Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Dynamic Forecasting and Control Algorithms of Glaucoma Progression for Clinician Decision Support

Jonathan E. Helm, Mariel S. Lavieri, Mark P. Van Oyen, Joshua D. Stein, David C. Musch

To cite this article:
Jonathan E. Helm, Mariel S. Lavieri, Mark P. Van Oyen, Joshua D. Stein, David C. Musch (2015) Dynamic Forecasting and Control Algorithms of Glaucoma Progression for Clinician Decision Support. Operations Research 63(5):979-999. https://doi.org/10.1287/opre.2015.1405

OPERATIONS RESEARCH
Vol. 63, No. 5, September–October 2015, pp. 979–999
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.2015.1405

CROSSCUTTING AREAS

Dynamic Forecasting and Control Algorithms of Glaucoma Progression for Clinician Decision Support

Jonathan E. Helm
Operations and Decision Technologies, Kelley School of Business, Indiana University, Bloomington, Indiana 47405,
helmj@indiana.edu

Mariel S. Lavieri, Mark P. Van Oyen
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48109
{lavieri@umich.edu, vanoyen@umich.edu}

Joshua D. Stein, David C. Musch
Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan 48105
{jdstein@umich.edu, dmusch@umich.edu}

In managing chronic diseases such as glaucoma, the timing of periodic examinations is crucial, as it may significantly impact patients' outcomes. We address the question of when to monitor a glaucoma patient by integrating a dynamic, stochastic state space system model of disease evolution with novel optimization approaches to predict the likelihood of progression at any future time. Information about each patient's disease state is learned sequentially through a series of noisy medical tests. This information is used to determine the best time to next test based on each patient's individual disease trajectory as well as population information. We develop closed-form solutions and study structural properties of our algorithm. While some have proposed that fixed-interval monitoring can be improved upon, our methodology validates a sophisticated model-based approach to doing so. Based on data from two large-scale, 10+ years clinical trials, we show that our methods significantly outperform fixed-interval schedules and age-based threshold policies by achieving greater accuracy of identifying progression with fewer examinations. Although this work is motivated by our collaboration with glaucoma specialists, the methodology developed is applicable to a variety of chronic diseases.

Keywords: linear Gaussian systems modeling; controlled observations; stochastic control; disease monitoring; medical decision making; glaucoma; visual field.
Subject classifications: healthcare: diagnosis, treatment; decision analysis: applications.
Area of review: Policy Modeling and Public Sector OR.
History: Received October 2011; revisions received March 2013, February 2014, November 2014, March 2015; accepted May 2015. Published online in *Articles in Advance* September 9, 2015.

1. Introduction

Glaucoma is a leading cause of visual impairment in the United States and worldwide. It is estimated that over 2.2 million Americans have glaucoma, and the number is expected to grow to more than 3 million by 2020 (see Friedman et al. 2004, Quigley and Broman 2006). Glaucoma is often asymptomatic early in the course of the disease, but if left untreated, it leads to gradual and progressive loss of vision, ultimately resulting in irreversible blindness. Early identification of progression and appropriate treatment can slow or halt the rate of vision loss (see NEI 2011).

Patients suffering from glaucoma are monitored periodically via noisy quantitative tests to determine whether the disease is stable or a change in treatment is warranted to slow glaucoma-related vision loss. There is often a clear trade-off between monitoring intervals that are too short (little information is gained between readings, and there is unnecessary cost and undue discomfort and/or anxiety for the patients) and too long (the patient's long-term outcomes may be affected adversely by the delay in detecting disease progression). However, no consensus exists as to the optimal frequency by which testing should take place, and the ideal frequency of testing can vary from patient to patient (see Jansonius 2006, 2007). Multiple factors (including age, family history, race, intraocular pressure levels, visual field variables, type 2 diabetes mellitus, medical history, and genetic factors among others) may affect the initial onset of the disease and its progression (Tielsch et al. 1990). With the movement toward patient-centered models of care (see Bensing 2000), monitoring guidelines that incorporate information from the patient's history are needed.

The standard for glaucoma care is to periodically measure intraocular pressure (IOP) (see Lee et al. 2007b, Musch et al. 2008) and peripheral vision, as captured by visual field (VF) testing (see Bengtsson et al. 2009, Diaz-Aleman et al. 2009, McNaught et al. 1995, Zahari et al. 2006) to determine if and when an intervention should be

performed to slow glaucoma-related vision loss. The IOP test measures the fluid pressure in the eye. A high IOP is an important risk factor that can lead to damage of the optic nerve and loss of peripheral vision. The automated VF test examines the sensitivity of the eye to light stimuli, which is a way of quantifying peripheral vision loss. Standard automated VF tests provide a quantitative metric on sensitivity to light throughout the field of vision, as well as a number of global indices comparing the patient's test performance to that of a healthy individual with no glaucoma (see Choplin and Edwards 1999). Two of the VF performance indices commonly used in clinical practice are mean deviation (MD) and pattern standard deviation (PSD). Testing noise is associated with both IOP readings and VF test results. During the VF test, patients can get nervous or tired, which can lead to false positive and false negative responses. Moreover, patients may experience fixation loss, which introduces error into test results. The VF test can be long, uncomfortable, and burdensome, particularly for elderly patients (see Gardiner and Demirel 2008). There is a clear trade-off understood by the Glaucoma provider community between monitoring intervals that are either too short (high cost and unnecessary discomfort) or too long (disease progression goes undetected) (see Glen et al. 2014; Lee et al. 2006, 2007a). Subject to the judgment and expertise of eye care providers, the frequency with which patients undergo testing may be as infrequent as every two years (see American Academy of Ophthalmology Glaucoma Panel 2010). This frequency depends on a variety of factors including disease severity and stability of the disease. The expense of conducting these tests can be significant for both the patients and the overall U.S. healthcare system (see Lee et al. 2006, Rein et al. 2006, Alliance for Aging Research 2011).

In addition to using data from perimetry (VF) and IOP to assess for glaucoma progression, there are also structural tests that assess for pathology to the optic nerve and retinal nerve fiber layer (e.g., optical coherence tomography (OCT), confocal scanning laser ophthalmoscopy, and scanning laser polarimetry). While these tests are becoming increasingly useful in clinical practice (see Schuman et al. 2012), unfortunately these tests were not commercially available when the clinical trials, on which our analysis is based, were carried out. Fortunately, our research models are scalable and will be able to accommodate data from structural testing in the future.

1.1. The Disease Monitoring Problem

Motivated by the nature of chronic disease management, this research explores solutions to the disease monitoring problem. The monitoring problem that we treat in this paper is distinctly different from disease screening and detection. Screening models serve to detect or rule out whether or not a person has a disease based on disease prevalence and possibly transmission models. The monitoring problem that we pursue in this paper focuses on the need to perform a series of ongoing tests over time to promptly identify time epochs at which patients who already have a disease are experiencing a progression/worsening of the disease. In contrast with screening problems, disease monitoring involves (1) tracking individual patients over time (rather than population level modeling), (2) gaining new and rich information about an evolving disease state with each test (as opposed to the yes/no result of a screening test), and (3) dynamic decisions of when to take tests based on the history of information learned about the patient up to the current time point. This class of problems poses different modeling challenges than the screening problem and opens the way for new operations research methods that have potential to positively impact longitudinal patient care.

Table 1. A description of our modeling paradigm and contributions to theory and clinical practice.

	Controlled Gaussian state space modeling approach
State space scalability	High
Stochasticity	Separate system noise and measurement noise
Patient centered model	Feedback driven; learns each patient's unique disease dynamics
Clinician interactive	Clinician can tailor model to each patient's needs
Solution approach	Closed-form solution enables techniques and real-time decision support
Generalizability to other diseases	High

In this paper we develop models and methods for determining the appropriate timing of monitoring tests based on the control of a linear Gaussian system for disease progression that is customizable for each patient. Table 1 summarizes some of the interesting and powerful features of our modeling approach.

In clinical practice, physicians monitor many chronic diseases by administering a set of quantifiable tests to gain information about a patient's disease state (such as VF and IOP). One- or two-dimensional state spaces are often insufficient to incorporate the richness of data involved in clinical decision making. This causes a problem for paradigms such as Markov decision process (MDP), which suffer from the curse of dimensionality. While approximate dynamic programming (ADP) techniques can deal with large state spaces, the need to incorporate noisy observations and the need for a continuous state makes the problem even more challenging. Continuous state space models characterized as first-order vector difference equations and white multivariate Gaussian noise inputs can easily accommodate large state spaces, noisy data, and rich data inputs. To capture a wide range of dynamic behavior, we include in the state not only a test measurement itself but its most important derivatives (e.g., first, second). The inclusion of variable derivatives in the state represents a departure from traditional disease modeling. The inclusion of an nth derivative

makes it possible to capture the *n*th order dynamics in the "position" variable, thereby allowing a first-order model to capture nonlinear behavior in key variables.

As with glaucoma, chronic disease monitoring typically involves both system noise (e.g., stochastic disease evolution) and measurement noise (e.g., testing errors). Capturing both types of noise is difficult in some stochastic control paradigms, yet these noise components are critical to capturing the true dynamics of chronic disease care. Our approach captures correlated multivariate Gaussian white noise that is present in many medical tests, including VF tests and IOP tests.

Our modeling framework is, to our knowledge, one of the first patient-centered decision support mechanisms for glaucoma monitoring. The model is feedback driven, which means that it learns about each patient's unique disease evolution dynamics via better state estimation as the clinician receives more test results, allowing our algorithm to fit the policy to the specific individual's disease. Another important feature that is gaining increasing attention in the clinical community is that two patients can experience the same symptoms very differently, so monitoring schedules should be tailored to the patient's experience of the symptoms and not just to the symptoms themselves (see Fowler et al. 1988). To avoid a "one size fits all" policy, we allow the clinician to adjust the algorithm based on three levels of aggressiveness to tailor the monitoring schedule to each individual patient's needs. For example, a clinician would likely prescribe a different treatment approach for an elderly patient with comorbidities versus a young, healthy patient with the same level of glaucoma.

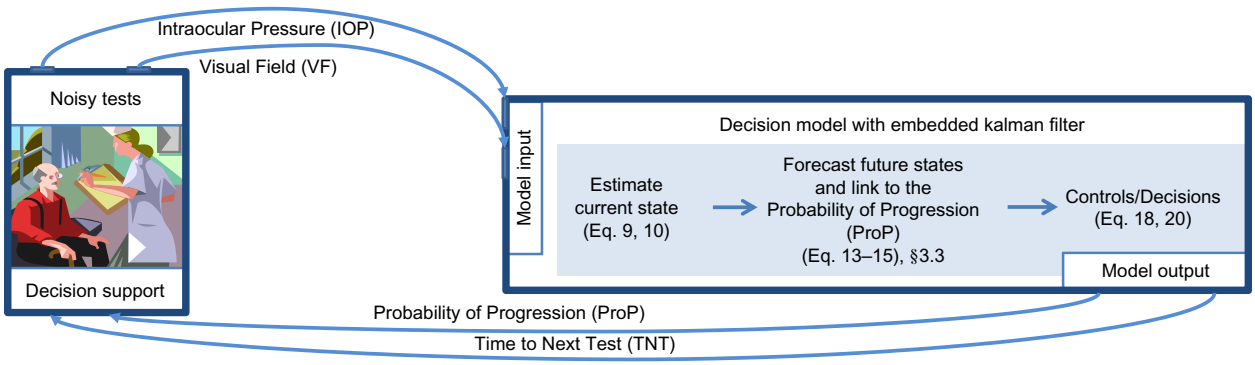
From an analytical perspective, we develop a closed-form solution to a nonlinear optimization over the multivariate Gaussian density describing future disease state. This optimization determines the monitoring frequency, which enables (1) efficient solutions for real-time clinical decision support and (2) identification of structural properties that provide important clinical insight into testing frequency. Our insights and results support recent clinical hypotheses on dynamic monitoring.

Beyond analytical results, clinical relevance and acceptance hinges on rigorous model validation. We tested our algorithm using data from two large-scale, 10+ year glaucoma clinical trials. We show that our methods significantly outperform current practice by achieving *greater accuracy* in identifying progression with *fewer examinations*.

Finally, the monitoring problem we address in this manuscript is not unique to glaucoma. Medical conditions that would benefit most from our approach are (1) asymptomatic early on in the disease; (2) effectively treatable to prevent morbidity and mortality if progression is detected early enough; (3) progressive and require patients to be followed over extended periods of time; (4) can lead to serious complications (such as blindness, kidney failure, stroke, and heart attack); and (5) have quantifiable measures (such as protein level measurements, blood pressure measurements, and viral load levels). Examples of chronic diseases for which physicians periodically monitor a number of quantifiable medical tests to capture progression include diabetes mellitus, connective tissue diseases, kidney diseases, and lupus. Given that chronic diseases affect almost one out of every two adults in the United States and account for 75% of U.S. healthcare spending (see CDC 2013), the proposed methodology has the potential for a broad impact on cost as well as on patients' quality of care and quality of life.

A high level view of our approach to disease monitoring is depicted in Figure 1. The model begins with an observation epoch at which the patient is given the required set of medical tests (e.g., VF and IOP tests). These tests may be perturbed by measurement noise. The noisy measurements are fed into a Kalman filter (also known as a Kalman-Bucy filter) model to obtain an estimate of the current disease state as well as a forecast specifying a distribution on the patient's future disease state. The forecasted disease state is then fed into a function—we term this the *probability of progression* (*ProP*) function—that converts the disease state into a probability that the patient will have progressed sufficiently to warrant a change in disease management. Finally, the time to next test (TNT) is given by a function that identifies the earliest time point that the patient's forecasted probability of progression will exceed a predetermined progression threshold.

Figure 1. (Color online) Decision support framework for chronic disease monitoring.



This paper's methodological contributions include the analysis of the interaction of the ProP function with the state space model parameters and the Kalman filter mean and covariance calculations. The structural properties analyzed generate new insights into the practice of monitoring patients, some of which have been hypothesized by physicians (see Jansonius 2007), but have yet to be mathematically modeled.

The remainder of the paper is organized as follows. Section 2 provides an overview of the relevant literature. Disease state estimation and forecasting are detailed in §3. In §4, we discuss our approach to determine the TNT and the solution and structural properties of our algorithm. Section 5 applies our models retrospectively to two large-scale, 10+ year glaucoma clinical trials for validation and demonstrates how our algorithm can deliver improved patient care with fewer tests compared to other policies that are similar to current practice. We describe how our algorithms may be integrated into current practice as well as discuss model limitations in §6. Finally, we discuss our results and future directions in §7.

2. Current State of Literature

There are three primary areas in the literature relevant to our approach: (1) medical examination models; (2) machine surveillance, inspection and maintenance; and (3) linear quadratic Gaussian (LQG) systems with controlled observations (i.e., control of measurement subsystems).

Medical Examination Models. Most research in the field focuses on performing discrete screenings to detect the first incidence of a disease, rather than monitoring an ongoing chronic disease. Denton et al. (2011) provides insight into some of the open challenges in this area, including those that we address here. The two main approaches are either cost based or assume a fixed number of examinations. Such models have been developed for cancer and diabetes mellitus among other chronic diseases (see Lincoln and Weiss 1964, Michaelson et al. 1999, Shwartz 1978, Baker 1998, Maillart et al. 2008, Rauner et al. 2010, Zelen 1993, Özekici and Pliska 1991, Hanin and Yakovlev 2001, Kirch and Klein 1974, Day and Walter 1984, Chhatwal et al. 2010). The recent work of Ayer et al. (2012) begins to explore personalizing testing schedules incorporating risk factors and history of tests similarly to our work. Lee and Wu (2009) develop disease classification and prediction approaches using math programming.

A second related research area involves monitoring and treatment decisions of an ongoing condition. Work has been done with regard to the timing of initial treatment (see Shechter et al. 2008, Denton et al. 2009, Shechter et al. 2010). The above research, however, does not incorporate multidimensional state spaces in feedback driven control loops to monitor patient-specific disease progression. For example, models have been developed for the treatment of HIV, diabetes, organ transplantation, cancer, and management of drug therapy dosages (see D'Amato et al. 2000, Lee et al. 2008, Alagoz et al. 2004, Lavieri et al. 2012, Hu et al. 1996). These approaches, however, only model a low dimensional health state with varying levels of degradation. In addition, existing models that consider frequency of monitoring decisions do not incorporate dynamic updating of information, rather making the assumption that all patients progress according to population statistics-driven transition functions. This is insufficient for the complex disease modeling we pursue in this work.

There is little work that seeks to model the complexities of a given disease by considering multiple interacting physiological indicators available. By using Gaussian state space models for disease progression and monitoring, our work is able to capture multidimensional, continuous state space models with correlated measurement noise in a tractable manner. This approach increases the scope of monitoring problems that can be solved and opens up the possibility of capturing complex and evolving diseases that are measured using a variety of different tests.

Machine Surveillance, Inspection and Maintenance. Extensive literature surveys of machine maintenance, inspection, and surveillance include Pierskalla and Voelker (1976), Sherif and Smith (1981), Barlow et al. (1996), and Wang (2002). These surveys propose that the literature can be divided into five primary modeling approaches: (1) age replacement models, (2) block replacement models, (3) delay-time models for inspection, (4) damage models, and (5) deterioration models.

Model types (3)–(5) are particularly relevant to the monitoring of chronic diseases. Damage models determine the properties of the failure time (e.g., disease progression), but do not consider the effect of inspections (see Nakagawa and Osaki 1974, Morey 1966). Deterioration and delay-time models assume that machine degradation can only be observed by inspecting the system. Inspection carries cost c_1, the current state of degradation carries a cost of c_2, and there is typically a cost for replacement and/or repair proportional to the state of deterioration (see Luss 1976, Yeh 1997, Ohnishi et al. 1986a, Mine and Kawai 1975, Derman and Sacks 1960, Bloch-Mercier 2002) or the length of time a failure goes undetected (see Keller 1974, Kander 1978, Munford and Shahani 1972, Donelson 1977, Luss 1983, Savage 1962, Barlow et al. 1963). These models, however, consider a one-dimensional state space with Markovian or semi-Markovian system dynamics and perfect observations, which is insufficient for our application.

In non-Markovian surveillance and inspection models (see Antelman and Savage 1965, Nakagawa and Yasui 1980, Kander and Raviv 1974, Chitgopekar 1974), the state space is still one-dimensional and the observations are assumed to be perfect. Papers that consider noisy or uncertain observations include Savage (1964), Noonan and Fain (1962), Rosenfield (1976), Eckles (1968), Ohnishi et al. (1986b). Again, the state space is one-dimensional and, while some models consider rich noise components, most consider only simple noise. Chronic disease progression

monitoring requires a multidimensional state space with both observation noise and correlated system noise. By incorporating these features, this paper expands the modeling approaches in inspection/surveillance and deterioration/damage modeling.

Linear Gaussian Systems. Linear Gaussian systems and LQG have been used in many different applications in dynamical systems modeling, estimation, and control theory. Our models, however, focus on systems *without fixed observation intervals*, which represents a major departure from the foundational models. Sensor scheduling research does investigate the question of how frequently, for a given set of available sensors, one should take measurements and from which sensors. However, our decisions on when to test and whether or not to declare progression fall outside the class of quadratic objective functions used in sensor scheduling. Work in sensor scheduling includes Mehra (1976), Oshman (1994), Wu and Arapostathis (2008); however, this literature typically assumes that a measurement is taken every period (though from different sensors). Control of measurement subsystems (see Meier III et al. 1967, Athans 1972, Lafortune 1985) is the area most closely related to ours. This work considers the problem of whether or not to take a measurement in each period. There is a cost for taking a measurement, a cost for system control, and a cost associated with each system state at every time instance. Our work extends the LQG control theory by formulating and analyzing the class of *monitoring problems* in combination with user input and employing nonstandard optimization approaches incorporating potentially complex disease progression functions.

3. State Space Modeling of Progression

We develop state space models for estimating and forecasting a patient's disease trajectory. In §3.1 we present our modeling approach, which is then applied in §3.2 to glaucoma patients from two major clinical trials, the Collaborative Initial Glaucoma Treatment Study (CIGTS) and Advanced Glaucoma Intervention Study (AGIS). Finally, §3.3 briefly describes the nature of the ProP estimator that converts a modeled disease state into a probability of progression. This component links the forecasting mechanisms developed in this section with the control on testing intervals presented in §4, which is illustrated in Figure 1.

3.1. Gaussian Continuous State Models of Disease Measurement Dynamics

Our vector continuous state space models are in the class of first-order stochastic difference equations with correlated Gaussian noise inputs such that the noise is independent from one period to the next (i.e., white). These first-order models are adequate for a surprisingly general class of systems, especially if state augmentation is used to linearize a nonlinear model by including the derivatives of key variables in the state space (see Bertsekas 1987, 2000a, b). This class of systems allows us to develop correlated multivariate Gaussian noise models for both (1) process noise, which can approximate the effect of unmodeled dynamics, and (2) measurement noise in medical test measurements. Our system model underlying the Kalman filter is comprised of a continuous, vector patient disease state and the system disease dynamics.

3.1.1. Patient Disease State. Current evidence indicates that a primary indicator of glaucoma progression is worsening of the visual field, and that IOP is a critical risk factor for future progression. In our model, we consider a nine-dimensional column vector to model the state of the patient, α_t:

$$\alpha_t = [MD, MD', MD'', PSD, PSD', PSD'', IOP, IOP', IOP'']^T, \quad (1)$$

where MD and PSD refer to two global measures of performance from the visual field test. Additional measures of performance from that test might be used contingent on data availability. Similarly, IOP represents the intraocular pressure measurement. The state variables MD' and MD'' refer to the first and second derivatives of the MD measure with respect to time: velocity and acceleration. Similar derivatives are taken of the PSD and IOP measures. Continuous time data on MD, PSD, and IOP measurements are not available because readings are at discrete time points, so we estimate the derivatives from the discrete time data. Given measurements x_1 at time t_1, x_2 at time t_2, and x_3 at time t_3, the first derivative is estimated via $(x_2 - x_1)/(t_2 - t_1)$ and the second derivative is estimated via $((x_2 - x_1)/(t_2 - t_1) - (x_3 - x_2)/(t_3 - t_2))/(t_3 - t_2)$. From the doctor's perspective, the visual field machine or IT system can estimate these derivatives from the history of observations. This estimate can then be combined with the underlying system dynamics that capture the derivative changes in the dynamical system model.

3.1.2. A Kalman Filter Model for the Disease Measurements. Our discrete-time disease model is recursive. In each period, there is a system transition and also a measurement of the system that can be taken by the observer/controller. The difference equation formulation of these system dynamics consists of a *state transition equation* and a *measurement equation*. The transition equation defines how the disease is progressing from one period to the next and the measurement equation describes the system's observation of disease state through medical testing. As an anchor for the recursive system equations, there is an *initial state* that is assumed prior to any observations, based on population characteristics found in the CIGTS and AGIS clinical trials.

State Transition Equation. In each period, t, the system moves to a new state at $t+1$ according to a state transition matrix $\mathbf{T}$ and a vector Gaussian white noise input η. The Gaussian noise represents unmodeled disease process noise. The recursive transition equation is given by

$$\alpha_t = \mathbf{T}\alpha_{t-1} + \eta \quad t = 1, \ldots, N, \tag{2}$$

where η is a Gaussian random vector with $\mathbf{E}[\eta] = 0$ and $\mathrm{Var}[\eta] = \mathbf{Q}$. Clearly the system state, α_t, is also a Gaussian random variable for all t since it is the result of a linear combination of Gaussian random variables.

Measurement Equation. In the measurement equation, $\mathbf{z}_t$ denotes the observation vector; i.e., the outcomes of the series of tests that are performed at each glaucoma patient's visit. Let $\mathbf{Z}$ denotes the matrix that models how components of the true state, α_t, are observed; ϵ is the Gaussian noise component that denotes the test noise described in §1. The measurement equation has the form

$$\mathbf{z}_t = \mathbf{Z}\alpha_t + \epsilon \quad t = 1, \ldots, N, \tag{3}$$

where ϵ is a Gaussian random variable with $\mathbf{E}[\epsilon] = 0$ and $\mathrm{Var}[\epsilon] = \mathbf{H}$. Again, clearly the observation $\mathbf{z}_t$ is a Gaussian random variable for all t.

Finally, let the initial state be a Gaussian random vector, X_0, with $\mathbf{E}[X_0] = \hat{\alpha}_0$ and covariance matrix $\mathrm{Var}[X_0] = \hat{\Sigma}_0$.

3.1.3. State Estimation and Prediction with the Kalman Filter. For the above model, the Kalman filter optimally estimates the mean and covariance parameters that completely characterize the state of the linear Gaussian system based on noisy observations. In each period, the Kalman filter performs two steps to generate state estimates: *prediction* and *update*. In the prediction step, the linear state transition model is used to estimate the mean and covariance of the next state. In the update step, new observations are used to optimally correct the model's prediction so as to minimize the mean squared error of the estimate: $\mathbf{E}[|\alpha_t - \hat{\alpha}_t|^2]$. Using the notation developed in §3.1.2, the Kalman filter approach (see Kalman 1960) is summarized below.

Prediction Step. The prediction step takes the most recent mean and covariance estimate with information up to time t, $\hat{\alpha}_{t|t}$ and $\hat{\Sigma}_{t|t}$, and uses the system dynamics model from Equation (2) to predict the future state as

$$\hat{\alpha}_{t+1|t} = \mathbf{T}\hat{\alpha}_{t|t} \tag{4}$$

$$\hat{\Sigma}_{t+1|t} = \mathbf{T}\hat{\Sigma}_{t|t}\mathbf{T}' + \mathbf{Q}, \tag{5}$$

where $\hat{\alpha}_{t+1|t}$ and $\hat{\Sigma}_{t+1|t}$ are the predicted mean and covariance at time $t+1$ given observations up to time t. Also note that the prime symbol, $'$, represents the matrix transpose.

Update Step. After the prediction step, a new observation, $\mathbf{z}_{t+1}$, is obtained and the error between the prediction and the observation is used to calculate the optimal new state estimate. In this step, first the measurement residual, $\tilde{\mathbf{y}}_{t+1}$, and the predicted covariance of the measurement, $\mathbf{S}_{t+1}$, are calculated as

$$\tilde{\mathbf{y}}_{t+1} = \mathbf{z}_{t+1} - \mathbf{Z}\hat{\alpha}_{t+1|t} \tag{6}$$

$$\mathbf{S}_{t+1} = \mathbf{Z}\hat{\Sigma}_{t+1|t}\mathbf{Z}' + \mathbf{H}. \tag{7}$$

The optimal Kalman gain, $\mathbf{K}_{t+1}$, is the solution to an optimization that minimizes the trace of the estimated covariance matrix (and thereby minimizes the mean squared error of the estimate). The optimal Kalman gain is given by

$$\mathbf{K}_{t+1} = \hat{\Sigma}_{t+1|t}\mathbf{Z}' \cdot \mathbf{S}_{t+1}^{-1}. \tag{8}$$

The optimal Kalman gain from Equation (8) is used to calculate the optimal new state estimate, $\hat{\alpha}_{t+1|t+1}$ and $\hat{\Sigma}_{t+1|t+1}$, for the Gaussian state random variable as

$$\hat{\alpha}_{t+1|t+1} = \hat{\alpha}_{t+1|t} + \mathbf{K}_{t+1} \cdot \tilde{\mathbf{y}}_{t+1} \tag{9}$$

$$\hat{\Sigma}_{t+1|t+1} = (I - \mathbf{K}_{t+1}\mathbf{Z})\hat{\Sigma}_{t+1|t}, \tag{10}$$

where I is the identity matrix. Equations (9) and (10) are the key equations that define the recursive Kalman estimator and will be relied upon in subsequent analysis.

Multiperiod Prediction. In our application, the condition of each patient varies from one patient to another, so the optimal time interval between tests will vary from one measurement to the next depending on the patient's measurement history. Therefore, our approach must predict sufficiently many periods ahead before applying the update step. By eliminating the update step for periods in which no observation is performed, the transition equation yields the l-step prediction equation (i.e., predicting l periods into the future) as

$$\hat{\alpha}_{t+l|t} = \mathbf{T}^l\hat{\alpha}_{t|t} \tag{11}$$

$$\hat{\Sigma}_{t+l|t} = \mathbf{T}^l\hat{\Sigma}_{t|t}(\mathbf{T}^l)' + \sum_{j=0}^{l-1}\mathbf{T}^j\mathbf{Q}\mathbf{T}^{j'}, \tag{12}$$

where α_{t+l} is the Gaussian state variable at time $t+l$ given that observations are available through time t (i.e., the observation history). The first element of the sum represents the multiperiod state transition and the second element of the sum in Equation (12) represents the multiperiod process noise accumulation.

3.2. Application to Two 10+ Year Clinical Trial Data Sets

In §§3.1.1–3.1.3 we presented the theoretical framework for modeling disease progression in glaucoma patients. To

validate our approach, we used real patient data from two 10+ years large randomized clinical trials of glaucoma patients, CIGTS and AGIS. CIGTS is a randomized clinical trial that followed 607 participants with newly diagnosed, mild to moderate glaucoma for up to 10 years. During the course of the trial, visual field and IOP readings were taken every six months. Participants were initially randomized to one of two treatment arms: medical or trabeculectomy (a surgical intervention). Participants who did not respond well to their treatment arm were given an additional treatment of argon laser trabeculoplasty (ALT).

AGIS followed 591 participants with advanced glaucoma for up to 11 years. Similar to CIGTS, measurements of VF and IOP for each participant were taken every six months. AGIS participants were randomized to one of two treatment sequences: one sequence began with ALT and the other began with trabeculectomy. Participants responding poorly to their initial treatment received the other treatment next. In both studies, for each participant a single eye was studied. The study eye was assigned prior to randomization based on the eligibility status of the eye. If both eyes were eligible, it was assigned based on the treating physician's selection.

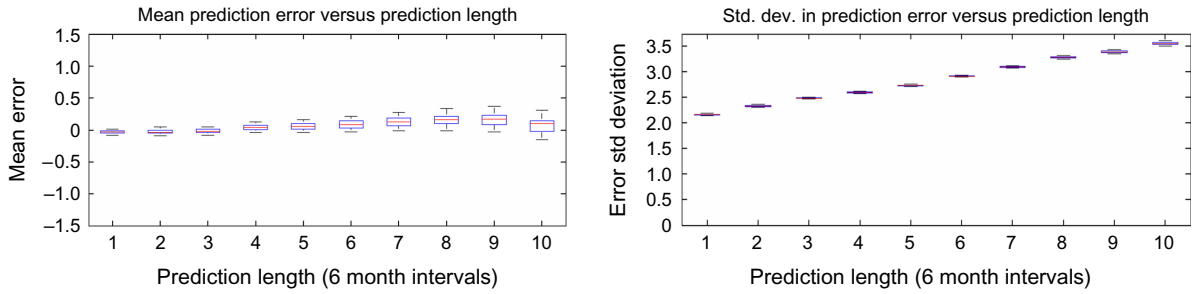
We combined the longitudinal data of the two randomized clinical trials into one data set. For our case study, we focused on participants from the clinical trials who were treated with medicine or ALT. Participants included in our study were randomly divided into equal size training and testing sets in a manner that maintained the original ratio between progressing and nonprogressing patients, as well as the mixture of mild, moderate, and advanced glaucoma patients and patients coming from each trial in both the training and testing sets. The time step for the linear Gaussian system was set to six months to match the time step of the data. Though the time step can be chosen to be any arbitrary length, we chose six months to avoid making assumptions about progression at points in time where data were not available. In other words, one transition moves the system forward in time six months. The training data were used to calibrate the model, employing the expectation maximization (EM) algorithm for parameter estimation of the Kalman filter and its implementation in Matlab (see Ghahramani and Hinton 1996, Digalakis et al. 1993, Murphy 1998) to find the matrices $\mathbf{T}$ (linear system dynamics), $\mathbf{Q}$ (process noise covariance), $\mathbf{Z}$ (the observation matrix that allows some or all of the states to be measured in a possibly altered form), $\mathbf{H}$ (measurement noise covariance), $\hat{\alpha}_0$ (initial state mean), and $\hat{\Sigma}_0$ (initial state covariance). Although the initial state, $(\hat{\alpha}_0, \hat{\Sigma}_0)$, is based on the population statistics, in practice when a new patient establishes with a glaucoma specialist (or is newly diagnosed), several baseline measurements are taken for MD, PSD, and IOP to assess the state of the disease. These baseline readings were then input as observations to the Kalman filter. Thus the initial state is used only as an initial condition for the recursion, which is then immediately tailored to the individual patient through several (often 2+) baseline readings before any testing schedules are generated. This allows the system to adjust to the individual patient (and away from the population mean) before results are generated. In our tests on the clinical trials, distance from the initial state did not significantly affect future state forecasts because the baseline readings were sufficient for the model to be tailored to the individual patient.

EM Algorithm. The EM algorithm has two steps that are performed iteratively until the algorithm converges: the E step and the M step. The system is initialized with an estimate of the matrices and vectors we want to fit. In the E step, the Kalman filter is run forward and backward (known as Kalman smoothing) on the data to provide the best estimate of the true system state at each time t given all the data available, including data coming before and after time t in the sequence. This yields the estimated Gaussian distribution for each time period. In the M step, the expected log likelihood of the set of observations is maximized by taking matrix derivatives with respect to the parameters to be estimated and setting them equal to zero (with the expectation taken over the Gaussian distribution from the E step). The parameters output from the EM algorithm for a particular training data set are given in Appendix B in the online appendix (available as supplemental material at http://dx.doi.org/10.1287/opre.2015.1405).

3.2.1. Model Fit and Normality. Our modeling approach enables us to efficiently capture system and measurement noise, but requires that we model our system as a set of stochastic difference (or differential) equations that are linear and have multivariate, correlated Gaussian noise. To ensure the robustness of the modeling approach and appropriateness in modeling glaucoma disease progression, we performed a sensitivity analysis. First, we randomly generated 25 training data sets (with the complement of training used for testing), while maintaining the proportion of different types of patients seen in the general population. Then, we parameterized the Kalman filter on each one of the 25 training data sets (using the EM algorithm) and tested it on the remaining test data. The box plots in Figure 2 are a result of these 25 separate parameterizations and runs of the Kalman filter. It can be seen from the tight box plots in Figure 2 that the model is quite robust to the patient data used to parameterize it.

After parameterizing the Kalman filter, for each participant in the test set, we used the Kalman filter model to predict MD values (MD being the most significant variable) for five years into the future for most patients. The prediction error (i.e., the predicted mean state minus the actual observation) was computed for each of the 25 training/test data set combinations mentioned above. The estimated average error and error standard deviation are given in the left and right plots, respectively, of Figure 2. These

Figure 2. (Color online) Kalman filter prediction error versus number of six month long periods into the future.



(a) Mean MD prediction error

(b) Standard deviation of prediction error

box and whisker plots show that our model for state prediction has very little bias. The red line is the median, the upper and lower edges of the box show the upper and lower quartiles of the data, and the whiskers show the maximum and minimum values observed. The fact that the boxes are very thin shows that the model is robust to the data used to parameterize the filter. Equivalent results have been obtained for other state variables.

Whereas we acknowledge that it is an approximation to model the process noise and state observation noise as both multivariate Gaussian random variables, numerical testing revealed that the Gaussian model is a reasonable fit. We analyzed the Kalman Filter residuals/innovations (the error between next step prediction and the actual observation) to test whether or not the system model is effective, evidenced by the residuals possessing a Gaussian distribution. For each element of the residual vector, the p-values of t-tests for unbiasedness (supporting the linearity assumption) as well as for the Shapiro test for normality support the case that these are normally distributed with zero mean. quantile-quantile (QQ) plots were used to compare the quantiles from the empirical distribution of the actual data to the quantiles of the hypothesized Gaussian distribution. For MD, PSD, and IOP, respectively, we have a match of the data to a Gaussian distribution for values within 2.5, 2.8, and 1.9 standard deviations of the mean (which is 95% of outcomes even in the worst case of IOP). With this good model fit and almost no bias (see Figure 2), we are confident the model is sufficiently capturing critical system dynamics.

3.3. Progression Models: Glaucoma ProP Function

Our next step is to match the Kalman Filter variables with treatment decisions. In glaucoma, as is the case with various chronic diseases, clinicians often face the challenge of interpreting multidimensional data to make decisions of how best to treat their patients (see Katz 1999). This can be difficult in practice because the amount of data is so large and is processed mentally without the aid of any similar decision support system. Identifying and properly utilizing this multidimensional space of information over a history of observations is the purpose of the ProP function. Specifically, the ProP function, f, maps the state space of physiological indicators, $\mathcal{S}$, to a measure of disease progression in $[0, 1]$: probability of progression.

In collaboration with subject matter experts and leveraging medical literature (e.g., Hodapp et al. 1993), we developed a glaucoma progression definition using the physiological indicators. Because there is no gold standard to measure glaucoma progression, our work has focused on identifying drops of three MD with respect to baseline that are validated in at least one instance into the future (see Musch et al. 2009). This definition has been compared against other progression definitions (such as Nouri-Mahdavi et al. 2004 and Hodapp-Anderson-Parrish from Hodapp et al. 1993) on a subset of patients for which sufficient data were available. Other definitions may be further explored in the future, contingent on data availability. All glaucoma progression instances were validated using the longitudinal data. After extensive testing of many approaches, we chose a logistic function, $f(x)$ where x is the disease state vector, to assess the probability of glaucoma progression for any patient at any given time:

$$f(\mathbf{x}) = \frac{1}{1 + e^{-w(\mathbf{x})}}, \tag{13}$$

$$w(\mathbf{x}) = b + \mathbf{ax}, \tag{14}$$

where $w(\mathbf{x})$ is a linear function of key risk factors, including MD, PSD, and IOP measures and can include other important factors such as structural changes to the optic nerve, age, race, family history, medical history, and genetic factors among others. The regression coefficients are captured in the *progression vector*, $\mathbf{a}$, which represents the n-dimensional direction of steepest ascent toward progression.

We further consulted with glaucoma specialists and the literature (e.g., Nouri-Mahdavi et al. 2004, De Moraes et al. 2011) to determine appropriate risk factors to consider in developing the ProP function. For our case study, we used generalized estimating equations with a logit link function on the training set of study participants to parameterize the ProP function. Starting with sex, age, race, baseline

MD, MD, MD velocity, MD acceleration, baseline PSD, PSD, PSD velocity, PSD acceleration, baseline IOP, IOP, IOP velocity, and IOP acceleration as our initial set of covariates, backward variable selection was performed with a significance level of 0.05 to determine the final set of covariates for our case study. In addition, for a subset of patients from the CIGTS trial, we also had available additional factors such as cardiac or vascular disease, disc hemorrhage, open angle glaucoma (OAG) diagnosis of both eyes (study eye and fellow eye) at baseline (i.e., none, primary open angle glaucoma, pseudoexfoliation, pigmentary, other). After performing forward and backward elimination on that subset of patients, we further concluded that none of the additional variables made a significant difference in our predictions. While some of the variables were significant in univariate analysis, they did not change our estimated area under the ROC curve (AUC) when we incorporated them into the models. Thus, these additional factors were not included in the ProP logistic regression function.

Unfortunately, we could not include information on the retinal nerve fiber layer as captured using optical coherence tomography because the technology to gather such clinical data were not available at the time the CIGTS and AGIS trials were carried out. Genetic factors were also not available to us from the clinical trials used for validation of our models. Incorporation of such factors, among others, may improve the accuracy of the models presented. The factors found to be relevant in our study were MD position (MD), velocity (MDV) and acceleration (MDA), PSD baseline value (PSDB), PSD position (PSD) and age. The coefficients we used are given as

$$\begin{aligned} w(\mathbf{x}) = &-6.0035 - 0.0568 \cdot MD - 4.0544 \cdot MDV \\ &- 1.1832 \cdot MDA - 0.1615 \cdot PSDB + 0.1536 \cdot PSD \\ &+ 0.0255 \cdot age, \end{aligned} \tag{15}$$

with a full description of the model and approach given in Schell et al. (2013). The AUC (obtained from the Mann-Whitney U statistic) for the ProP function applied to the testing set was 0.919, which is clinically considered to be very good. Additional covariates (including structural changes to the optic nerve, diabetes mellitus, medical history, and genetic factors) may improve our estimations and should be considered in future implementations of our models. Although IOP and its derivatives were not used as a factor in the ProP function (Equation (15)), it was found to be important in the Kalman filter modeling of test measurement evolution because IOP interacts with VF and PSD. A thorough treatment of the key factors involved in glaucoma progression can be found in Musch et al. (2009) and Schell et al. (2013).

4. Time to Next Test (TNT)

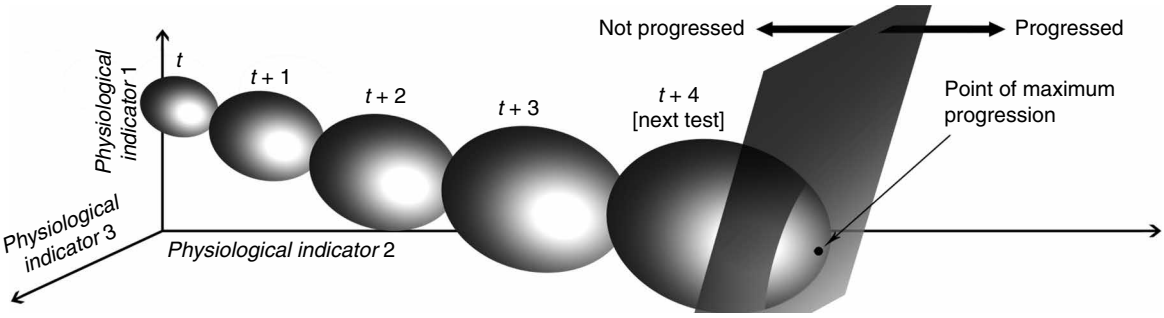
The idea behind our approach is to balance testing frequently to catch progression early against the cost, discomfort, and inconvenience associated with testing. We capture this trade-off by delaying testing until the point in time at which the model indicates that we can no longer be statistically "confident" that the patient has not progressed. Specifically, to determine the TNT, we forecast the patient disease state trajectory into the future until the ProP function hits an optimized threshold indicating sufficient likelihood of progression for a test to be performed. The TNT interval of time to the next test is therefore determined by the length of time it takes for the disease state forecast to reach the optimized progression threshold. We develop a stochastic point of maximum progression (POMP) function that maximizes the deterministic ProP function over the n-dimensional Gaussian density of the forecasted state. This yields the "worst case" point, or the point of maximum progression, within a confidence region around the mean state vector; a conservative estimate of the patient's probability of progression. Experimental testing is used to tune the parameters controlling the size of the confidence region (ρ) and the optimized progression threshold (τ) to capture the trade-off between catching progression and cost of testing mentioned above.

Figure 3 is a conceptual representation of this approach for a three-dimensional state space. In this figure t is the current period and the ellipsoid at period t represents the $100\rho\%$ confidence region around the state estimate. As we forecast the patients disease state further into the future (e.g., periods $t+1, t+2, \ldots$), the center of the confidence region (i.e., the forecasted mean state) moves in accordance with the disease dynamics (i.e., transition matrix $\mathbf{T}$). In addition, the confidence region expands as the covariance around the forecasted mean grows the further into the future the state is predicted. The time of the next test occurs at the first period in which the forecasted confidence region intersects or exceeds the progression threshold (an n-dimensional hyperplane), illustrated by the plane in Figure 3; in this case period $t+4$.

Our model has two parameters, ρ and τ, that control how aggressively to test a given patient. τ sets a threshold on the probability of progression (the plane in Figure 3). At the same time that the probability that a patient has progressed exceeds τ the algorithm recommends taking another test at that time. Smaller values of τ indicate a lower tolerance for missing progression because the patient reaches the threshold more quickly, generating more frequent testing. ρ adjusts the size of the confidence region around the predicted mean disease state (the ball in Figure 3), with larger values generating more frequent tests. For clinician usability, we present in §§5.2 and 5.3 an intuitive three-level aggressiveness scale (low, medium, high) to be selected by clinicians that set ρ and τ to realize the desired monitoring aggressiveness.

In practice, if clinicians receive a suspicious/unreliable test result it is common to schedule a follow-up test in the near future to confirm the results, because the test results may not be informative and thereby would be ignored if

Figure 3. Depiction of the confidence region point of maximum progression time to next test approach, POMP TNT.



unconfirmed. Our optimal estimation and TNT scheduling algorithm support this clinical process in the following ways. First, the filtered ProP reading obtained immediately upon receiving the exam results would give the clinician an indication of whether there is concern regarding the patient's condition. The Kalman filtering helps to reduce the noise in the testing giving the clinician a clearer picture of the patient's status. If the clinician feels the VF exam results are suspicious/unreliable (either because of the filtered ProP estimate or because of some of the error checking in the Humphrey Visual Field Analyzer), the clinician will schedule a subsequent (follow-up) test for the near future to either confirm or invalidate the suspicious/unreliable test. This test will be done off line, and when the clinician is satisfied, the nonsuspicious result will be added to the Kalman filter algorithm and used to calculate the time of the next regular test. Alternatively, the average (or weighted average) of the results might be included in the algorithm. In the case of the two clinical trial studies on which we tested our algorithm, only nonsuspicious results were included, which provides the same results as the approach we describe for using this system in clinical practice.

4.1. Point of Maximum Progression Time to Next Test Approach

In this section we develop a closed-form solution to the optimization of the ProP function over the Gaussian prediction region. Mathematically we can define the $100\rho\%$ prediction region for the Gaussian random variable with mean $\hat{\alpha}_{t+l|t}$ and covariance $\hat{\Sigma}_{t+l|t}$ for l periods into the future as

$$\mathscr{D}_\rho(\hat{\alpha}_{t+l|t}, \hat{\Sigma}_{t+l|t}) = \{\mathbf{x}\colon (\mathbf{x}-\hat{\alpha}_{t+l|t})'\hat{\Sigma}^{-1}_{t+l|t}(\mathbf{x}-\hat{\alpha}_{t+l|t}) \leqslant \chi^2(1-\rho, n)\}, \quad (16)$$

where $\hat{\alpha}_t$ and $\hat{\Sigma}_t$ represent our current estimate of the mean and covariance of the disease state at time t (see Chew 1966). Also, $\chi^2(1-\rho, n)$ is the $1-\rho$ quantile of the chi-square distribution with n degrees of freedom.

The goal is to associate the state estimate with ProP by using function f. A logical and conservative approach is to find the maximum value of the ProP function, f, over the prediction region, $\mathscr{D}_\rho(\hat{\alpha}_{t+l|t}, \hat{\Sigma}_{t+l|t})$. Given the current state estimate, $\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}$, the stochastic POMP function, h_ρ, with respect to the ProP function, f, for the l-step state forecast is given by

$$h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, l) = \max_{\mathbf{x}\in\mathscr{D}_\rho(\hat{\alpha}_{t+l|t}, \hat{\Sigma}_{t+l|t})} f(x), \quad (17)$$

where $\hat{\alpha}_{t+l|t}, \hat{\Sigma}_{t+l|t}$ are obtained from $\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}$ through Equations (11) and (12). We first observe that the prediction region, $\mathscr{D}_\rho(\hat{\alpha}_{t+l|t}, \hat{\Sigma}_{t+l|t})$, defined by Equation (16) is convex.

It is possible that for many chronic illnesses, as with glaucoma, the ProP function will be a logistic regression as described in §3.3. Therefore, maximizing the ProP function is equivalent to maximizing $w(\mathbf{x})$ (see Equation (13)), which is a linear function of $\mathbf{x}$. Thus finding the point of maximum progression is then a convex optimization problem. To solve this optimization problem, we rely on the Karush-Kuhn-Tucker (KKT) conditions.

Recall that $\mathbf{a}$ is the progression vector of risk factors from Equation (14). The optimization of the ProP function over the prediction region has a closed-form solution given by Theorem 1, which is proved in the online appendix. The closed-form solution was determined using a two-stage approach based on the observation that the KKT conditions are both necessary and sufficient. First we solved the KKT stationarity conditions for an arbitrary coefficient of the constraint gradient. The resulting solution was input into the complementary slackness conditions to determine the appropriate coefficient.

THEOREM 1. *Given the l-step prediction region $\mathscr{D}_\rho(\hat{\alpha}_{t+l|t}, \hat{\Sigma}_{t+l|t})$ defined by Equation (16) with $\rho \in (0, 1)$ and progression vector $\mathbf{a}$, the maximum value of the ProP function, h_ρ, and the associated disease state, $\tilde{h}_\rho$, have a closed-form solution,*

$$h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, l) = \max_{\mathbf{x}\in\mathscr{D}_\rho(\hat{\alpha}_{t+l|t}, \hat{\Sigma}_{t+l|t})} \mathbf{a}'\mathbf{x} = \mathbf{a}'\hat{\alpha}_{t+l|t} + \sqrt{\chi^2(1-\rho, n)\mathbf{a}'\hat{\Sigma}_{t+l|t}\mathbf{a}} \quad (18)$$

$$\tilde{h}_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, l) = \underset{\mathbf{x}\in\mathcal{D}_\rho(\hat{\alpha}_{t+l|t}, \hat{\Sigma}_{t+l|t})}{\arg\max} \mathbf{a}'\mathbf{x}$$
$$= \hat{\alpha}_{t+l|t} + \left(\sqrt{\frac{\chi^2(1-\rho, n)}{\mathbf{a}'\hat{\Sigma}_{t+l|t}\mathbf{a}}}\right)\cdot\hat{\Sigma}_{t+l|t}\mathbf{a}. \quad (19)$$

Finally, given a progression threshold of τ, the time to next test is determined by the TNT function, $F_{\rho,\tau}(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t})$, where $F_{\rho,\tau}$: $\mathbb{R}^n \times (\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{N}$, maps the current state to the time interval between the current observation and the next observation:

$$F_{\rho,\tau}(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}) = \min_{l\in\mathbb{Z}^+} l \quad \text{s.t. } h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, l) \geqslant \tau. \quad (20)$$

In the next section we prove that the POMP function, h_ρ, is monotonically increasing in l, therefore the TNT function can be solved quickly and easily with iterative search techniques. For a problem with n possible testing epochs, a simple binary search that divides the search space in half at each iteration can solve this problem in the worst case on order of $O(\log(n))$, because the terms are monotonically increasing in l. Even when the search space is large, the algorithm will find the solution quickly. For example, imagine a disease that can be monitored on intervals of one second over the course of a year (a total of 31,449,600 possibilities). Our search method requires at worst 25 function evaluations plus comparisons to solve the optimization, which would be nearly instantaneous.

In §5, we compare the performance of our TNT algorithm with currently accepted medical practice. We also present in §4.2 structural insights from our approach that have been hypothesized by researchers and clinicians but, to our knowledge, have not yet been rigorously validated. The first (see Jansonius 2007) is that testing intervals for glaucoma should be variable rather than fixed. Our approach goes even further by showing how the testing interval can be determined using the key physiological indicators and providing an indication of the benefits.

4.2. Structural Properties of the TNT Algorithm

In this section we discuss the structural properties of the TNT algorithm and the insights they provide. Property 1, given in Theorem 2, says that the further into the future we wait before testing, the more uncertain we are about whether the patient has progressed or not and the more likely the patient has gotten worse, and thus are more likely to test. Property 2, given in Lemma 1, states that the more patient observations the model has, the smaller the estimated covariance is in the *direction of progression*, $\mathbf{a}$ (i.e., the direction of the progression vector $\mathbf{a}$ from Equation (14)). Property 3, given in Theorem 3, states that the system will test more frequently when there is less information about a patient. Property 4, given in Theorem 4, states that the worse off (i.e., closer to progression) a patient is, the more frequently they will be tested.

For many chronic diseases, called degenerative diseases, the disease tends to get worse over time. Some clear examples include Alzheimer's, Parkinson's, and ALS among others. For glaucoma, lost sight cannot be recovered. Mathematically the progressive nature of chronic disease can be captured by the following condition on the system transition matrix, $\mathbf{T}$.

Definition 1. We call a linear transformation $\mathbf{T}$, a *progressing transformation* with respect to progression vector $\mathbf{a} \in \mathbb{R}^n$, if and only if

(i) $\mathbf{a}'\mathbf{T}\alpha \geqslant \mathbf{a}'\alpha$ for all states $\alpha \in \mathcal{S}$, and

(ii) for any matrix $\mathbf{B}$ such that $\mathbf{a}'\mathbf{B}\mathbf{a} \geqslant 0$, it follows that $\mathbf{a}'\mathbf{T}\mathbf{B}\mathbf{T}'\mathbf{a} \geqslant \mathbf{a}'\mathbf{B}\mathbf{a}$.

The intuition behind Definition 1 is as follows. Note that $\mathbf{a}$ is a vector representing the direction of progression (in n dimensions). For (i), the linear transition matrix representing disease dynamics, $\mathbf{T}$, transforms the state α. If $\mathbf{a}'\mathbf{T}\alpha \geqslant \mathbf{a}'\alpha$ then for any current state α, applying the linear transformation will always result in a state that is larger in the direction of progression. This captures the medical property that patients with glaucoma do not regain lost sight (i.e., get "better"). Condition (ii) is the quadratic version of condition (i) for capturing the progression concept with respect to the covariance matrix.

Property 1 (Prediction Uncertainty) *from Theorem 2, shows that as the Kalman filter predicts the patient's state further into the future, it monotonically approaches the threshold, τ, for scheduling a next test. It supports the intuition that, the further into the future we wait before testing, the more uncertain we are about the patient's disease state.*

Theorem 2. *If the linear system transformation, $\mathbf{T}$, is a progressing transformation, then for any state $(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t})$, the function*

$$h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, l) = \mathbf{a}'\hat{\alpha}_{t+l|t} + \sqrt{\chi^2(1-\rho, n)\mathbf{a}'\hat{\Sigma}_{t+l|t}\mathbf{a}}$$

is monotone increasing in l.

Property 2 (Number of Observations vs. Uncertainty) *shows that the covariance around the disease state estimate in the direction of progression is decreasing in the number of observations. Thus, the more information the system has about a patient, the less uncertainty there is in the disease state estimate with respect to whether the patient has progressed. For a rigorous statement of this property, we present notation and three definitions of properties of the covariance matrix.*

We consider a system where there is an initial observation at time t_s and a final observation at time t_f. Let $\Pi_n([t_s, t_f])$ be the set of open loop policies with n observations at times $s_1, s_2, \ldots, s_n$, where the first observation is at time $t_s = s_1$ and the final observation is at time $t_f = s_n$. Let $\hat{\Sigma}^{\pi_n}_{s_j|s_{j-1}}$ be the covariance estimate at time s_j given information up through time s_{j-1} under policy π_n—which can be

determined from $\hat{\Sigma}^{\pi_n}_{s_{j-1}|s_{j-1}}$ using the $(s_j - s_{j-1})$-step prediction Equation (12). Finally, let $\mathbf{K}^{\pi_n}_{s_j|s_{j-1}}$ be the $(s_j - s_{j-1})$-step Kalman gain under policy π_n defined by replacing the one-step covariance estimate with the $(s_j - s_{j-1})$-step covariance matrix in Equations (7) and (8).

Definition 2. Given an open loop observation schedule $\pi_n = \{s_1, s_2, \ldots, s_n\} \in \Pi_n([s_1, s_n])$, we define the *covariance estimate adjustment* at time $s_j \in \pi_n$ to be $\mathbf{C}^{\pi_n}_{s_j, s_{j-1}} = \mathbf{K}^{\pi_n}_{s_j|s_{j-1}} \mathbf{Z} \cdot \hat{\Sigma}^{\pi_n}_{s_j|s_{j-1}}$.

In other words, the covariance estimate adjustment at time s_j under policy π_n is simply the amount by which the covariance is reduced as a result of having an observation at time s_j, given prior observations at $s_1, \ldots, s_{j-1}$. This is the matrix that is subtracted as the second term of Equation (10) in the Kalman filter update step.

Definition 3. For arbitrary square matrices $\mathbf{M}$ and $\mathbf{N}$ of the same dimension n, for any $\mathbf{a} \in \mathbb{R}^n$, we let $\mathbf{M} \succeq_a \mathbf{N}$ mean that $\mathbf{a}'(\mathbf{M} - \mathbf{N})\mathbf{a} \geqslant 0$.

Definition 3 is similar to the matrix equivalent of "greater than" for scalars, but is tied to a specific multiplier $\mathbf{a}$. The final definition will enable us to define a relationship between the cumulative covariance estimate adjustment over the entire schedule, π_n, of systems with different observation schedules.

Definition 4. We call a matrix sequence, $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$, *a-monotone* if $\mathbf{A}_n \succeq_a \mathbf{A}_{n-1} \succeq_a \cdots \succeq_a \mathbf{A}_1$.

It can be shown that systems with uncorrelated noise components have the a-monotonicity property for the sum of covariance estimate adjustments. For correlated noise, this property is difficult to show analytically but can be checked numerically for any system using some simple code (we used Matlab). This has been checked and clearly holds for the system parameterized by our clinical trial data described in §5. In discussions with our clinical collaborators, it is expected that this property will hold for a variety of chronic diseases. The following lemma, which is proved in the online appendix, shows that if more patient observations are available to the system the covariance will be smaller in the direction of progression.

Lemma 1. *Let $\pi_m \in \Pi_m([t_s, t_f])$ and let $\pi_n = \pi_m \cup \pi_{n-m} \in \Pi_n([t_s, t_f])$ be a policy that calls for all the observations of π_m but also has an additional $n - m$ observations within the interval (t_s, t_f). Under the assumption that the matrix sequence $(\sum_{j=2}^{k} \mathbf{C}_j^{\pi_k}$ for $k = 2, 3, \ldots$ such that $\pi_2 \subset \pi_3 \subset \cdots \subset \pi_k)$ is a-monotone in k, the covariance matrix $\Sigma^{\pi_m}_{t_f|t_f} \succeq_a \Sigma^{\pi_n}_{t_f|t_f}$ for $n > m$.*

The result from Lemma 1 supports both Property 2—more patient observations correlates with more certainty about whether the patient has progressed—and Property 3—the length of the testing interval is shorter when the system has less information about a patient.

Property 3 (Number of Observations vs. Testing Frequency) *shows that the length of the testing interval is shorter (i.e., tests are scheduled more frequently) when the system has less information. This property mirrors physician behavior in that a glaucoma specialist will often see the patient more frequently when they have less information about the patient (e.g., a new patient), but if the patient has been stable for a long time the specialist will begin to increase the interval between tests. The following theorem, proved in the online appendix, supports this intuition analytically.*

Theorem 3. *Given open loop testing policies $\pi_n \in \Pi_n([t_s, t_f])$ and $\pi_m \in \Pi_m([t_s, t_f])$ such that $n > m$ and $\pi_m \subset \pi_n$, under the assumption that the covariance estimate updates are a-monotone, $F_{\rho,\tau}(\hat{\alpha}_{t|t}, \hat{\Sigma}^{\pi_m}_{t|t}) \leqslant F_{\rho,\tau}(\hat{\alpha}_{t|t}, \hat{\Sigma}^{\pi_n}_{t|t})$, where $F_{\rho,\tau}(\cdot, \cdot)$ is given by Equation (20).*

Property 4 (Disease State vs. Testing Frequency) *shows that a patient who is "worse off" will be tested more frequently than a patient who is "doing well." The following theorem supporting Property 4 is proved in the online appendix.*

Theorem 4. *Given two patients at time t with mean state vectors $\hat{\alpha}_1$ and $\hat{\alpha}_2$ and covariance matrices $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, if $\mathbf{a}'\hat{\alpha}_1 > \mathbf{a}'\hat{\alpha}_2$ and $\hat{\Sigma}_1 \succeq_a \hat{\Sigma}_2$ then patient 1 will be tested no later than patient 2.*

The next section illustrates how our approach can benefit clinicians by applying the POMP TNT algorithm to two 10+ year clinical trials (AGIS and CIGTS).

5. POMP TNT Algorithm Applied to AGIS and CIGTS Clinical Trials

We begin by describing the design of the experiment and then we present the results comparing the POMP TNT algorithm with fixed-interval schedules that are common in practice. Starting from a cost-based optimization in which there are costs for testing and costs for missed progression, we identify a simple three-zone aggressiveness scale that allows clinicians to tailor their treatment to match the needs of a patient in a manner that is simple and can be related to a traditional fixed-interval testing approach. We then compare the associated Pareto improving schedules with fixed-interval testing schemes and age-based threshold policies.

5.1. Model Usage and Design of Experiment

Data and model parameterization using AGIS and CIGTS clinical trial data are described in §3.2. After parameterizing the Kalman filter and ProP function with the training set as described in §§3.2 and 3.3, POMP TNT was used to dynamically generate a monitoring schedule for each patient in the test data set. For both POMP TNT and fixed-interval methods, the scheduling process was terminated either at the end of the trial or when progression was detected, where progression was determined by the criteria described

in §3.3. Based on input from our clinical coauthors, we compared POMP TNT with fixed-interval schedules using three performance measures: (1) average number of tests per patient (number of tests, lower is better); (2) fraction of samples among progressing patients (our data are discrete and form a sequence of measurement samples spaced apart by six month intervals) at which the data indicate progression and for which the algorithm called for a test (accuracy, higher is better); (3) average number of periods (where a period is six months) that a patient's progression went undetected (diagnostic delay, lower is better).

The POMP TNT algorithm has two parameters that influence the testing aggressiveness: (1) the threshold τ for determining whether progression has occurred using the logistic regression from Equation (13) and (2) the size ρ of the prediction region (i.e., confidence level). Using the training data, for each interval length (i.e., 1, 1.5, and 2 years) we found the τ and ρ combination that generated a POMP TNT schedule with approximately the same average number of tests per patient as the corresponding fixed-interval schedule while either (1) maximizing accuracy or (2) minimizing diagnostic delay. To do so, we performed a two-dimensional search on the training data as follows. Let $TPP_{TNT}(\tau, \rho)$ be the average number of tests per patient for the training data set using the POMP TNT algorithm with parameters τ and ρ. Let TPP_n be the average number of tests per patient for fixed interval testing with testing interval length of n years. Now for each interval length $n = 1, 1.5, 2$ years we perform the following two steps:

1. For each τ on a discrete grid between 0 and 1, compute $\rho_n(\tau)$ as the largest value of ρ (also on a discrete grid between 0 and 1) such that $TPP_{TNT}(\tau, \rho_n(\tau)) \leqslant TPP_n$.

2. Find $\tau_n^* = \arg\min_\tau DiagnosticDelay(\rho_n(\tau))$, where τ is optimized over the discrete grid from (1) above.

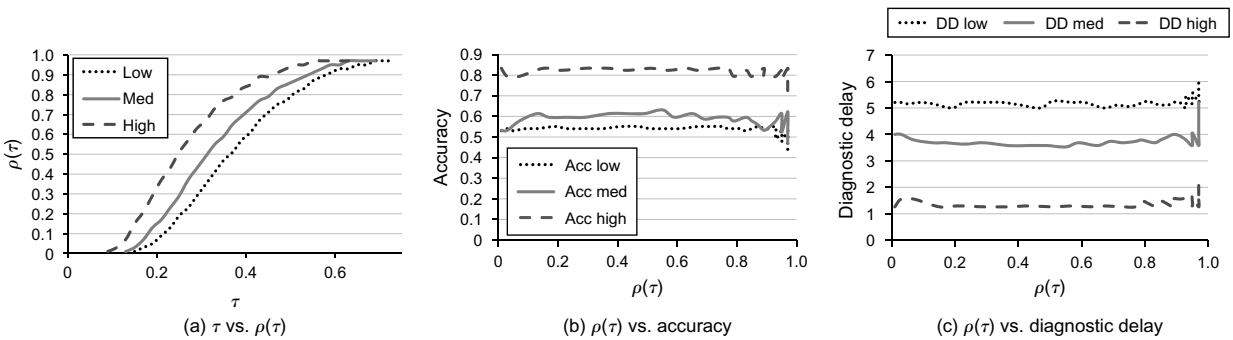
The same search can be performed to maximize accuracy. Finding $\rho_n(\tau)$ can be done very quickly using a binary search because of the fact that, for any given value of τ, the number of tests per patient is monotone increasing in the size of ρ. Monotonicity in ρ can be verified quickly by considering Equations (18)–(20). As will be seen in §§5.2 and 5.3, we align the POMP TNT parameters with these three fixed intervals common in current practice, $n = 1, 1.5$, and 2 years, to enable us to suppress the ρ and τ parameters and instead provide clinicians with aggressiveness levels (zones) that they can adjust to tailor their treatment: low $(\tau_2^*, \rho_2(\tau_2^*))$, med $(\tau_{1.5}^*, \rho_{1.5}(\tau_{1.5}^*))$, and high $(\tau_1^*, \rho_1(\tau_1^*))$.

Robustness of Parameter Choice. A nice property of this approach to choosing algorithm parameters ρ and τ is that the accuracy and diagnostic delay resulting from the choice are relatively insensitive to the initial choice of τ and to the combination of $(\tau, \rho(\tau))$ for each regime, n. Figure 4(a) shows how ρ changes with τ for each level of aggressiveness (low, medium, high) as a result of the search described above. Figures 4(b) and (c) show the key performance metrics of accuracy and diagnostic delay, which for each level of aggressiveness, are very robust to the choice of τ and ρ. As long as the initial τ is selected from a large range in the middle—not too close to zero or one—the accuracy and diagnostic delay resulting from $(\tau_n, \rho_n(\tau_n))$ is nearly identical for any initial choice of τ. From our experiments applying the algorithm to the two clinical trial data sets (see Figure 4), it appears that algorithm is robust in terms of accuracy and diagnostic delay for any choice of $\rho(\tau)$ between 0.2 and 0.8.

5.2. A Cost Model and Zone-Based Method for Clinician Model Control

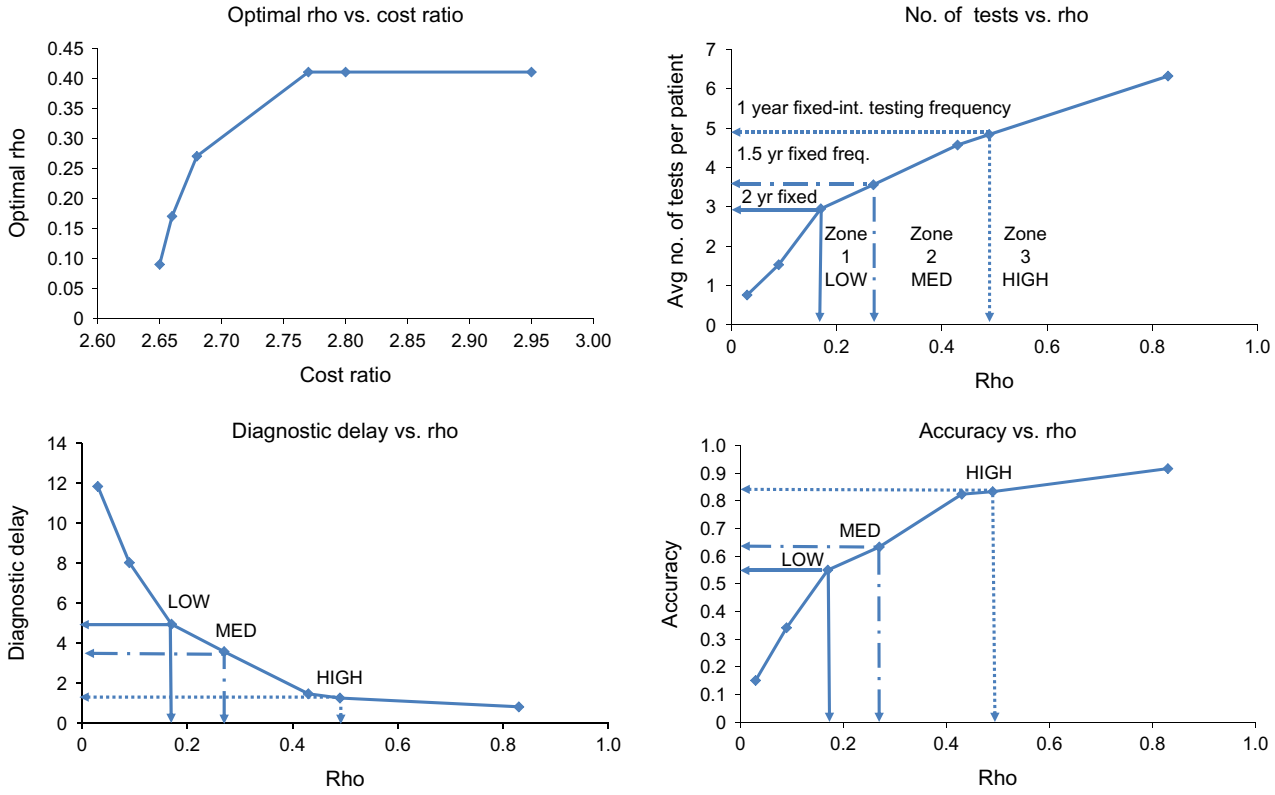
From a healthcare policy perspective, it is important to consider the trade-off between the cost of undetected glaucoma progression (per unit of time), c_p, and the cost per test performed, c_t. Let $\pi = c_p/c_t$ be the cost ratio of progression cost to testing cost. A low cost ratio implies a desire to avoid overtesting, whereas a high cost ratio implies a preference for more aggressive testing in the hopes of early detection of progression. As mentioned previously, different patients may experience differently the discomfort of both testing and the disease's symptoms. The cost ratio can capture the sense of how burdensome the testing procedures are relative to disease progression and produce a schedule tailored to each patient's preference. This ratio can also be

Figure 4. The robustness of parameter choices for ρ and $\rho(\tau)$ is presented for low, medium, and high aggressiveness settings as follows: (a) the calculated $\rho(\tau)$ for each value of τ, (b) the accuracy of setting the time to next test at a stage in which progression occurred vs. $\rho(\tau)$, (c) the diagnostic delay vs. $\rho(\tau)$.



(a) τ vs. $\rho(\tau)$ (b) $\rho(\tau)$ vs. accuracy (c) $\rho(\tau)$ vs. diagnostic delay

Figure 5. (Color online) Performance measures as a function of the cost ratio.



used by the clinician to capture how aggressively they feel the disease should be treated in each individual patient.

We assessed the total cost per patient as $c_p \times (\textit{diagnostic delay}) + c_t \times (\textit{no. of tests})$. Using a similar procedure to the one described in §5.1, we performed a search on the training set to identify the ρ and τ combination that minimized the average total cost per patient for each cost ratio. The upper left graph in Figure 5 presents the conversion between the cost ratio and the optimal value of ρ, where the optimal ρ and τ were determined from the training data. If testing was allowed on a continuous time scale then one would expect this plot to be monotonically increasing. However, because it is only possible to test at discrete time points (six month intervals) the same ρ value may be optimal for multiple cost ratios. Further, ρ and τ are determined jointly by maximizing performance on the training data. For these reasons, it is possible for the same rho to be optimal for different cost ratios. In the remaining three graphs of Figure 5 we plot the performance metrics on the testing data versus the ρ values obtained from the cost optimization.

In the upper right graph of Figure 5, we have marked three ρ zones related to how aggressively the algorithm will test a patient: low, medium, and high. These zones are found by comparing the testing frequency with the frequency of the 1, 1.5, and 2 year fixed interval testing schemes. The frequencies of the fixed-interval testing schemes are shown with the three different arrow types (zone three is reached at a cost ratio beyond the upper limit shown in the upper left plot). The two year fixed schedule does not result in an exactly proportional reduction in the number of tests per year because of the nature of the end of horizon effects of our CIGTS and AGIS data sets. The result suggests an intuitive zone-based method for adjusting the POMP TNT algorithm to tailor the testing schedule to each patient. These zones will be investigated in the next section as a simple three-zone system for clinician interactive model control.

Remark 1. According to our clinical collaborator, a high aggressiveness testing schedule would likely be aligned with six month testing intervals, medium to one year, and low to two years; however, CIGTS and AGIS data are available only every six months, so a six month testing scheme would not yield meaningful results for comparing POMP TNT with fixed-interval testing. Although more frequent VF testing can be appropriate for some patients, this is uncommon in part because such tests are time consuming and tiring to the patients (see Glen et al. 2014). In practice, most of the patients have no more than one to two tests per year on average (see Fung et al. 2013, Stein et al. 2012). Moreover, frequent fixed-interval testing regimens (e.g., four month and six month) have been shown to be worse at estimating rates of MD loss and glaucoma progression than variable-spaced follow-ups (see Crabb et al. 1997), so this evidence and other studies provide further

support to the need to investigate personalized follow-up schemes (see Jansonius 2006, 2007).

5.3. Pareto Improving Schedules

In this section we show how the POMP TNT algorithm dominates both fixed-interval schedules as well as an optimal age-based threshold policy in a Pareto sense. To test the fixed-interval schedules, which we call "FI," we scheduled tests at fixed frequencies (i.e., periods of 1, 1.5, and 2 years). These intervals were chosen because they are multiples of six months, ensuring that whenever a test was called for, the data in the data set were available to evaluate whether or not the patient had progressed, and hence whether or not any given test caught an instance of progression and, if so, how quickly it did so. Given the time step of the data, it is not very informative to consider a six month fixed interval because this implies testing in every possible period and its accuracy and diagnostic delay cannot be analyzed under our current definitions. If the data were available every three months, then we would update our definition of accuracy and diagnostic delay, leading a six month fixed-interval scheme to achieve only an accuracy of 50% and diagnostic delay of 1.5 months. If data were available, we could also update the time step for the Kalman filter to be three months instead of six months and be able to compare the two methods, though unfortunately data were not available every 1.5 months.

VF follow-ups longer than one year are common in practice, as discussed by Stein et al. (2012). Additionally, we have been able to further support our fixed-interval choices of 1, 1.5, and 2 years from data that we collected from patients being treated at the Kellogg Eye Center by various clinicians. We randomly selected 34 patients seen at the Kellogg Eye Center between January 1, 1990 and July 31, 2013 with similar physiological characteristics as the patients upon which our models were parameterized. IRB was obtained for this study, and all data were manually entered with two people analysts assigned to each entry session to ensure reliability of the information gathered and avoid data entry errors. The median time in between readings was 370 days (i.e., 50% of the patients had over a year in between VF readings), and all patients sampled had over six months in between VF readings (the minimum was 217 days between readings). This suggests that our selection of 1 year, 1.5 years, and 2 years is a good benchmark for comparing our algorithms with current practice.

To avoid introducing any bias, we varied when the first test of the sequence began, alternately choosing the first test to be the patients first visit, second visit, third visit, and so on. We also tested a policy that used information on the patient's age as well. This policy employs two age thresholds dividing the patients into young, middle age, and old age groups. Each group was assigned its own testing frequency. For example, one may start out testing once every six months in the young group, then switch to once a year for the middle-aged group, and finally to once every two years for the older group. To find the best such set of policies, which we term "OPT TH," we performed an exhaustive search over the training data by changing (1) the two age thresholds that divided the three groups, and (2) the testing frequencies assigned to each group. We were then able to find the thresholds and frequency assignments that performed efficiently (lying along the Pareto frontier). As with the FI schedules, we varied the starting time of the first test within each age group to avoid bias introduced by choosing an arbitrary starting point for the sequence of tests.

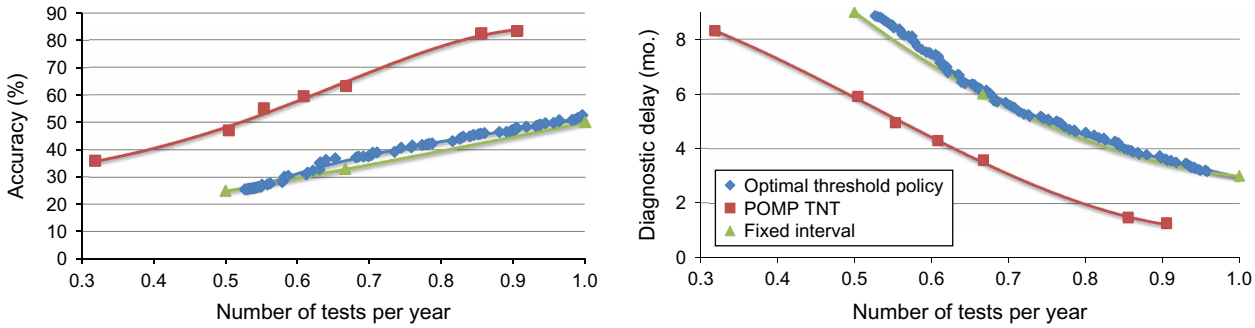
Whereas choosing continuous variables ρ and τ in applying POMP TNT requires a greater depth of understanding, choosing a level of aggressiveness from among predefined zones (i.e., low, med, high) is both intuitive and easy. Further, aligning the zones with a particular fixed-interval schedule allows clinicians to relate the zones to their previous experiences in treating patients. This increases the ease of adoption into clinical practice. For the purposes of this paper, it is appropriate to define the terms *low*, *medium*, *high* to specifically refer to levels of aggressiveness in patient testing that have an equivalent average frequency to the 1, 1.5, and 2 year fixed testing intervals. Table 2 displays the accuracy and diagnostic delay (averaged across all patients in the test data set) of FI schedule (1, 1.5, 2 years), the optimal threshold policies (OPT TH), and the corresponding POMP TNT schedule (high, med, low).

Table 2 shows that POMP TNT dominates the 1, 1.5, and 2 year fixed-interval schedules by providing higher accuracy and lower diagnostic delay with close to the same testing frequency on average. Perhaps surprisingly, the optimal age-based threshold policies barely outperform fixed-interval policies. This implies that the information update from new test results used to make POMP TNT testing decisions has a greater impact than the age of the patient used in isolation. Also note that POMP TNT yields better accuracy than the one year fixed interval schedule

Table 2. Performance of fixed-interval testing schedules and POMP TNT algorithm.

	High freq (1 yr)			Med freq (1.5 yr)			Low freq (2 yr)		
	FI	OPT TH	TNT	FI	OPT TH	TNT	FI	OPT TH	TNT
No. of tests per year	1.0	0.96	0.91	0.67	0.66	0.67	0.50	0.53	0.55
Accuracy%	50	50	83	33	37	63	25	26	55
Diag. delay (mo.)	3.0	3.17	1.26	6.00	6.15	3.58	9.00	8.86	4.95

Figure 6. (Color online) Fixed-interval and POMP TNT accuracy and diagnostic delay versus average tests per patient.



while using approximately as few tests as the two year fixed interval schedule. The Pareto curve in Figure 6 further suggests that POMP TNT in fact would be able to dominate all fixed-interval schedules and optimal threshold policies across all dimensions. By adjusting model parameters in Figure 6, we designed schedules over a finer range of levels of aggressiveness than simply low, medium, and high. For equivalent frequency FI and OPT TH schedules, the POMP TNT schedule with the equivalent number of tests per patient yields between 30%–33% *increase* in accuracy and between 40%–58% *decrease* in diagnostic delay. As a final note, the model was tested on different subgroups of patients. The algorithm outperformed yearly fixed-interval testing for all subgroups. As expected, it scheduled more tests per year on progressing (versus nonprogressing) patients, on AGIS (versus CIGTS) patients and on African-American (versus Caucasian) patients (significant at $p = 0.05$, full results can be found in Schell et al. 2014).

We acknowledge that, if a sufficiently low cost of testing is provided (or a high cost of progression), our algorithm will eventually call for testing every six months (or at every period at which testing is possible). We also acknowledge that a fixed-interval schedule may be preferred by some patients and/or practitioners. For example, the patient may prefer knowing that he or she will always be tested every 12 months rather than having to remember when he or she will be tested next. On the other hand, it may be easier for providers to forecast resource utilization if all patients are tested at fixed intervals of time. Also, some clinicians may prefer testing all patients at fixed intervals of time, rather than having to rely on algorithms to determine when patients should be tested next. Note that for every dynamic policy, there exists a fixed-interval policy with frequent monitoring, which performs better than the dynamic policy (e.g., one in which the monitoring interval is equal to the shortest possible time step in the dynamic policy). Therefore, for convenience, physicians can continue to use fixed-interval policies. Notice, though, that testing at fixed and frequent intervals of time will come at the cost of increased testing and hence cost.

6. Integrating a Data-Driven Decision Support Tool Into Glaucoma Clinical Practice

In treating patients with open-angle glaucoma, clinicians are faced with the task of quickly and efficiently processing the results from a number of quantitative tests including visual fields, intraocular pressures, and results from structural measurements of glaucoma such as optical coherence tomography (a topic for future research). Current practice often requires ophthalmologists or optometrists to make gestalt judgments based on their experience and expertise as to whether glaucoma progression has occurred or not and when future testing should be performed.

Our glaucoma decision aid tool could enhance current practice by providing clinicians with personalized, dynamically updated recommendations regarding follow-up visits and diagnostic testing. For each glaucoma patient, this method would compute the probability of progression of the patient as a function of time in the future. In practice, past test results from a patient's medical record would be entered into the POMP TNT tool. As new tests are taken at subsequent visits, these test results too would be entered into the tool, which would update in real time the Kalman filter model estimates of key variables used to estimate the future probability of progression over time. The decision support tool would provide the eye care provider with (1) the current ProP rating (signaling whether or not progression has occurred at that particular visit), and (2) a suggested time length into the future for the patient to return for their next VF and IOP tests, depending on the aggressiveness level (e.g., low, medium, or high) that the clinician and patient deem appropriate. If greater detail is desired, our method can forecast the projected ProP trajectory years into the future (with estimates on the variance of these forecasts). This tool could be further enhanced in the future to incorporate additional parameters not presently available to the investigators, e.g., central corneal thickness and OCT results (see De Moraes et al. 2011).

Starting from a cost-based optimization in which there are costs for testing and costs for missed progression, we identified a simple three-zone aggressiveness scale that allows clinicians to tailor their treatment to the individual

patient. For each aggressiveness level, the tool would use corresponding model parameters, which are precomputed by the analysis software based on historical data at the population level (e.g., see Figure 5 for an example of parameters set to match on average the expected testing frequency of a fixed-interval schedule). The clinician may choose the aggressiveness level based on their clinical experience in managing patients with glaucoma along with consultation with the patient of his or her preferences. One guiding factor in choosing an aggressiveness level could be, for example, how patients feel about their disease (e.g., Burr et al. 2007). Factors that may go into the decision include the age of the patient, the underlying severity of the glaucoma and perceived likelihood of progression to blindness (patients with more severe disease tend to require closer monitoring), the status of the other eye (monocular patients are often monitored more aggressively), the general health of the patient (patients who have limited life expectancy may not require aggressive monitoring as they will likely die before they go blind from glaucoma), input from the patient (some patients may be unwilling to undergo frequent monitoring or live many hours away from the eye care professional and it would be infeasible to monitor them very aggressively), and other factors. Clinicians are trained to make just such assessments and to choose an aggressiveness level of monitoring that is appropriate for each patient. Combining expert judgment, consultation with, and knowledge of the patient, the clinician may determine how aggressively to test the patient. Our framework would then assist in the decision of when to schedule the tests based on the desired level of aggressiveness. Notice that these are decisions that clinicians already routinely make in conjunction with their patients. Since each aggressiveness level is associated with a fixed-interval testing frequency, the clinicians are able to relate this choice back to testing schedules that they are familiar with. Although we have mapped aggressiveness levels to cost ratios, future work may also link the choice of aggressiveness level to the glaucoma-specific health status of the patient and the patients utility/disutility from the disease (such as Burr et al. 2007). The analytics our algorithm provides can support and inform these decisions.

6.1. Model Limitations and Areas for Future Exploration

There are various limitations that are either not considered in this work or not possible using existing systems science. The first limitation lies in the scope of factors that are incorporated in our glaucoma decision aid. Although factors such as medical comorbidities (lower systolic perfusion pressure, lower systolic blood pressure, cardiovascular disease), central corneal thickness, and presence of beta-zone parapapillary atrophy have been found in some studies and univariate analyses to be risk factors associated with glaucoma progression, these data were not available in the data sets used in the present analysis. Other factors, such as bilaterality of disease, exfoliation syndrome (a subtype of glaucoma), and presence of hemorrhages around the optic disc were available in only a subset of patients that we had access to, and were found not to be significant in predicting progression through backward and forward elimination. Although we were not able to account for some factors considered in other studies, we were able to incorporate demographic characteristics of the patients such as age and race into the algorithm and to show that other factors did not improve algorithm performance. Even without these additional factors, the algorithm outperforms yearly fixed-interval and optimal age-based dynamic interval testing strategies. In the future we hope to acquire other data sources that contain information regarding some of the additional risk factors to incorporate them into the decision aid tool. Additionally, data availability only at six month intervals prevented us from testing our approach on finer intervals. As the testing intervals become shorter, the benefit of our approach in terms of absolute accuracy relative to a fixed-interval policy may decrease. Along these lines, it should be noted that for every dynamic policy, there is a fixed policy that will perform better in terms of accuracy. For example, one could set the fixed interval to be the smallest step of the dynamic policy to achieve better accuracy, though more frequent testing will come at higher costs.

Second, we performed our analysis on data from patients who agreed to participate in a randomized clinical trial. Though we would not expect substantial differences in performance on other patients with glaucoma who are receiving care outside the setting of a clinical trial, we acknowledge that participants in clinical trials may be a biased sample. The fact that POMP TNT performed well on participants in two different clinical trials, though, suggests that it should perform well on patients with different severities of glaucoma. Additional work is required to validate our decision aid tool on patients who are receiving care outside a clinical trial setting, especially those with tests taken at varying time intervals. Third, we note that the assumption of Gaussian noise is necessary to perform the computations of our linear Gaussian systems model. While we validated the Gaussian assumption for our clinical trials patients, it is possible that other systems may not follow strictly Gaussian noise distributions. In this case, the Kalman filter remains unbiased but the estimator no longer minimizes the variance of the estimate, therefore the resulting schedules would be more conservative (higher variance means more frequent testing). Thus, in this case, some efficiency would be sacrificed, but the patients would benefit from earlier detection. Fourth, we do not directly consider patient utilities (for testing versus progression), and we would rather leave this subjective assessment up to the clinician when they choose whether to use a low, medium, or high aggressiveness parameter setting to incorporate into when next to test each patient. Though one might try to estimate patient utility functions directly, we feel that it is

best for a clinician to decide this based on their knowledge of each patient and his or her circumstances. Fifth, we consider patient heterogeneity through updating disease trajectories through the tests that are received over time. The underlying transition matrix is not changed over time. Including a learning component in the transition dynamics is the subject of future research. Sixth, we use a regression-based smoothing method to estimate velocities and accelerations. While these functioned sufficiently well in our case study, other methods such as fitting splines to the data may prove useful in this and other contexts. Finally, in our analysis, we assume full compliance with the schedule generated and do not specifically model compliance in our algorithm. This is out of scope for this work but represents an area for future exploration.

7. Conclusions and Future Work

This paper contributes a new modeling paradigm for the monitoring of glaucoma and other chronic diseases. In contrast to disease detection models, chronic diseases often require monitoring a number of key physiological indicators that provide rich and dynamic information about a patient's changing condition. To take full advantage of this data rich environment, we developed a multivariate state space model of disease progression based on the Kalman filter to forecast the disease trajectory. Then the ProP function was optimized over the Gaussian density of the Kalman filter to determine the TNT.

Beyond the ability to handle multidimensional state spaces, a key benefit of this approach is that the model output summarizes the full distribution on the patient's current state via the mean vector and the covariance matrix of a Gaussian random variable. This allows the incorporation of both patient system noise and testing noise into the state space model and yields a far richer characterization of the patient's health state than simpler estimation and forecasting methods. Our decision support approach is flexible enough to allow clinician judgment by setting model aggressiveness levels to complement their medical knowledge with the advanced statistical predictions. This approach will benefit both eye care professionals and their glaucoma patients, and it will potentially translate to other chronic diseases.

Our validation study was based on data from the two 10+ year clinical trials, CIGTS and AGIS. It demonstrated that POMP TNT was able to outperform fixed-interval regimens in terms of accuracy—30%–33% *better* than comparable fixed-interval schedules—and diagnostic delay—40%–58% *better*. This confirms a hypothesis within the medical community that variable intervals may in fact outperform fixed-interval testing. POMP TNT also provides a rigorous, analytical tool for harnessing large amounts of historical data to determine the appropriate variable interval lengths between tests. We believe that this research approach will be useful to clinical practice and provide a theoretical framework for addressing the unique features of *monitoring problems*.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/opre.2015.1405.

Acknowledgments

The authors gratefully acknowledge the statistical analyses performed by Greggory Schell. The authors thank the anonymous reviewers and the associate editor for many valuable comments that have led to significant improvements in the paper. This research was supported in part by NSF [Grant CMMI-1161439] and Clinical and Translational Science Awards (CTSA) from NIH [Grant MICHR (UL 1RR024986)] .

References

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2004) The optimal timing of living-donor liver transplantation. *Management Sci.* 50(10): 1420–1430.

Alliance for Aging Research (2011) The economic burden of vision loss. Accessed September 1, 2011, http://www.silverbook.org/browse.php?id=84.

American Academy of Ophthalmology Glaucoma Panel (2010) Preferred practice pattern guidelines. Accessed January 10, 2011, http://www.aao.org/ppp.

Antelman GR, Savage IR (1965) Surveillance problems: Wiener processes. *Naval Res. Logist.* 12(1):35–55.

Athans M (1972) On the determination of optimal costly measurement strategies for linear stochastic systems. *Automatica* 8(4):397–412.

Ayer T, Alagoz O, Stout NK (2012) OR forum: A POMDP approach to personalize mammography screening decisions. *Oper. Res.* 60(5): 1019–1034.

Baker RD (1998) Use of a mathematical model to evaluate breast cancer screening policy. *Health Care Management Sci.* 1(2):103–113.

Barlow RE, Hunter LC, Proschan F (1963) Optimum checking procedures. *J. Soc. Indust. Appl. Math.* 11(4):1078–1095.

Barlow RE, Proschan F, Hunter LC (1996) *Mathematical Theory of Reliability* (Society for Industrial Mathematics, Philadelphia).

Bengtsson B, Patella VM, Heijl A (2009) Prediction of glaucomatous visual field loss by extrapolation of linear trends. *Arch. Ophthalmology* 127(12):1610–1615.

Bensing J (2000) Bridging the gap: The separate worlds of evidence-based medicine and patient-centered medicine. *Patient Ed. Counseling* 39(1):17–25.

Bertsekas DP (1987) *Dynamic Programming: Deterministic and Stochastic Models* (Prentice-Hall, Englewood Cliffs, NJ).

Bertsekas DP (2000a) *Dynamic Programming and Optimal Control*, Third ed., Vol. 1 (Athena Scientific, Belmont, MA).

Bertsekas DP (2000b) *Dynamic Programming and Optimal Control*, Vol. 2 (Athena Scientific, Belmont, MA).

Bloch-Mercier S (2002) A preventive maintenance policy with sequential checking procedure for a Markov deteriorating system. *Eur. J. Oper. Res.* 142(3):548–576.

Burr JM, Kilonzo M, Vale L, Ryan M (2007) Developing a preference-based glaucoma utility index using a discrete choice experiment. *Optometry Vision Sci.* 84(8):E797–E809.

CDC, Centers for Disease Control and Prevention (2013) Chronic disease prevention and health promotion. Accessed February 1, 2014, http://www.cdc.gov/chronicdisease/.

Chew V (1966) Confidence, prediction, and tolerance regions for the multivariate normal distribution. *J. Amer. Statist. Assoc.* 61(315): 605–617.

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

Chitgopekar SS (1974) A note on the costly surveillance of a stochastic system. *Naval Res. Logist.* 21(2):365–371.

Choplin NT, Edwards RP (1999) *Visual Field Testing with the Humphrey Field Analyzer: A Text and Clinical Atlas* (Slack, Thorofare, NJ).

Crabb DP, Fitzke FW, McNaught AI, Edgar DF, Hitchings RA (1997) Improving the prediction of visual field progression in glaucoma using spatial processing. *Ophthalmology* 104(3):517–524.

D'Amato RM, D'Aquila RT, Wein LM (2000) Management of antiretroviral therapy for HIV infection: Analyzing when to change therapy. *Management Sci.* 46(9):1200–1213.

Day NE, Walter SD (1984) Simplified models of screening for chronic disease: Estimation procedures from mass screening programmes. *Biometrics* 40(1):1–13.

De Moraes CG, Juthani VJ, Liebmann JM, Teng CC, Tello C, Susanna R Jr, Ritch R (2011) Risk factors for visual field progression in treated glaucoma. *Arch. Ophthalmology* 129(5):562–568.

Denton BT, Alagoz O, Holder A, Lee EK (2011) Medical decision making: Open research challenges. *IIE Trans. Healthcare Systems Engrg.* 1(3):161–167.

Denton BT, Kurt M, Shah ND, Bryant SC, Smith SA (2009) Optimizing the start time of statin therapy for patients with diabetes. *Medical Decision Making* 29(3):351–367.

Derman C, Sacks J (1960) Replacement of periodically inspected equipment. (An optimal optional stopping rule). *Naval Res. Logist.* 7(4): 597–607.

Diaz-Aleman VT, Anton A, de la Rosa MG, Johnson ZK, McLeod S, Azuara-Blanco A (2009) Detection of visual-field deterioration by glaucoma progression analysis and threshold noiseless trend programs. *British J. Ophthalmology* 93(3):322–328.

Digalakis V, Rohlicek JR, Ostendorf M (1993) ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition. *IEEE Trans. Speech Audio Processing* 1(4): 431–442.

Donelson J (1977) Cost model for testing program based on nonhomogeneous Poisson failure model. *IEEE Trans. Reliability* 26(3):189–194.

Eckles JE (1968) Optimum maintenance with incomplete information. *Oper. Res.* 16(5):1058–1067.

Fowler FJ, Wennberg JE, Timothy RP, Barry MJ, Mulley AG, Hanley D (1988) Symptom status and quality of life following prostatectomy. *J. Amer. Medical Assoc.* 259(20):3018–3022.

Friedman DS, Wolfs RC, O'Colmain BJ, Klein BE, Taylor HR, West S, Leske MC, Mitchell P, Congdon N, Kempen J (2004) Prevalence of open-angle glaucoma among adults in the United States. *Arch. Ophthalmology* 122(4):532–538.

Fung SSM, Lemer C, Russell RA, Malik R, Crabb DP (2013) Are practical recommendations practiced? A national multi-centre cross-sectional study on frequency of visual field testing in glaucoma. *British J. Ophthalmology* 97(7):843-847.

Gardiner SK, Demirel S (2008) Assessment of patient opinions of different clinical tests used in the management of glaucoma. *Ophthalmology* 115(12):2127–2131.

Ghahramani Z, Hinton GE (1996) Parameter estimation for linear dynamical systems. University of Toronto Technical Report CRG-TR-96-26.

Glen FC, Baker H, Crabb DP (2014) A qualitative investigation into patients views on visual field testing for glaucoma monitoring. *BMJ Open* 4(1):e003996.

Hanin AD, Yakovlev AY (2001) Optimal schedules of cancer surveillance and tumor size at detection. *Math. Comput. Model.* 33(12–13): 1419–1430.

Hodapp E, Parrish II RK, Anderson DR, Perkins TW (1993) *Clinical Decisions in Glaucoma* (Mosby, St. Louis).

Hu C, Lovejoy WS, Shafer SL (1996) Comparison of some suboptimal control policies in medical drug therapy. *Oper. Res.* 44(5):696–709.

Jansonius NM (2006) Towards an optimal perimetric strategy for progression detection in glaucoma: From fixed-space to adaptive inter-test intervals. *Graefe's Arch. Clinical Experiment. Ophthalmology* 244(3): 390–393.

Jansonius NM (2007) Progression detection in glaucoma can be made more efficient by using a variable interval between successive visual field tests. *Graefe's Arch. Clinical Experiment. Ophthalmology* 245(11):1647–1651.

Kalman RE (1960) A new approach to linear filtering and prediction problems. *J. Basic Engrg.* 82(1):35–45.

Kander Z (1978) Inspection policies for deteriorating equipment characterized by *N* quality levels. *Naval Res. Logist.* 25(2):243–255.

Kander Z, Raviv A (1974) Maintenance policies when failure distribution of equipment is only partially known. *Naval Res. Logist.* 21(3): 419–429.

Katz J (1999) Scoring systems for measuring progression of visual field loss in clinical trials of glaucoma treatment. *Ophthalmology* 106(2):391–395.

Keller JB (1974) Optimum checking schedules for systems subject to random failure. *Management Sci.* 21(3):256–260.

Kirch RLA, Klein M (1974) Surveillance schedules for medical examinations. *Management Sci.* 20(10):1403–1409.

Lafortune S (1985) *On Stochastic Optimal Control Problems with Selection Among Different Costly Observations* ((Memorandum) Electronics Res Lab, COE, University of California, Berkeley).

Lavieri MS, Puterman ML, Tyldesley S, Morris WJ (2012) When to treat prostate cancer patients based on their PSA dynamics. *IIE Trans. Healthcare Systems Engrg.* 2(1):62–77.

Lee EK, Wu TL (2009) Classification and disease prediction via mathematical programming. *Handbook of Optimization in Medicine* (Springer, New York), 1–50.

Lee CP, Chertow GM, Zenios SA (2008) Optimal initiation and management of dialysis therapy. *Oper. Res.* 56(6):1428–1449.

Lee PP, Walt JW, Rosenblatt LC, Siegartel LR, Stern LS (2007b) Association between intraocular pressure variation and glaucoma progression: Data from a United States chart review. *Amer. J. Ophthalmology* 144(6):901–907.

Lee PP, Levin LA, Walt JG, Chiang T, Katz LM, Dolgitser M, Doyle JJ, Stern LS (2007a) Cost of patients with primary open-angle glaucoma: A retrospective study of commercial insurance claims data. *Ophthalmology* 114(7):1241–1247.

Lee PP, Walt JG, Doyle JJ, Kotak SV, Evans SJ, Budenz DL, Chen PP, et al. (2006) A multicenter, retrospective pilot study of resource use and costs associated with severity of disease in glaucoma. *Arch. Ophthalmology* 124(1):12–19.

Lincoln TL, Weiss GH (1964) A statistical evaluation of recurrent medical examinations. *Oper. Res.* 12(2):187–205.

Luss H (1976) Maintenance policies when deterioration can be observed by inspections. *Oper. Res.* 24(2):359–366.

Luss H (1983) An inspection policy model for production facilities. *Management Sci.* 29(9):1102–1109.

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

McNaught AI, Hitchings RA, Crabb DP, Fitzke FW (1995) Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefe's Arch. Clinical Experiment. Ophthalmology* 233(12):750–755.

Mehra RK (1976) Optimization of measurement schedules and sensor designs for linear dynamic systems. *IEEE Trans. Automatic Control* 21(1):55–64.

Meier III L, Peschon J, Dressler R (1967) Optimal control of measurement subsystems. *IEEE Trans. Automatic Control* 12(5):528–536.

Michaelson JS, Halpern E, Kopans DB (1999) Breast cancer: Computer simulation method for estimating optimal intervals for screening. *Radiology* 212(2):551–560.

Mine H, Kawai H (1975) An optimal inspection and replacement policy. *IEEE Trans. Reliability* 24(5):305–309.

Morey RC (1966) Some stochastic properties of a compound-renewal damage model. *Oper. Res.* 14(5):902–908.

Munford AG, Shahani AK (1972) A nearly optimal inspection policy. *Oper. Res. Quart.* 23(3):373–379.

Murphy K (1998) *Kalman Filter Toolbox for Matlab* (Comp Sci and AI Lab., MIT, Cambridge, MA).

Musch DC, Gillespie BW, Lichter PR, Niziol LM, Janz NK (2009) CIGTS study investigators. Visual field progression in the collaborative initial glaucoma treatment study the impact of treatment and other baseline factors. *Ophthalmology* 116(2):200–207.

Musch DC, Gillespie BW, Niziol LM, Cashwell LF, Lichter PR (2008) Factors associated with intraocular pressure before and during 9 years of treatment in the collaborative initial glaucoma treatment study. *Ophthalmology* 115(6):927–933.

Nakagawa T, Osaki S (1974) Some aspects of damage models(cumulative processes in reliability physics). *Microelectronics Reliability* 13: 253–257.

Nakagawa T, Yasui K (1980) Approximate calculation of optimal inspection times. *J. Oper. Res. Soc.* 31(9):851–853.

NEI National Eye Institute (2011) Facts about glaucoma. Accessed March 12, 2012, http://www.nei.nih.gov/health/glaucoma/glaucoma_facts.asp.

Noonan GC, Fain CG (1962) Optimum preventive maintenance policies when immediate detection of failure is uncertain. *Oper. Res.* 10(3):407–410.

Nouri-Mahdavi K, Hoffman D, Coleman AL, Liu G, Li G, Gaasterland D, Caprioli J (2004) Predictive factors for glaucomatous visual field progression in the advanced glaucoma intervention study. *Ophthalmology* 111(9):1627–1635.

Ohnishi M, Kawai H, Mine H (1986a) An optimal inspection and replacement policy for a deteriorating system. *J. Appl. Probab.* 23(4):973–988.

Ohnishi M, Kawai H, Mine H (1986b) An optimal inspection and replacement policy under incomplete state information. *Eur. J. Oper. Res.* 27(1):117–128.

Oshman Y (1994) Optimal sensor selection strategy for discrete-time state estimators. *IEEE Trans. Aerospace Electr. Systems* 30(2):307–314.

Özekici S, Pliska SR (1991) Optimal scheduling of inspections: A delayed Markov model with false positives and negatives. *Oper. Res.* 39(2):261–273.

Pierskalla WP, Voelker JA (1976) A survey of maintenance models: The control and surveillance of deteriorating systems. *Naval Res. Logist.* 23(3):353–388.

Quigley HA, Broman AT (2006) The number of people with glaucoma worldwide in 2010 and 2020. *British J. Ophthalmology* 90(3): 262–267.

Rauner MS, Gutjahr WJ, Heidenberger K, Wagner J, Pasia J (2010) Dynamic policy modeling for chronic diseases: Metaheuristic-based identification of Pareto-optimal screening strategies. *Oper. Res.* 58(5): 1269–1286.

Rein DB, Zhang P, Wirth KE, Lee PP, Hoerger TJ, McCall N, Klein R, Tielsch JM, Vijan S, Saaddine J (2006) The economic burden of major adult visual disorders in the United States. *Arch. Ophthalmology* 124(12):1754–1760.

Rosenfield D (1976) Markovian deterioration with uncertain information. *Oper. Res.* 24(1):141–155.

Savage IR (1962) Surveillance problems. *Naval Res. Logist.* 9(3–4): 187–209.

Savage IR (1964) Surveillance problems: Poisson models with noise. *Naval Res. Logist.* 11(1):1–13.

Schell GJ, Lavieri MS, Van Oyen MP, Helm JE, Liu X, Musch D, Stein JD (2014) Using filtered forecasting techniques to determine personalized monitoring schedules for patients with open-angle glaucoma. *Ophthalmology* 121(8):1539–1546.

Schell GJ, Lavieri MS, Stein JD, Musch DC (2013) Filtering data from the collaborative initial glaucoma treatment study for improved identification of glaucoma progression. *BMC Medical Informatics Decision Making* 13(1):137–145.

Schuman JS, Puliafito CA, Fujimoto JG, Duker JS(2012) *Optical Coherence Tomography of Ocular Diseases*, Third ed. (Slack Incorporated, Thorofare, NJ).

Shechter SM, Alagoz O, Roberts MS (2010) Irreversible treatment decisions under consideration of the research and development pipeline for new therapies. *IIE Trans.* 42(9):632–642.

Shechter SM, Bailey MD, Schaefer AJ, Roberts MS (2008) The optimal time to initiate HIV therapy under ordered health states. *Oper. Res.* 56(1):20–33.

Sherif YS, Smith ML (1981) Optimal maintenance models for systems subject to failure—A review. *Naval Res. Logist.* 28(1):47–74.

Shwartz M (1978) A mathematical model used to analyze breast cancer screening strategies. *Oper. Res.* 26(6):937–955.

Stein JD, Talwar N, LaVerne AM, Nan B, Lichter PR (2012) Trends in use of ancillary glaucoma tests for patients with open-angle glaucoma from 2001 to 2009. *Ophthalmology* 119(4):748–758.

Tielsch JM, Sommer A, Witt K, Katz J, Royall RM, et al. (1990) Blindness and visual impairment in an American urban population: The Baltimore eye survey. *Arch. Ophthalmology* 108(2):286–290.

Wang H (2002) A survey of maintenance policies of deteriorating systems. *Eur. J. Oper. Res.* 139(3):469–489.

Wu W, Arapostathis A (2008) Optimal sensor querying: General Markovian and LQG models with controlled observations. *IEEE Trans. Automatic Control* 53(6):1392–1405.

Yeh RH (1997) Optimal inspection and replacement policies for multistate deteriorating systems. *Eur. J. Oper. Res.* 96(2):248–259.

Zahari M, Mukesh BN, Rait JL, Taylor HR, McCarty CA (2006) Progression of visual field loss in open angle glaucoma in the Melbourne visual impairment project. *Clinical Experiment. Ophtalmology* 34(1):20–26.

Zelen M (1993) Optimal scheduling of examinations for the early detection of disease. *Biometrika* 80(2):279–293.

Jonathan E. Helm is an assistant professor of operations and decision technologies at the Kelley School of Business, Indiana University. His research interests include capacity management in healthcare networks, patient flow modeling and optimization, patient monitoring for chronic disease, and distribution of medical aid in developing countries. Recently, he has been studying mechanisms to reduce the impact of hospital readmissions at both the operational and policy levels. His awards and honors include an invitation to give a Showcase Presentation at the 2014 and 2015 POMS CHOM Mini Conference, he was finalist (second place) in the 2013 INFORMS Data Mining Best Student Paper Competition, and also received a second place award in the INFORMS Public Sector OR best paper competition, first prize in the 2012 INFORMS "Doing Good with Good OR," first prize in 2011 and 2015 and finalist in 2012 for the POMS CHOM best paper competition.

Mariel S. Lavieri is an assistant professor in the Department of Industrial and Operations Engineering at the University of Michigan. Her research interests are in dynamic programming, stochastic control, optimization, data analytics, and partially observable state space models applied to healthcare. She is the recipient of the 2006 Bonder Scholarship, and an honorary mention in the 2010 George B. Dantzig Dissertation Award. She has also received the 2009 Pierskalla Award and the 2013 Young Participant with Most Practical Impact Award from the International Conference on Operating Research. She has guided work that won the 2012 INFORMS "Doing Good with Good OR Award" and the 2013 Medical Decision Making Lee Lusted Award.

Mark Van Oyen is a professor of industrial and operations engineering at the University of Michigan. His research includes the analysis, design, control, and management of operational systems, with emphasis on performance, flexibility, and controlled queueing network models. Healthcare operations research and medical decision making are key application areas. His awards with student advisees include first prize in the 2012 MSOM Student Paper Competition, the 2012 INFORMS "Doing Good with Good OR" first prize, and two finalist papers with MSOM. He also coauthored papers that won the 2010 Pierskalla Award, second prize in the 2013 POMS CHOM best paper competition, and first and second prize winning papers in the 2011 POMS CHOM competition. He was the IOE Department Faculty of the Year for 2008–2009, an ALCOA Manufacturing Systems Faculty

Fellow, and the 2003 Researcher of the Year at Loyola University Chicago's School of Business.

Joshua D. Stein is an associate professor in the Department of Ophthalmology and Visual Sciences at the University of Michigan. He is a board-certified ophthalmologist and a fellowship-trained glaucoma specialist with over 10 years of clinical experience caring for patients with different types and severities of glaucoma. In addition to his clinical experience, as a health services researcher some of his research interests include patterns of eye care in the United States, outcomes of glaucoma surgery, and quality of life of patients with ocular disorders.

David C. Musch is a professor in the Department of Ophthalmology and Visual Sciences and holds a joint appointment in the Department of Epidemiology at the University of Michigan. His research focuses on epidemiological studies (both observational and interventional designs) on common ophthalmic conditions, with a particular focus on glaucoma treatment, corneal transplantation outcomes, and studies of childhood eye disorders.