p = 0.8$, is reached in less than 10 visits with a strong initial belief about a high sensitivity value. In contrast, a level of confidence $p = 0.65$ is reached after about 20 visits in Figure 6. *Our recommendation to physicians is to allocate a significant effort in pre-diagnosing the patient, for example, via preliminary medical tests or by soliciting additional patient information, to identify unique patient profiles and form a better initial belief about patient sensitivity; this would lead to substantially reducing the length of the initiation stage.*

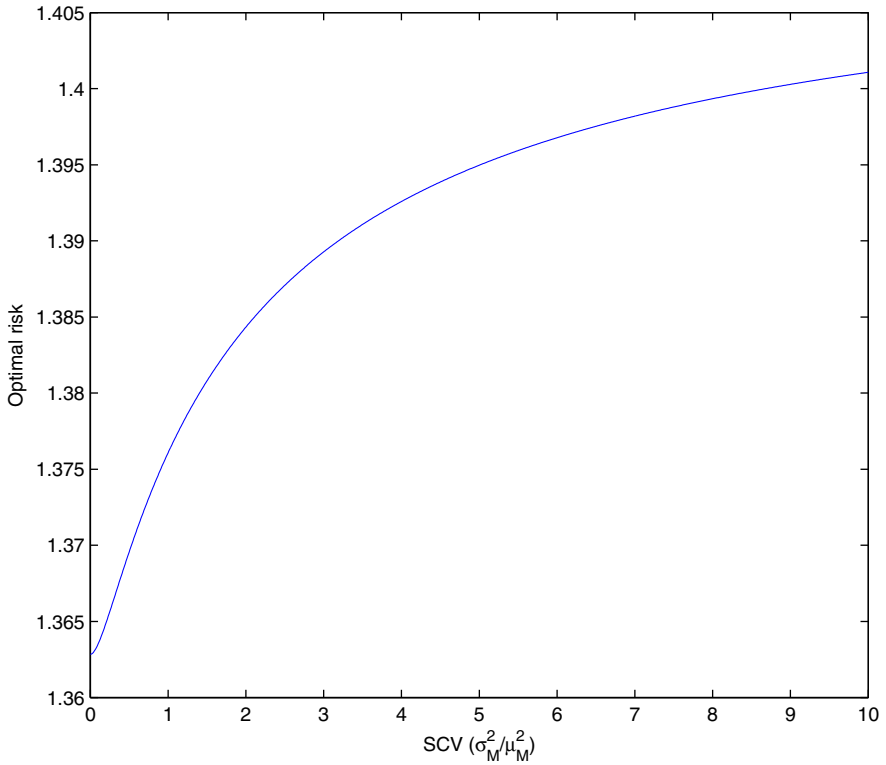
7.2. The Maintenance Stage

We now turn to the maintenance stage of treatment. Our objective is to demonstrate, numerically, the importance of effectively learning about γ during the initiation stage, and to quantify the effect of ignoring a previous stroke on future treatment. We assume that the physician has formed a belief about γ during the initiation stage; this belief remains unchanged during the maintenance stage. In particular, we assume that the belief about γ at the beginning of the maintenance stage is normally distributed with mean μ_M and variance σ_M^2. In this section, we plot and discuss corresponding figures; we include detailed numerical results in section S4.2 of the supplement.

Correct assessment of patient sensitivity. In Figure 8, we consider a medium sensitivity $\gamma_{true} = 0.11$, and assume that $\mu_M = \gamma_{true} = 0.11$. We then vary the squared coefficient of variation (SCV) of the belief about γ, $\text{SCV} \equiv \sigma_M^2/\mu_M^2$, and study the effect of this variation on the optimal risk in the maintenance stage (corresponding to the myopically optimal dose). Figure 8 shows that *the optimal risk in the maintenance stage increases as the uncertainty about γ increases*. In Figure 9, we assume that $\sigma_M^2 = 0$ and, once more, a medium sensitivity $\gamma_{true} = 0.11$. We vary the value of μ_M and find, as expected, that the lowest value of the optimal risk is achieved at $\mu_M = \gamma_{true} = 0.11$ (mean equal to true value). Interestingly, having $\mu_M = \gamma_{true}$ may not always be risk minimizing, as we show next.

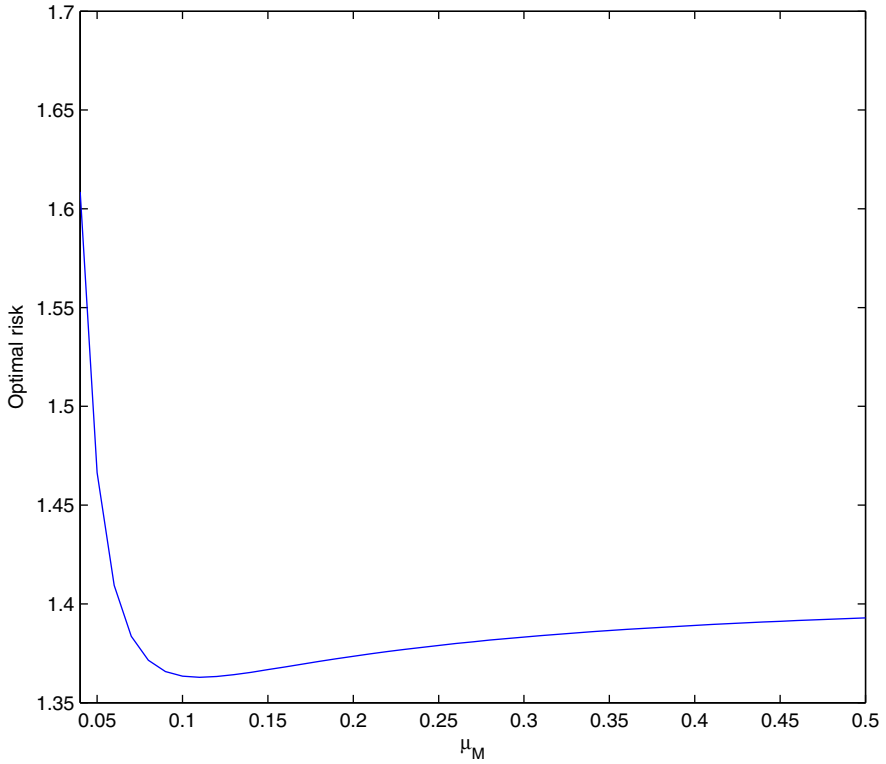
In Figure 10, we assume that the physician's belief about γ is uncertain. In particular, we assume that $\gamma_{true} = 0.11$ but $\sigma_M^2 = 5$. Figure 10 shows that with such a high σ_M^2 value, the optimal risk does not, surprisingly, correspond to $\mu_M = \gamma_{true} = 0.11$. Indeed, the optimal risk is monotone increasing in μ_M, so that underestimating sensitivity results in a smaller risk value. In other words, we can formulate the following recommendation: *If the physician remains highly uncertain about the true value of γ after the initiation stage (e.g.,*

Figure 8 Optimal Myopic Risk as a Function of the SCV for fixed $\mu_M = \gamma_{true} = 0.11$ and Alternating Values of σ_M^2



Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

Figure 9 Optimal Myopic Risk as a Function of μ_M for $\sigma_M^2 = 0$ and $\gamma_{true} = 0.11$



in exceptional cases where the patient state is unusually variable), then it may be risk minimizing to underestimate γ.

Effect of ignoring previous stroke events. We now study the effect of ignoring a previous stroke when treating a patient in the maintenance stage. For this, we assume a sensitivity level $\gamma = 0.12$ and compare the following three scenarios for different values of the SCV and of μ_M: (1) the patient does not experience any previous strokes; (2) the patient experiences a stroke and the physician updates the fixed effect β in Equation (2) accordingly after this stroke; and (3) the patient experiences a stroke and the physician does not update β. We present our results in Table 2. Table 2 shows that *risk can considerably increase by not taking into account the occurrence of a previous stroke event, and that this increase is particularly strong when the uncertainty, σ_M^2, about γ is low*. This is intuitively clear since in this case the physician is "confident" about a wrong model for the patient's INR. For example, for an SCV equal to 0.01 and $\mu_M = 0.02$, the optimal risk increases from 5.1 to 14.1 when ignoring a previous stroke in designing future treatment.

8. Concluding Remarks

In this study, we proposed an individualized design for the anticoagulation treatment of patients with AF. This heart condition puts patients at an increased risk of occurrence of a stroke. Patients with AF are typically prescribed warfarin, a popular anticoagulant drug, as a preventive measure. Despite its popularity, warfarin is notoriously difficult to manage.

We modeled the initiation stage of treatment using a POMDP model (section 3). In the POMDP modeling framework, we proposed a systematic way to learn about patient sensitivity, γ, which strongly influences patient response to warfarin and is typically unobservable.

We modeled the maintenance stage of treatment, using an MDP model (section 4). In the MDP modeling framework, we assumed some pre-formed belief about γ and derived several analytical results. We analyzed data provided by a hematology clinic in Montreal (section 5). Through direct analysis and numerical experiments, we investigated issues faced by doctors in practice. For the initiation stage of treatment, we derived a closed-form expression for belief updates in the POMDP framework (section 3.3). We showed that the complexity of the POMDP lies in balancing risk minimization with faster learning about γ, at each decision epoch. In section 6, we fit our POMDP model to data. In section 7, we discretized our problem, consistently with medical practice, and presented simulation results based on the numerical solution of the POMDP. Our experiments showed that the performance of the myopically optimal policy is close, in terms of both average risk and TTR, to the optimal POMDP policy.

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

Figure 10 Optimal Myopic Risk as a Function of μ_M for $\sigma_M^2 = 5$ and $\gamma_{true} = 0.11$

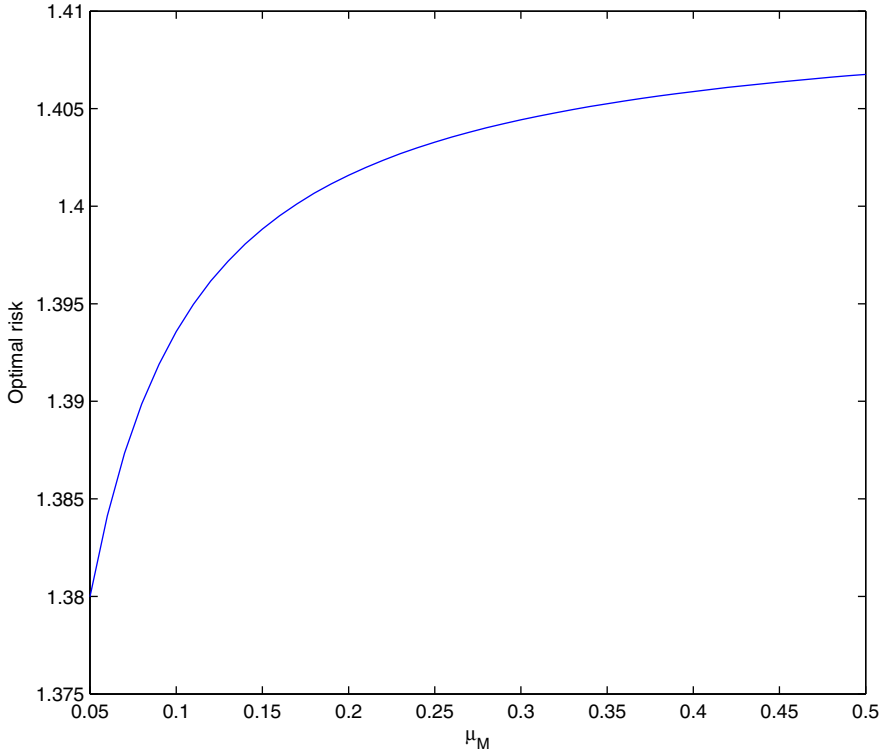


Table 2 Expected Risk Values in the Maintenance Stage under Different Scenarios and $\gamma = 0.12$. We include in parentheses the optimal dosages required to minimize the risk criterion

SCV for γ \ μ_M	0.02	0.04	0.08	0.1	0.12
			No previous stroke		
0.01	11.9 (12.7)	2.06 (6.4)	1.39 (3.2)	1.37 (2.5)	1.37 (2.1)
0.5	3.22 (8.2)	1.48 (4.1)	1.36 (2.1)	1.37 (1.6)	1.38 (1.4)
5	1.37 (2.6)	1.38 (1.3)	1.41 (0.64)	1.41 (0.52)	1.42 (0.43)
			Previous stroke taken into account		
0.01	5.13 (9.3)	1.69 (4.7)	1.38 (2.3)	1.36 (1.9)	1.36 (1.5)
0.5	2.19 (6.0)	1.42 (3.0)	1.36 (1.5)	1.37 (1.2)	1.37 (1.0)
5	1.36 (1.8)	1.37 (0.90)	1.39 (0.45)	1.39 (0.36)	1.40 (0.30)
			Previous stroke ignored, that is, use same dosages as without a stroke		
0.01	14.1	2.34	1.44	1.39	1.37
0.5	3.77	1.57	1.37	1.36	1.36
5	1.39	1.36	1.38	1.39	1.39

We proved that the myopic policy is optimal for the MDP model (section 4.2). This shows that it suffices to solve the myopic problem, at every decision epoch, during the maintenance stage. Moreover, there is a unique optimal dose for all decision epochs during that stage. This is consistent with medical practice where such a dose is called the stable dose.

We formulated sufficient conditions for the existence and uniqueness of that stable, risk-minimizing, dose for the MDP model (section 4.3). We also analyzed how that dose affects other clinical measures commonly used in medical practice, such as the TTR (section 4.4). We showed that doses which are optimal under the TTR maximization criterion may be very different from doses which are optimal under the risk minimization criterion. This is insightful since those two criteria are often assumed to be equivalent in medical practice.

Numerically, we investigated the required length of the initiation stage, and showed that the greatest decrease in risk typically results from the initial visits to the clinic. This substantiates that it suffices to have a short initiation stage in practice. We also studied how that length is affected by the physician's initial belief about patient sensitivity. Consistent with intuition, we found that the length of the initiation stage

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

increases sharply as the desired level of accuracy about γ increases. We also found that the initial belief about γ strongly impacts the required length for the initiation stage.

At the onset of the maintenance stage, the physician has already formed a reliable belief about patient sensitivity. We showed the importance of forming a correct belief about γ at the beginning of the maintenance stage, and showed that if the physician is uncertain about the value of patient sensitivity, then it may be best to underestimate it. We also studied the effect of ignoring a previous stroke on subsequent dosing decisions and risks.

In both our POMDP and MDP frameworks, we took the physician's perspective and excluded costs that patients may incur upon each visit to the clinic. A natural extension of our models would be to include both visiting costs and the effect of time between successive visits. Then, the decision to be made at each epoch concerns both the dosage and the time until the next visit. It would be interesting to investigate structural properties of the solutions of our decision-making problems in that case.

In this study, we assumed that patients take their doses as prescribed, since we did not have any additional information about how patients manage their prescriptions. It is important to try to gather such data in the future, in order to develop more reliable models. Finally, there is a need to conduct a randomized trial to test the effectiveness of our proposed treatment procedure, and to compare it to treatments that are currently used in practice.

References

Alagoz, O., L. M. Maillart, A. J. Schaefer, M. S. Roberts. 2007. Choosing among living-donor and cadaveric livers. *Management Sci.* **53**(11): 1702–1715.

American Heart Association. 2014. New guidelines for the primary stroke prevention: A closer step toward personalized medicine. Available at http://my.americanheart.org/ (accessed April 6, 2015).

American Stroke Association. 2015. Anti-clotting agents explained. Available at http://www.strokeassociation.org/ (accessed April 6, 2015).

Ansell, J., J. Hirsh, J. Dalen, 2001. Managing oral anticoagulant therapy. *Chest* **119**(1): 22S–38S.

Attaya, S., T. Bornstein, N. Ronquillo, R. Volgman, L. T. Braun, R. Trohman, A. Volgman. 2012. Study of warfarin patients investigating attitudes toward therapy change (SWITCH Survey). *Am. J. Ther.* **19**(6): 432–435.

Ayer, T., O. Alagoz, N. K. Stout. 2012. A POMDP approach to personalize mammography screening decisions. *Oper. Res.* **60** (5): 1017–1021.

Ayvaci, M. U., O. Alagoz, E. S. Burnside. 2012. The effect of budgetary restrictions on breast cancer diagnostic decisions. *Manuf. Serv. Oper. Manag.* **14**(4): 600–617.

Bellman, R. 1957. *Dynamic Programming*. Princeton University Press, Princeton, NJ.

Bhatt, S. 2014. Update on genetic testing and warfarin. *Curr. Emerg. Hosp. Med. Rep.* **2**(3): 133–137.

Bishop, C. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ.

Botton, M., E. Bandinelli, L. Rohde, L. Amon, M. Hutz. 2011. Influence of genetic, biological and pharmacological factors on warfarin dose in a Southern Brazilian population of European ancestry. *Br. J. Clin. Pharmacol.* **72**(3): 442–450.

Braziunas, D. 2003. Pomdp solution methods. Technical report, University of Toronto.

Cassandra, T. 2014. Partially observable Markov decision processes. Available at http://www.pomdp.org/ (accessed April 16, 2014).

Chhatwal J., O. Alagoz, E. S. Burnside. 2010. Optimal breast biopsy decision-making based on mammographic features and demographic factors. *Oper. Res.* **58**(6): 1577–1591.

Coroian, D. and K. Hauser. 2015. Learning stroke treatment progression models for an MDP clinical decision support system, in SIAM Int'l Conference on Data Mining, April 2015.

Denton, B. T., O. Alagoz, A. Holder, E. K. Lee. 2011. Medical decision making: Open research challenges. *IIE Trans. Healthc. Syst. Eng.* **1**(3): 161–167.

Ginsburg, G. and D. Voora. 2010. The long and winding road to warfarin pharmacogenetic testing. *J. Am. Coll. Cardiol.* **55**(25): 2813–2815.

Goulionis, J. E., A. Vozikis. 2009. Medical decision making for patients with Parkinson disease under average cost criterion. *Aust. New Zealand Health Policy* **6**(1): 15.

Hamburg, M. A., F. S. Collins. 2010. The path to personalized medicine. *N. Engl. J. Med.* **363**(4): 301–304.

Hauskrecht, M., H. Fraser. 2000. Planning treatment of ischemic heart disease with partially observable Markov decision processes. *Artif. Intell. Med.* **18**(3): 221–244.

He, M., L. Zhap, W. B. Powell. 2010. Optimal control of dosage decisions in controlled ovarian hyperstimulation. *Ann. Oper. Res.* **178**(1): 223–245.

Holbrook, A., S. Schulman, D. M. Witt, P. O. Vandvik, J. Fish, M. J. Kovacs, P. J. Svensson, D. L. Veenstra, M. Crowther, G. H. Guyatt. 2012. Evidence-based management of anticoagulant therapy: Antithrombotic therapy and prevention of thrombosis: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest J.* **141**(Suppl. 2): e152S–e184S.

Hwang, J. M., J. Clemente, K. P. Sharma, T. N. Taylor, C. L. Garwood. 2011. Transportation cost of anticoagulation clinic visits in an urban setting. *J. Manag. Care Pharm.* **17**(8): 635–640.

Hylek, E. M., A. S. Go, Y. Chang, N. G. Jensvold, L. E. Henault, J. V. Selby, D. E. Singer. 2003. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. *N. Engl. J. Med.* **349**(11): 1019–1026.

Hylek, E. M., C. Evans-Molina, C. Shea, L. E. Henault, S. Regan. 2007. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation* **115**(21): 2689–2696.

Institute for Safe Medication Practices. 2012. Leading drug safety issues of 2012. Internet Document: [21 pages]. Available at http://www.ismp.org/quarterwatch (accessed October 17, 2013).

Jaffer A., L. Bragg. 2003. Practical tips for warfarin dosing and monitoring. *Clevel. Clin. J. Med.* **70**(4): 361–370.

Leshno, M., Z. Halpern, N. Arber. 2003. Cost-effectiveness of colorectal cancer screening in the average risk population. *Health Care Manag. Sci.* **6**(3): 165–174.

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514

Maillart, L., J. S. Ivy, S. Ransom, K. Diehl. 2008. Assessing dynamic breast cancer screening policies. *Oper. Res.* **56**(6): 1411–1427.

Millican, E., P. Lenzini, P. Milligan, L. Grosso, C. Eby, E. Deych, G. Grice, J. Clohisy, R. Barrack, S. Burnett, D. Voora, S. Gatchel, A. Tiemeier, B. Gage. 2007. Genetic-based dosing in orthopedic patients beginning warfarin therapy. *Blood* **110**(5): 1511–1515.

Moyer, T. P., D. J. Okane, L. M. Baudhuin. 2009. Warfarin sensitivity genotyping: A review of the literature and summary of patient experience. *Mayo Clin. Proc.* **84**(12): 1079–1094.

Oden A, M. Fahlen, R. G. Hart. 2006. Optimal INR for prevention of stroke and death in the atrial fibrillation: A critical appraisal. *Thromb. Res.* **117**(5): 493–499.

Oldgren, O., M. Alings, H. Darius, H. C. Diener, J. Eikelboom, M. D. Ezekowitz, G. Kamensky, P. A. Reilly, S. Yang, S. Yusuf, L. Wallentin, S. J. Connolly. 2011. Risks for stroke, bleeding, and death in patients with atrial fibrillation receiving dabigatran or warfarin in relation to the CHADS2 score: A subgroup analysis of the RE-LY trial. *Ann. Intern. Med.* **155**(10): 660–667.

Patrick, A. R., J. Avorn, N. K. Choudhry. 2009. Cost-effectiveness of genotype-guided warfarin dosing for patients with atrial fibrillation. *Circ. Cardiovasc. Qual. Outcomes* **2**(5): 429–436.

Personalized Medicine Coalition. 2015. Annual 2012 report: Change agent. Available at http://www.personalizedmedicinecoalition.org/ (accessed April 7, 2015).

Rosendaal F. R., S. C. Cannegieter, F. J. M. van der Meer, E. Briet. 1993. A method to determine the optimal intensity of Oral Anticoagulant Therapy. *Thromb. Haemot.* **69**(3): 236–239.

Schaefer, A., M. Bailey, S. Shechter, M. Roberts. 2004. Modeling medical treatment using markov decision processes. M. L. Brandeau, F Sainfort, W. P. Pierskalla, eds. *Operations Research and Health Care: A Handbook of Methods and Applications*. Kluwer, Boston, MA, 593–612.

Shechter, S., M. Bailey, A. J. Shaefer, M. S. Roberts. 2008. The optimal time to initiate HIV therapy under ordered health states. *Oper. Res.* **56**(1): 20–33.

Smallwood, R., E. Sondik. 1973. Optimal control of partially observable processes over a finite horizon. *Oper. Res.* **21**(5): 1071–1088.

The International Warfarin Pharmacogenetics Consortium. 2009. Estimation of the warfarin dose with clinical and pharmacogenetic data. *New Engl. J. Med.* **360**(8): 753–764.

The National Institute for Health and Care Excellence. 2014. Available at http://www.pulsetoday.co.uk/ [Retrieved March 30, 2015].

United States Department of Health and Human Services. 2014. Available at: http://www.guideline.gov/content.aspx?id=48562. [Retrieved March 30, 2014].

White, P. J. 2010. Patient factors that influence warfarin dose response. *J. Pharm. Pract.* **23**(3): 194–204.

Witt, D., T. Delate, N. Clark, C. Martell, T. Tran, M. Crowther, D. Garcia, W. Ageno, E. Hylek. 2009. Outcomes and predictors of very stable INR control during chronic anticoagulation therapy. *Blood* **114**(5): 952–956.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1: Personalized anticoagulation therapy.

Please Cite this article in press as: Ibrahim, R., et al. Designing Personalized Treatment: An Application to Anticoagulation Therapy. *Production and Operations Management* (2015), doi 10.1111/poms.12514