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

(Authors' names blinded for peer review)

Glaucoma is a chronic condition that causes blindness and affects an estimated 2.2 million Americans. An important clinical topic receiving increasing attention is the timing of periodic examinations for glaucoma patients. There is a clear tradeoff between monitoring intervals that are either too short (high cost) and too long (disease progression goes undetected). We address this question by integrating a dynamic linear Gaussian systems model of disease progression 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. We develop closed form solutions and study structural properties of our algorithm. In the process we develop a new framework that extends classical theory of positive semi-definiteness, quadratic forms and partial ordering of matrices. The insights from this work confirm the validity of changes to patient monitoring policies that have been proposed by some clinicians, but have yet to be rigorously validated. Based on data from a large-scale glaucoma clinical trial, we show that our methods significantly outperform current practice by achieving *greater accuracy* of identifying progression with *fewer examinations*. Our methodology is applicable to a variety of chronic diseases and can contribute to more efficient use of healthcare resources and more effective patient care.

Key words: linear Gaussian systems modeling, controlled observations, stochastic control, disease monitoring, medical decision making, glaucoma, visual field

---

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 (28)). 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 (28)).

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 tradeoff 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. With the movement towards patient-centered models of care (see (4)), monitoring guidelines that incorporate information from the patient's history are needed.

The standard for glaucoma care is to periodically measure intraocular pressure (IOP) and peripheral vision, as captured by visual field (VF) testing (see (10)) 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 (10)). 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 and burdensome, particularly for elderly patients (see (10)). 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 (2)). 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 US healthcare system (see (20)).

In this paper we develop advanced models and methods for determining the appropriate frequency of monitoring chronic disease customizable for each patient. Table 1 summarizes how our approach contributes to both theory and clinical practice.

In clinical practice, physicians typically monitor 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 insufficient to incorporate the richness of data involved in clinical decision making. This causes a problem for such traditional paradigms as Markov Decision Process (MDP), which suffer from the curse of dimensionality. We propose a novel approach using linear Gaussian systems to model disease progression. Our models can easily accommodate large state spaces and rich data inputs. These models are continuous state models with dynamics specified

	Traditional paradigm	This paper
Modeling Approach	MDP	Controlled Linear Gaussian System
State Space Scalability	Low	High
Stochasticity	System Noise	System Noise + Measurement Noise
Disease Dynamics	Population-based	Incorporates Patient Outcomes
Use of Patient Data	Low	High
Model	Static	Feedback Driven
Decision Informed By	Model	Model + Clinician Input
Clinical Value	Structural Insights	Structural Insights, Data-oriented, Flexible
Solution Approach	Optimization Techniques	Closed Form Solution Enables Fast Simple Search Techniques
Validation	Mostly Lit. Survey Based	Clinical Trial Data
Generalizability to Other Diseases	Low	High

Table 1 A comparison of our paradigm with traditional approaches to patient testing.

by a first order vector difference equations. By including in the state not only a test measurement itself, but its most important derivatives (e.g., second, third, and fourth in our model) and noise dynamics, existing methods can be harnessed to parameterize a rich model of progression capturing a wide range of dynamic behavior.

As demonstrated by the case of glaucoma, chronic disease monitoring typically involves both system noise (e.g. stochastic disease evolution) and measurement noise (e.g. testing errors). This is not easily captured in traditional approaches, yet these noise components are critical to capturing the true dynamics of chronic disease care. Our approach enables us to capture correlated multivariate white noise that is present in many medical tests, including VF tests and IOP tests.

This work also breaks ground by moving from a paradigm that is population-based to a paradigm that is focused on individual patient outcomes. Traditional approaches often formulate the problem using transition probabilities derived from a general population. Our approach incorporates not only a population-based a priori model, but also the information from each successive test to learn about the patient's disease dynamics and tailor the policy to the specific individual. This means that our model is driven by rich patient information that accumulates as more tests are scheduled. As will be seen, our model considers the value of obtaining only the necessary information about a patient to detect disease progression subject to the clinician's specified level of confidence about the disease state.

In contrast with traditional approaches, our approach produces a dynamic policy that is adaptable to specific patient needs. Traditional policies are often based on solving an optimization model off-line and using the output or structural insights to guide population-wide testing decisions in a static manner. While we also obtain structural insights, our algorithm is a feedback control-based

approach that dynamically updates the policy as additional patient data are obtained. Furthermore, our approach forecasts the patient disease state into the future until their is no longer statistical confidence that the forecasted disease state has not passed a progression threshold, at which point another test is recommended. Our decision support algorithm allows the clinician to control both the confidence level and the progression threshold to tailor the treatment to specific patient needs. This creates a sharp and clinically relevant contrast to traditional optimization approaches that produce a "one size fits all" policy, that is, a policy that treats two patients with the same symptoms in the same way. As noted in (12), two patients can experience the exact same symptoms very differently and therefore treatments should be tailored to the patient's experience of the symptoms and not just to the symptoms themselves. For example, a clinician would likely prescribe a different treatment approach for a sick elderly patient versus a young healthy patient if they had the same level of glaucoma.

Clinical relevance and acceptance also hinges on model validity. Traditional modeling efforts typically take a high level approach to validating their models by calculating transition probabilities based on population data often obtained from literature surveys. We take a more detailed approach by testing our algorithm on individual patients from a large 10 year clinical trial and comparing its outcomes with those based on currently accepted testing practices for the same patients.

Finally, the monitoring problem we address in this manuscript is not unique to glaucoma. There is a need for novel methodology to assist clinicians caring for patients with an array of medical conditions. Those 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 (blindness, kidney failure, stroke, heart attack, etc.), and (5) have quantifiable measures (protein level measurements, blood pressure measurements, viral load levels, etc.). Examples of chronic diseases for which physicians periodically monitor a number of quantifiable medical tests to capture progression include diabetes mellitus, connective tissue diseases, and kidney diseases. Given that chronic illnesses affect almost one out of every two adults in the United States (see (7)), the new ideas in this research have the potential for broad impact on cost to our health system 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 where 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 model to obtain an estimate of the current disease state as well as a

forecast of 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.

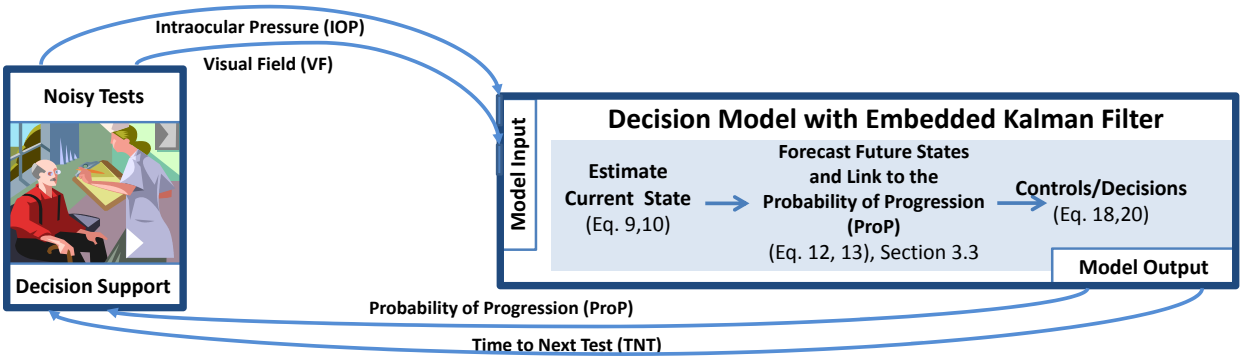


Figure 1 Decision support framework for chronic disease monitoring.

This paper's methodological contributions are derived from the analysis of the interaction of the ProP function with the linear Gaussian system dynamics parameters and the Kalman filter mean and covariance calculations. This work generates a new framework that extends classical theory of positive semi-definiteness, quadratic forms and partial ordering of matrices. The structural properties analyzed generate new insights into the practice of monitoring patients, some of which have been hypothesized by physicians (see (13)), but have yet to be rigorously validated.

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 Section 3. In Section 4, we discuss our approach to determine the Time to Next Test(TNT) and the solution and structural properties of our algorithm. Section 5 applies our models retrospectively to a large-scale glaucoma clinical trial for validation and to demonstrates how our algorithm can deliver improved patient care with fewer tests. Finally, we discuss our results and future directions in Section 6.

2. Literature

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

Medical Examination and Screening Models: Most research in the field focuses on performing discrete screenings or examinations to detect the first incidence of a disease, rather than monitoring an ongoing chronic disease. 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 (21, 33, 9)). Work has also been done with regard to the timing of initial treatment (see (36, 11, 35)). The above research, however, does not incorporate multi-dimensional state spaces in feedback driven control loops to monitor patient-specific disease progression.

A second related research area involves monitoring and treatment decisions of an ongoing condition. For example, models have been developed for the treatment of HIV, diabetes, organ transplantation, cancer, and how to manage drug therapy dosages (see (19, 1)). 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. Much of the current work relies on Markov Decision Processes (MDP) or simulation for determining optimal monitoring schedules. MDP models suffer from the curse of dimensionality, which hinders the potential for modeling multiple dimensions of patient disease state. Tractable optimization, on the other hand, is difficult or impossible in simulation. By using controlled linear Gaussian models for disease progression and monitoring, our work significantly advances the state-of-the-art in the modeling and monitoring of chronic diseases.

Machine Surveillance, Inspection and Maintenance Models: Extensive surveys of the literature in machine maintenance, inspection and surveillance are discussed by (32) and (37). These surveys propose that the literature can be divided into 5 primary modeling approaches (1) Age Replacement Models, (2) Block Replacement Models, (3) Delay-time Models for inspection, (4) Damage Models, (5) Deterioration Models.

Model types (3), (4), and (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 (26, 23)). 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 (39, 6)), or the length of time a failure

goes undetected (see (15, 24)). These models, however, consider a one-dimensional state space with Markovian or semi-Markovian system dynamics and perfect observations, which is often insufficient for chronic disease monitoring.

In non-Markovian surveillance and inspection models (see (27, 16)), the state space is still one-dimensional and the observations are assumed to be perfect. Papers that consider noisy or uncertain observations include (29, 30). 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 multi-dimensional state space with both observation noise and system. Our research thereby significantly expands the modeling approaches in inspection/surveillance and deterioration/damage modeling.

Linear Gaussian Systems: Linear Gaussian systems and linear quadratic Gaussian control (LQG) have been used in many different applications in dynamical systems modeling, estimation, and control theory. A major departure from the foundational models, our particular interest is in linear Gaussian systems *without fixed observation intervals*. Sensor scheduling is a related areas as it assesses: given a set of available sensors, how frequently should one take measurements and from which sensors? Work in this area includes (31, 38); however, this literature typically assumes that a measurement is taken every period (though from different sensors). Control of measurement subsystems (see (22, 3, 18)) 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 non-standard optimization approaches.

3. State Space Modeling of Progression

We develop state space models for estimating and forecasting disease state and demonstrate their application to modeling the progression of glaucoma. This approach allows us to incorporate both (1) process noise, which can approximate the effect of unmodeled disease dynamics, and (2) measurement noise in medical test measurements. In Sec. 3.1 we present a linear Gaussian systems approach to modeling disease dynamics, which is then applied in Sec. 3.2 to glaucoma patients from a major clinical trial, the Collaborative Initial Glaucoma Treatment Study (CIGTS). Finally, Sec. 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 Sec. 4.

3.1. Linear Gaussian Systems Disease Model

State space models are adequate for a surprisingly general class of systems, especially if state augmentation is used to linearize them (see (5)). 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. The linear Gaussian systems model is comprised of a *patient disease state* and the particular *linear system disease dynamics*.

3.1.1. Patient Disease State. Current evidence indicates that the primary indicator of glaucoma progression is worsening of the Visual Field (VF), and that Intraocular Pressure (IOP) is a critical risk factor for future progression. In our model, we consider an eight-dimensional vector to model the state of the patient, $\alpha_t = \left[VF, \frac{\partial VF}{\partial t}, \frac{\partial^2 VF}{\partial t^2}, \frac{\partial^3 VF}{\partial t^3}, IOP, \frac{\partial IOP}{\partial t}, \frac{\partial^2 IOP}{\partial t^2}, \frac{\partial^3 IOP}{\partial t^3} \right]$, where VF refers to a global measure of performance from the visual field test. $\frac{\partial VF}{\partial t}$, $\frac{\partial^2 VF}{\partial t^2}$, and $\frac{\partial^3 VF}{\partial t^3}$ refer to the first three moments of the VF measure with respect to time: velocity, acceleration and jerk. Similarly, IOP represents the intraocular pressure measurement. The third moment, jerk, was included after preliminary research revealed that the model with only the first two moments was missing sudden fluctuations in IOP and VF performance that are commonly associated with jerk in physics.

3.1.2. Linear System Disease Dynamics. Our discrete-time linear systems models are recursive models where, in each period, there is a system transition and also a measurement of the system that is available to the observer/controller. The mathematical 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 of the recursive system equations, there is an *initial state* that is assumed prior to any observations, based on population characteristics found in the CIGTS clinical trial.

State Transition Equation. In each period, t, the system transitions to a new state according to a linear 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

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

where η is a Gaussian random vector with $\mathbf{E}[\eta] = 0$ and $\mathbf{Var}[\eta] = \mathbf{Q}$. 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. Non-white noise is present in some applications, which could be accounted for

by generalizing the observation model (see Eq. 2) or by using state augmentation to model colored noise by passing the white noise through a linear system model (e.g. see (5)).

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

$$\mathbf{z}_t = \mathbf{Z}\alpha_t + \epsilon \qquad t = 1, \ldots, N, \tag{2}$$

where ϵ is a Gaussian random variable with $\mathbf{E}[\epsilon] = 0$ and $\mathbf{Var}[\epsilon] = \mathbf{H}$. 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 $\mathbf{Var}[X_0] = \hat{\Sigma}_0$. As will be seen in the Kalman filter approach to state estimation for linear Gaussian systems, the goal is to estimate the mean and covariance parameters of the Gaussian state variable.

3.1.3. State Estimation and Prediction with the Kalman Filter. One of the classical frameworks for state estimation and prediction for linear Gaussian systems is the Kalman filter. 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 operations (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 Sec. 3.1.2, the Kalman filter approach (see (14)) 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 linear system dynamics model from Eq. 1 to predict the future state as

$$\hat{\alpha}_{t+1|t} = \mathbf{T}\hat{\alpha}_{t|t} \tag{3}$$

$$\hat{\Sigma}_{t+1|t} = \mathbf{T}\hat{\Sigma}_{t|t}\mathbf{T}' + \mathbf{Q}, \tag{4}$$

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 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 around the measurement, $\mathbf{S}_{t+1}$, are calculated as

$$\tilde{\mathbf{y}}_{t+1} = \mathbf{z}_{t+1} - \mathbf{Z}\hat{\alpha}_{t+1|t} \tag{5}$$

$$\mathbf{S}_{t+1} = \mathbf{Z}\hat{\Sigma}_{t+1|t}\mathbf{Z}' + \mathbf{H}. \tag{6}$$

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

$$\mathbf{K}_{t+1} = \hat{\Sigma}_{t+1|t}\mathbf{Z}' \cdot \mathbf{S}_{t+1}^{-1}. \tag{7}$$

The optimal Kalman gain from Eq. 7 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

$$\hat{\alpha}_{t+1|t+1} = \hat{\alpha}_{t+1|t} + \mathbf{K}_{t+1} \cdot \tilde{\mathbf{y}}_{t+1} \tag{8}$$

$$\hat{\Sigma}_{t+1|t+1} = (I - \mathbf{K}_{t+1}\mathbf{Z})\,\hat{\Sigma}_{t+1|t}, \tag{9}$$

where I is the identity matrix. Eq. 8 and 9 are the key equations that define the recursive Kalman estimator and will be relied upon in subsequent analysis.

Multi-Period Prediction. In our application, we would like to consider a dynamic schedule that depends on the condition of each patient which can vary from one patient to another. Thus the optimal time interval between tests will vary from one measurement to the next depending on the stability of a given patient's disease progression at prior time points. Therefore, our approach must predict an arbitrary number of periods before applying the update step. By eliminating the update step for periods in which no observation is performed, it is possible to apply the linear transition equation recursively to obtain the ℓ-step prediction equation (i.e. predicting ℓ periods into the future) as

$$\hat{\alpha}_{t+\ell|t} = \mathbf{T}^{\ell}\hat{\alpha}_{t|t} \tag{10}$$

$$\hat{\Sigma}_{t+\ell|t} = \mathbf{T}^{\ell}\hat{\Sigma}_{t|t}(\mathbf{T}^{\ell})' + \sum_{j=0}^{\ell-1} \mathbf{T}^{j}\mathbf{Q}\mathbf{T}^{j'}, \tag{11}$$

where $\alpha_{t+\ell}$ is the Gaussian state variable at time ℓ given that observations are available through time t (i.e., the observation history). The first element of the sum represents the multi-period linear state transition and the second element of the sum represents the multi-period process noise accumulation.

3.2. Application of Linear Systems Models to the CIGTS Clinical Trial Data

In Sections 3.1.1, 3.1.2, and 3.1.3 we presented the theoretical framework that we use to model disease progression in glaucoma patients. This section demonstrates the necessary steps to translate theory into a functional model using real patient data from a previously conducted clinical trial. An important first step to developing an effective estimation and forecasting model is to properly parameterize the linear systems model. Based on historical data, the goal is to determine the matrices **T** (linear system dynamics), **Q** (process noise covariance), **Z** (the observation matrix that allows some or all of the states to be measured in a possibly altered form), **H** (measurement noise covariance), $\hat{\alpha}_0$ (initial state mean), and $\hat{\Sigma}_0$ (initial state covariance). To determine these parameters, we rely on the expectation maximization (EM) algorithm for parameter estimation of linear stochastic systems and its implementation in Matlab (see (25)).

First, we isolated a set of training data and used it to calibrate the model with the EM algorithm. Then we tested the estimation model on the remaining patients who were enrolled in the CIGTS trial. For each patient in the test set, we ran the Kalman filter for the first two years of their time in the trial (as a warm-up) and then made predictions of the future state of their glaucoma from the two year point through the end of the patient's trial. Since most forms of glaucoma tend to progress very slowly over the course of many years, a two year warm up was deemed reasonable for this disease, but the transient period may be modified to fit the specific disease in question. Although patients in the CIGTS had variable follow-up, we were able to test our predictions up to *six years* into the future for most patients. We then measured the prediction error by comparing the predicted mean state with the actual observations. The overall results presented in Fig. **??** and Table 2 show that our linear systems model for state prediction has very little bias.

	Months Predicted into the Future									
	6 mo	12 mo	18 mo	...	42 mo	48 mo	54 mo	60 mo	66 mo	72 mo
Mean	0.06	0.12	0.16	...	0.16	0.19	0.22	0.25	0.19	0.07
Var	2.72	6.05	7.47	...	13.28	15.29	17.19	18.38	18.28	19.53

Table 2 Kalman filter absolute value of prediction error over the CIGTS population versus number of periods predicted.

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 (17)). This can be difficult in practice because the amount of data is so large and is processed mentally without

the aid of any decision support system. Identifying and properly utilizing this multidimensional space of information over a history of observations is the purpose of the Probability of Progression (ProP) function. Specifically, the ProP function is a mapping, f, that maps the state space of physiological indicators, $\mathcal{S}$, to a measure of disease progression in $[0,1]$: probability of progression.

Current evidence indicates that the primary factors predicting glaucoma progression are abnormalities of the VF and the level of IOP. The first step involves defining ProP using these indicators for two key points in time: (1) the current moment (immediately after a new measurement is taken) and (2) a future moment at which the next measurement should be taken. The ProP function links patient disease state to disease progression by combining the widely-accepted Hodapp-Parrish-Anderson (HPA) criteria to quantify that progression has occurred with a loss of 3 dB of mean deviation (the mean deviation is a summary measure of VF loss which is referred to as the "VF score" in this paper) with respect to the patient's baseline. We define the ProP function as a logistic regression that links the key physiological factors (e.g. VF and IOP) to the probability of progression as in (34):

$$f(\mathbf{x}) \;=\; \frac{1}{1+e^{-z(\mathbf{x})}} \tag{12}$$

$$z(\mathbf{x}) = b + \mathbf{a}\mathbf{x} \tag{13}$$

where $z(\mathbf{x})$ is a linear function of key risk factors, including VF and IOP measures and can include other important factors such as age, race and others depending on the specific disease. These factors are captured by the *progression vector*, $\mathbf{a}$, which is so called because it represents the n-dimensional direction of steepest ascent toward progression. A thorough treatment of the key factors involved in glaucoma progression is can be found in (34).

4. Time to Next Test (TNT)

The idea behind our approach is that a test only needs to be performed when the physician is no longer sufficiently confident that the patient has not progressed. To determine the Time to Next Test (TNT), we forecast the patient disease state trajectory into the future until the ProP function hits a threshold indicating sufficient likelihood of progression for a test to be performed. The optimal interval of time in between tests is therefore determined by the length of time it takes for the disease state forecast to reach the progression threshold. The complicating factor is that the future state is not a deterministic point in n-dimensional space, but rather an n-dimensional Gaussian random variable. Thus, we develop a stochastic Point of Maximum Progression (POMP) function that maximizes the deterministic ProP function over the Gaussian density of the forecasted state.

This yields the "worst" 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.

Fig. 2 is a conceptual representation of this approach for a 3-dimensional state space. In this figure t is the current period and the ellipse 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 projected. The time of the next test occurs at the first period in which the forecasted confidence region intersects or exceeds the progression threshold (and n-dimensional hyperplane), illustrated by the plane in Fig. 2; in this case period $t+4$.

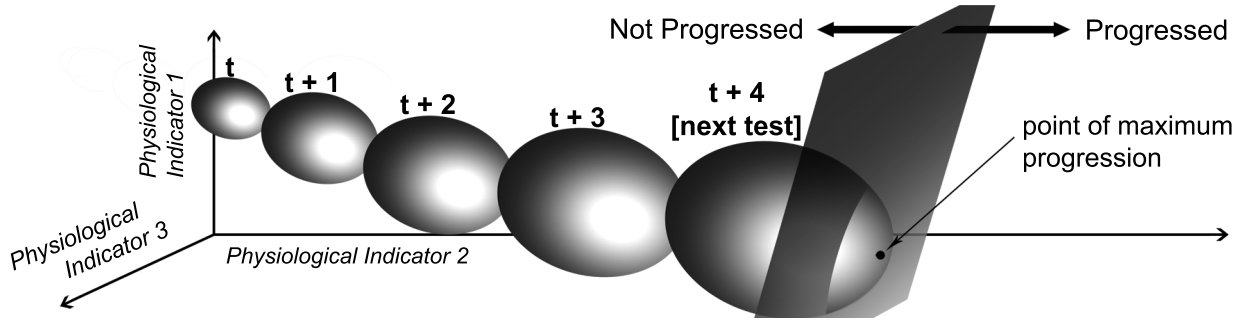


Figure 2 Depiction of the confidence region point of maximum progression time to next test approach

In our approach the clinician has two parameters that can be adjusted to customize the treatment schedule to each patient, as summarized in Table 3. The parameter τ controls the progression threshold. If for a given state the ProP function produces a value greater than $1/(1+e^{-\tau})$ the patient is considered to have progressed. Hence the smaller the τ, the more frequently the model will recommend tests. The parameter ρ adjusts the size of the confidence region around the predicted mean disease state, thereby allowing the clinician to determine the confidence level that there has been no progression. The larger the ρ, the more frequently the model will recommend tests.

Parameter	Definition	Effect on Testing
$\tau \in (-\infty, \infty)$;	Progression threshold (w.r.t. POMP function);	$\tau\downarrow$= more frequent
$\rho \in (0, 1)$;	Size of confidence region around the mean state estimate;	$\rho\uparrow$= more frequent

Table 3 Model parameters that can be adjusted by the clinician to customize treatment.

The benefit of the confidence region approach is that the confidence level can be adjusted by the clinician based on treatment strategy to be more aggressive (e.g. 90%) or less aggressive (e.g.

60%). For example, a doctor caring for an older glaucoma patient who has other illnesses may opt to be less aggressive in monitoring due to the low likelihood the patient will go blind in his/her remaining years and the risks of additional treatment may outweigh the potential benefits. This will lead to longer intervals between tests and thus less frequent testing. Alternatively, with a young, healthy glaucoma patient the doctor will increase ρ so that testing is more frequent to avoid missing detection of irreversible vision loss.

4.1. Point of Maximum Progression (POMP) Time to Next Test (TNT) Approach

In this section we develop a closed form solution to the optimization of the ProP function over the Gaussian prediction region. The form of prediction region that we choose to work with is an n-dimensional ellipsoid where n is the dimension of the state space. Mathematically we can define the $100\rho\%$ prediction region for the Gaussian random variable with mean $\hat{\alpha}_{t+\ell|t}$ and covariance $\hat{\Sigma}_{t+\ell|t}$ for ℓ periods in the future as

$$\mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t}) = \{\mathbf{x} : (\mathbf{x} - \hat{\alpha}_{t+\ell|t})' \hat{\Sigma}^{-1}_{t+\ell|t} (\mathbf{x} - \hat{\alpha}_{t+\ell|t}) \leq \chi^2(1-\rho, n)\}, \tag{14}$$

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 (8)). 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. However, the prediction region represents a *set* of points (rather than a single point) that we expect the true state to lie. If a conservative estimate of ProP is desired, then the approach would be to find the maximum value of the ProP function, f, over the prediction region, $\mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})$. Given the current state estimate, $\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}$, the stochastic Point of Maximum Progression (POMP) function, h_ρ, with respect to the ProP function, f, for the ℓ-step state forecast is given by

$$h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, \ell) = \max_{\mathbf{x} \in \mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})} f(x), \tag{15}$$

where $\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t}$ are obtained from $\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}$ through Eq. 10 and 11. To solve the optimization problem formulated in Eq. 15, we first analyze the structure of the prediction region.

LEMMA 1. *The prediction region, $\mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})$, defined by Eq. 14 is convex.*

Proof. To prove that the prediction region $\mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})$ is convex, let $y = \mathbf{x} - \hat{\alpha}_{t+\ell|t}$. Let $g(\mathbf{y}) = \mathbf{y}' \hat{\Sigma}^{-1}_{t+\ell|t} \mathbf{y} - \chi^2(1-\rho, n)$. Then the constraint $g(y) \leq 0$ is convex iff the Hessian of g is positive semi-definite. It can easily be shown that the Hessian of g is given by $H[g, \mathbf{y}] = 2\hat{\Sigma}^{-1}_{t+\ell|t}$, which is positive semi-definite because $\hat{\Sigma}^{-1}_{t+\ell|t}$ is psd. □

It is easily conceivable that in many chronic illnesses, as in glaucoma, the ProP function will be a logistic regression as described in Sec. 3.3. Therefore, maximizing the ProP function is equivalent to maximizing $z(\mathbf{x})$ (see Eq. 12), which is a linear function of $\mathbf{x}$. From this and Lemma 1, 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 Eq. 13. The optimization of the ProP function over the prediction region has closed form solution given by the 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 $\mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})$ defined by Eq. 14 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,*

$$h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, \ell) = \max_{\mathbf{x} \in \mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})} \mathbf{a}'\mathbf{x} = \mathbf{a}'\hat{\alpha}_{t+\ell|t} + \sqrt{\chi^2(1-\rho, n)\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}} \tag{16}$$

$$\widetilde{h}_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, \ell) = \operatorname*{arg\,max}_{\mathbf{x} \in \mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})} \mathbf{a}'\mathbf{x} = \hat{\alpha}_{t+\ell|t} + \left(\sqrt{\frac{\chi^2(1-\rho, n)}{\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}}} \right) \cdot \hat{\Sigma}_{t+\ell|t}\mathbf{a}. \tag{17}$$

Proof. To show that h_ρ has closed form solution with optimal value $\tilde{h}_\rho$, we begin with the fact that the feasible region is convex by Lemma 1. Further, the objective function can be reformulated to an equivalent objective that is linear in the decision variable. It is clear that maximizing $f(\mathbf{x}) = \frac{1}{1+e^{-z(\mathbf{x})}}$ is equivalent to maximizing the linear function $z(\mathbf{x}) = b + \mathbf{a}\mathbf{x}$, where $\nabla z(\mathbf{x}) = \mathbf{a}$. We reformulate maximization problem to the equivalent minimization problem to match the standard KKT conditions:

$$\max_{\mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})} \mathbf{a}'\mathbf{x} = \min_{\mathcal{D}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})} -\mathbf{a}'\mathbf{x} \tag{18}$$

The linear objective and the convex constraints guarantee that the KKT conditions are necessary and sufficient. Thus any solution that satisfies the KKT conditions will be optimal.

First note that $\hat{\Sigma}_{t+\ell|t}$ is positive semi-definite so $\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a} \geq 0$. Secondly, if $\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a} = 0$ then we would have a perfect prediction of the patient's future state without any uncertainty, which is not realistic so without loss of generality we let $\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a} > 0$. This eliminates any degenerate cases for taking square roots or dividing by $\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}$.

Stationarity Conditions If we let f represent the objective function of the minimization problem (Eq. 18) and g represent the constraint function then the stationarity conditions are

$$\begin{aligned}\nabla f + u \cdot \nabla g &= -\mathbf{a}' + 2u(\mathbf{x} - \hat{\alpha}_{t+\ell|t})'\hat{\Sigma}^{-1}_{t+\ell|t} = -\mathbf{a}' + 2u\sqrt{\frac{\chi^2(1-\rho,n)}{\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}}} \cdot \left(\hat{\Sigma}_{t+\ell|t}\mathbf{a}\right)' \hat{\Sigma}^{-1}_{t+\ell|t} \\ &= -\mathbf{a}' + 2u\left(\sqrt{\frac{\chi^2(1-\rho,n)}{\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}}}\right) \cdot \mathbf{a}'\end{aligned}$$

where the first equality follows by taking the respective gradients and the second equality follows by plugging in the proposed optimal solution, $\mathbf{x}^* = \tilde{h}_\rho(\hat{\alpha}_{t+\ell|t}, \hat{\Sigma}_{t+\ell|t})$, for $\mathbf{x}$. If we then let

$$u = \frac{1}{2} \cdot \left(\sqrt{\frac{\chi^2(1-\rho,n)}{\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}}}\right)^{-1}, \tag{19}$$

clearly the stationarity conditions, $\nabla f + u \cdot \nabla g = 0$, will be satisfied

Complementary Slackness Conditions As before, we let $\mathbf{x}^* = \tilde{h}_\rho$ be the proposed solution to the optimization problem

$$\begin{aligned}g(\mathbf{x}^*) &= \sqrt{\frac{\chi^2(1-\rho,n)}{\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}}} \cdot \left(\hat{\Sigma}_{t+\ell|t}\mathbf{a}\right)' \hat{\Sigma}^{-1}_{t+\ell|t} \sqrt{\frac{\chi^2(1-\rho,n)}{\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}}} \cdot \hat{\Sigma}_{t+\ell|t}\mathbf{a} - \chi^2(1-\rho,n) \\ &= \frac{\chi^2(1-\rho,n)}{\mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a}} \cdot \mathbf{a}'\hat{\Sigma}_{t+\ell|t}\mathbf{a} - \chi^2(1-\rho,n) = 0.\end{aligned}$$

Therefore, the complementary slackness conditions are satisfied.

Dual Feasibility: By Eq. 19 it is clear the $u \geq 0$. □

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.

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

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. As an example, a simple binary search that divides the search space in half at each iteration can solve this problem for a problem with n possible testing epochs in the worst case on order of $O(\log(n))$, because the terms are monotonically increasing in ℓ. This is faster in the worst case than traditional binary search because of the special structure. Thus 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 1 second over the course of a year. The search space is 31,449,600 seconds/year.

The binary search will find the optimal monitoring time in at worst 25 function evaluations plus comparisons, which a modern computer could solve instantaneously.

In Section 5, we compare the performance of our TNT algorithm with currently accepted medical practice. We also present in Section 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. One major conclusion is that we are able to confirm the belief held by some (see (13)) that testing intervals for glaucoma should be variable rather than fixed. Our approach goes even further by showing that the length of the testing interval can be determined using the key physiological indicators of progression based on the patient's history of tests.

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 about the monitoring and testing of chronic disease patients. 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 thus are more likely to test. Property 2, given in Lemma 2, 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 Eq. 13). 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.

One of the primary features of many chronic diseases is that the disease tends to get worse over time. That is, the overall disease trajectory is a progressive one. Some clear examples include Alzheimer's, Parkinson's, and ALS among others. For glaucoma this is manifest in the fact that once a patient has lost sight they can never again regain it. Mathematically the progressive nature of chronic disease can be captured by a condition on the system transition matrix, $\mathbf{T}$.

DEFINITION 1. We call a linear transformation $\mathbf{T}$, a **progressing transformation** for progression vector $\mathbf{a} \in \mathbb{R}^n$, if

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

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

Property 1 (Prediction Uncertainty) shows that as the Kalman filter projects the patient's state further into the future, it monotonically approaches the threshold, τ, for scheduling a next test. This property supports the intuition that, the further into the future we wait before testing, the more uncertain we are about the patient's disease state. Property 1 is given by the following theorem, which is proved in the Online Appendix.

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

Proof. We prove monotonicity of $h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, \ell)$ given progression vector $\mathbf{a}$ via induction. For the base case, consider $h_\rho(\hat{\alpha}_{t|t}, \Sigma_{t|t}, 1)$ compared with $h_\rho(\hat{\alpha}_{t|t}, \Sigma_{t|t}, 0)$. The following relationships hold due to the Kalman prediction equations, Eq.'s 3 and 4

$$\mathbf{a}'\hat{\alpha}_{t+1|t} = \mathbf{a}'\mathbf{T}\hat{\alpha}_{t|t} \tag{21}$$

$$\mathbf{a}'\hat{\Sigma}_{t+1|t}\mathbf{a} = \mathbf{a}'\mathbf{T}\hat{\Sigma}_{t|t}\mathbf{T}'\mathbf{a} + \mathbf{a}'\mathbf{Q}\mathbf{a}. \tag{22}$$

First note that because $\mathbf{T}$ is a progressing transformation then $\mathbf{a}'\mathbf{T}\hat{\alpha}_{t|t} \geq \mathbf{a}'\hat{\alpha}_{t|t}$. Using the Cholesky decomposition on the positive semi-definite covariance matrix, $\hat{\Sigma}_{t|t} = \mathbf{L}\mathbf{L}'$, we get the following relationships

$$\mathbf{a}'\mathbf{T}\hat{\Sigma}_{t|t}\mathbf{T}'\mathbf{a} = \mathbf{a}'\mathbf{T}\mathbf{L}\mathbf{L}'\mathbf{T}'\mathbf{a} \tag{23}$$

$$\mathbf{a}'\hat{\Sigma}_{t|t}\mathbf{a} = \mathbf{a}'\mathbf{L}\mathbf{L}'\mathbf{a}. \tag{24}$$

Note that the right hand side (RHS) of Eq.'s 23 and 24 are symmetric with the right half of the RHS being the transpose of the left half of the RHS (e.g. $\mathbf{a}'\mathbf{T}\mathbf{L} = (\mathbf{L}'\mathbf{T}'\mathbf{a})'$). Rewriting the left half of the RHS of Eq.'s 23 and 24 we get

$$\mathbf{a}'\mathbf{T}\mathbf{L} = [\mathbf{a}'\mathbf{T}\mathbf{L}_1, \mathbf{a}'\mathbf{T}\mathbf{L}_2, \ldots, \mathbf{a}'\mathbf{T}\mathbf{L}_n] \tag{25}$$

$$\mathbf{a}'\mathbf{L} = [\mathbf{a}'\mathbf{L}_1, \mathbf{a}'\mathbf{L}_2, \ldots, \mathbf{a}'\mathbf{L}_n], \tag{26}$$

where $\mathbf{L}_i$ represents the i^{th} column of the $\mathbf{L}$ matrix, which is the matrix square root of $\Sigma_{t|t}$. Invoking the properties of the progressing transformation, $\mathbf{T}$, it is clear that

$$\mathbf{a}'\mathbf{T}\mathbf{L}_i \geq \mathbf{a}'\mathbf{L}_i \quad \text{for } i = 1, \ldots, n, \tag{27}$$

which implies that each entry of the $\mathbf{a}'\mathbf{T}\mathbf{L}$ vector, the left half of Eq. 23, is larger than or equal to each entry of the $\mathbf{a}'\mathbf{L}$ vector, the right half of Eq. 24. Further, we have that $\mathbf{L}'\mathbf{T}'\mathbf{a} = (\mathbf{a}'\mathbf{T}\mathbf{L})'$ and $\mathbf{L}'\mathbf{a} = (\mathbf{a}'\mathbf{L})'$. Combining Eq.'s 23, 24, 25, 26, and 27,

$$\begin{aligned}
\mathbf{a}'\mathbf{T}\hat{\Sigma}_{t|t}\mathbf{T}'\mathbf{a} &= (\mathbf{a}'\mathbf{T}\mathbf{L}_1)^2 + (\mathbf{a}'\mathbf{T}\mathbf{L}_2)^2 + \ldots + (\mathbf{a}'\mathbf{T}\mathbf{L}_n)^2 \\
&\geq (\mathbf{a}'\mathbf{L}_1)^2 + (\mathbf{a}'\mathbf{L}_2)^2 + \ldots + (\mathbf{a}'\mathbf{L}_n)^2 \\
&= \mathbf{a}'\hat{\Sigma}_{t|t}\mathbf{a}.
\end{aligned} \tag{28}$$

The first equality follows from Eq. 25 and the fact that the right half of the Cholesky decomposition is just the transpose of the left half. The inequality follows by applying the properties of the progressing transformation, $\mathbf{T}$, to each term, $(\mathbf{a}'\mathbf{T}\mathbf{L}_1)^2$, of the sum. The final equality follows from the same arguments as the first equality. This result could have been shown by invoking Def. 1 (ii); however, we chose the preceding approach to show that this property holds in general even without requiring the second property of a progressing transformation. In fact, Eq. 28 will hold for any positive semi-definite matrix as long as Def 1 (i) holds, and since the covariance matrix is positive semi-definite it follows from the arguments of Eq. 28.

Finally since $\mathbf{Q}$ is positive semi-definite, $\mathbf{a}'\mathbf{Q}\mathbf{a} \geq 0$. Now we have shown that $\mathbf{a}'\mathbf{T}\hat{\Sigma}_{t|t}\mathbf{T}'\mathbf{a} \geq \mathbf{a}'\hat{\Sigma}_{t|t}\mathbf{a}$ and $\mathbf{a}'\mathbf{T}\hat{\alpha}_{t|t} > \mathbf{a}'\hat{\alpha}_{t|t}$. It follows directly from Eq. 22 and 21 that

$$h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, 1) = \mathbf{a}'\mathbf{T}\hat{\alpha}_{t|t} + \sqrt{\chi^2(1-\rho, n)\mathbf{a}'(\mathbf{T}\hat{\Sigma}_{t|t}\mathbf{T}' + \mathbf{Q})\mathbf{a}} \tag{29}$$

$$\geq \mathbf{a}'\hat{\alpha}_{t|t} + \sqrt{\chi^2(1-\rho, n)\mathbf{a}'\hat{\Sigma}_{t|t}\mathbf{a}} \quad = \; h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}, 0) \tag{30}$$

The base case has been proven. For the induction step, assume the claim is true for l, it follows that the claim holds for $l+1$ directly using the same arguments and the fact that

$$\mathbf{a}'\hat{\alpha}_{t+\ell+1|t} = \mathbf{a}'\mathbf{T}\hat{\alpha}_{t+\ell|t} \tag{31}$$

$$\mathbf{a}'\hat{\Sigma}_{t+\ell+1|t}\mathbf{a} = \mathbf{a}'\mathbf{T}\hat{\Sigma}_{t+\ell|t}\mathbf{T}'\mathbf{a} + \mathbf{a}'\mathbf{Q}\mathbf{a}. \quad \square \tag{32}$$

Property 2 (No. 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. To state this property rigorously, we first present some notation as well as 3 definitions regarding pertinent 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 Eq. 11. 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 Eq.'s 6 and 7.

Definition 2. Given 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 Eq. 9 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 $\mathbf{a}'(\mathbf{M} - \mathbf{N})\mathbf{a} \geq 0$.

Definition 3 is similar to the matrix equivalent of the greater than symbol 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 $\mathbf{A_n} \succeq_a \mathbf{A_{n-1}} \succeq_a \ldots \succeq_a \mathbf{A_1}$

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 extremely 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 Sec. 5. For reasonable systems of progressive disease, a-monotonicity should hold because lack of a-monotonicity leads to a counter intuitive result. Every time the Kalman filter obtains an observation, an adjustment is made that reduces the covariance around the state estimate *independent of the actual observed value.* If a-monotonicity did not hold, then it would be possible that adding an extra observation would reduce the overall covariance adjustment to the system. 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 2. *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 $\Sigma_{t_f|t_f}^{\pi_m} \succeq_a \Sigma_{t_f|t_f}^{\pi_n}$ for $n > m$.*

Proof. To prove that $\Sigma_{t_f|t_f}^{\pi_m} \succeq_a \Sigma_{t_f|t_f}^{\pi_n}$ for $n > m$ and $\pi_n \supset \pi_m$ we instead consider, without loss of generality, systems over the interval $[1, t]$ with a starting observation $t_s = 1$ and a final

observation at $t_f = t \in \{3,4,5,\ldots\}$. Now consider two policies, $\pi_m \subset \pi_n$. In policy π_m there are m observations at times $s_1,\ldots,s_m$, and in policy π_n there are $n > m$ observations at times $t_1,\ldots,t_n$ such that $t_1 = s_1 = 1$ and $t_n = s_m = t$ and $\forall i\ \exists j$ such that $s_i = t_j$. Let the covariance estimate adjustment for observation at time s_j given the last observation was at time s_i under policy π_m be denoted by $\mathbf{C}^{\pi_m}_{s_j,s_i}$. The key observation to make is that the covariance matrix at the final time s_m under observation schedule π_m can be shown after some algebra to have the following form:

$$\begin{aligned}
\Sigma^{\pi_m}_{s_m|s_m} &= \Sigma^{\pi_m}_{s_m|s_{m-1}} - \mathbf{C}^{\pi_m}_{s_m,s_{m-1}} = \mathbf{T}^{s_m-s_{m-1}}\Sigma^{\pi_m}_{s_{m-1}|s_{m-1}}\mathbf{T}^{s_m-s_{m-1}\prime} + \mathbf{Q} - \mathbf{C}^{\pi_m}_{s_m,s_{m-1}} \\
&= \mathbf{T}^{s_m-s_{m-1}}\left(\mathbf{T}^{s_{m-1}-s_{m-2}}\Sigma^{\pi_m}_{s_{m-2}|s_{m-2}}\mathbf{T}^{s_{m-1}-s_{m-2}\prime} + \mathbf{Q} - \mathbf{C}^{\pi_m}_{s_{m-1},s_{m-2}}\right)\mathbf{T}^{s_m-s_{m-1}\prime} + \mathbf{Q} - \\
&\quad \mathbf{C}^{\pi_m}_{s_m,s_{m-1}} \\
&= \cdots = \mathbf{T}^{s_m-1}\Sigma^{\pi_m}_{1|1}\mathbf{T}^{s_m-1\prime} + \sum_{j=0}^{s_m-1}\mathbf{T}^j\mathbf{Q}\mathbf{T}^{\mathbf{j}\prime} - \sum_{j=2}^{m}\mathbf{T}^{s_m-s_j}\mathbf{C}^{\pi_m}_{s_j,s_{j-1}}\mathbf{T}^{s_m-s_j\prime} \\
&= \Sigma^{\pi_m}_{s_m|1} - \sum_{j=2}^{m}\mathbf{T}^{t-s_j}\mathbf{C}^{\pi_m}_{s_j,s_{j-1}}\mathbf{T}^{t-s_j\prime}. \qquad (33)
\end{aligned}$$

The ellipsis in the above equation represents the further expansion of the covariance matrix estimate. Similarly we have that

$$\Sigma^{\pi_n}_{t_n|t_n} = \Sigma^{\pi_n}_{t_n|1} - \sum_{j=2}^{n}\mathbf{T}^{t-t_j}\mathbf{C}^{\pi_n}_{t_j,t_{j-1}}\mathbf{T}^{t-t_j\prime}$$

Note that $\Sigma^{\pi_n}_{t_n|1} = \Sigma^{\pi_m}_{s_m|1}$ because $t_n = s_m = t$ and the policy has no effect on the t-step prediction since no observations are incorporated in the predicted covariance. It is now clear that to show $\Sigma^{\pi_m}_{t|t} \succeq_a \Sigma^{\pi_n}_{t|t}$, it is sufficient to show

$$\sum_{j=2}^{n}\mathbf{T}^{t-t_j}\mathbf{C}^{\pi_n}_{t_j,t_{j-1}}\mathbf{T}^{t-t_j\prime} \succeq_a \sum_{j=2}^{m}\mathbf{T}^{t-s_j}\mathbf{C}^{\pi_m}_{s_j,s_{j-1}}\mathbf{T}^{t-s_j\prime} \qquad (34)$$

First we show the result for an arbitrary feasible number of observations, m, that adding one extra observation to the schedule will yield a covariance matrix that is quadratically smaller with respect to the progression vector a. That is for all $\pi_{m+1} \in \Pi_{m+1}([1,t])$ and $\pi_m \in \Pi_m([1,t])$ where $\pi_{m+1} \supset \pi_m$,

$$\sum_{j=2}^{m+1}\mathbf{T}^{t-t_j}\mathbf{C}^{\pi_{m+1}}_{t_j,t_{j-1}}\mathbf{T}^{t-t_j\prime} \succeq_a \sum_{j=2}^{m}\mathbf{T}^{t-s_j}\mathbf{C}^{\pi_m}_{s_j,s_{j-1}}\mathbf{T}^{t-s_j\prime}.$$

In this case all the observations occur at the same time points except that the policy π_{m+1} with $m+1$ observations will have an extra observation in between two of the observations from policy π_m. Without loss of generality let the extra observation occur at time s, where $s_{j-1} < s < s_j$ and

s_{j-1} and s_j are the $j-1$ and j observations in policy π_m. All of the covariance estimate updates prior to observations s remain unchanged between schedule π_m and schedule π_{m+1} and thus can be canceled out. Therefore it remains to show that

$$\mathbf{T}^{t-s}\mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}}\mathbf{T}^{t-s'} + \mathbf{T}^{t-s_j}\mathbf{C}^{\pi_{m+1}}_{s_j,s} + \mathbf{T}^{t-s_j\prime} + \sum_{i=j+1}^{m}\mathbf{T}^{t-s_i}\mathbf{C}^{\pi_{m+1}}_{s_i,s_{i-1}}\mathbf{T}^{t-s_j\prime} \succeq_a$$
$$\mathbf{T}^{t-s_j}\mathbf{C}^{\pi_m}_{s_j,s_{j-1}}\mathbf{T}^{t-s_j\prime} + \sum_{i=j+1}^{m}\mathbf{T}^{t-s_i}\mathbf{C}^{\pi_m}_{s_i,s_{i-1}}\mathbf{T}^{t-s_j\prime}. \tag{35}$$

To show the relationship from Eq. 35 it is best to break the LHS and RHS into smaller components and show how each component of the LHS dominates the corresponding component of the RHS recursively building up to the entire equation. The way we do so is by factoring powers of $\mathbf{T}$ out of each term as follows. Recalling that $t = s_m$, the LHS can clearly be rewritten as

$$\begin{aligned}
LHS =& \left(\prod_{i=j+1}^{m}\mathbf{T}^{s_i-s_{i-1}}\right)\mathbf{T}^{s_j-s}\mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}}\mathbf{T}^{s_j-s'}\left(\prod_{i=j+1}^{m}\mathbf{T}^{s_i-s_{i-1}\prime}\right) + \\
&\left(\prod_{i=j+1}^{m}\mathbf{T}^{s_i-s_{i-1}}\right)\mathbf{C}^{\pi_{m+1}}_{s_j,s}\left(\prod_{i=j+1}^{m}\mathbf{T}^{s_i-s_{i-1}\prime}\right) + \\
&\sum_{k=j+1}^{m}\left(\prod_{i=k+1}^{m}\mathbf{T}^{s_i-s_{i-1}}\right)\mathbf{C}^{\pi_{m+1}}_{s_i,s_{i-1}}\left(\prod_{i=k+1}^{m}\mathbf{T}^{s_i-s_{i-1}\prime}\right) \\
=& \mathbf{T}^{t-s_{m-1}}\Big(\mathbf{T}^{s_{m-1}-s_{m-2}}\cdot\Big(\cdots \\
&\cdot\Big(\mathbf{T}^{s_{j+2}-s_{j+1}}\left[\mathbf{T}^{s_{j+1}-s_j}\left\{\mathbf{T}^{s_j-s}\mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}}\mathbf{T}^{s_j-s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s}\right\}\mathbf{T}^{s_{j+1}-s_j\prime} + \mathbf{C}^{\pi_{m+1}}_{s_{j+1},s_j}\right]\mathbf{T}^{s_{j+2}-s_{j+1}\prime}\Big) \\
&+\cdots+\mathbf{C}^{\pi_{m+1}}_{s_{m-2},s_{m-3}}\Big)\mathbf{T}^{s_{m-1}-s_{m-2}\prime} + \mathbf{C}^{\pi_{m+1}}_{s_{m-1},s_{m-2}}\Big)\mathbf{T}^{t-s_{m-1}\prime} + \mathbf{C}^{\pi_{m+1}}_{s_m,s_{m-1}}.
\end{aligned} \tag{36}$$

The RHS of Eq. 35 follows the same form as Eq. 36 except with one fewer term.

$$\begin{aligned}
RHS =& \mathbf{T}^{t-s_{m-1}}\Big[\mathbf{T}^{s_{m-1}-s_{m-2}}\cdot\Big(\cdots \\
&\cdot\Big(\mathbf{T}^{s_{j+2}-s_{j+1}}\left[\mathbf{T}^{s_{j+1}-s_j}\left\{\mathbf{C}^{\pi_m}_{s_j,s_{j-1}}\right\}\mathbf{T}^{s_{j+1}-s_j\prime} + \mathbf{C}^{\pi_m}_{s_{j+1},s_j}\right]\mathbf{T}^{s_{j+2}-s_{j+1}\prime}\Big) \\
&+\cdots+\mathbf{C}^{\pi_m}_{s_{m-2},s_{m-3}}\Big)\mathbf{T}^{s_{m-1}-s_{m-2}\prime} + \mathbf{C}^{\pi_m}_{s_{m-1},s_{m-2}}\Big]\mathbf{T}^{t-s_{m-1}\prime} + \mathbf{C}^{\pi_m}_{s_m,s_{m-1}}.
\end{aligned} \tag{37}$$

We begin by comparing the inner most terms (denote by curly brackets) of Eq. 36, $\{\mathbf{T}^{s_j-s}\mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}}\mathbf{T}^{s_j-s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s}\}$, with the inner most term of Eq. 37, $\{\mathbf{C}^{\pi_m}_{s_j,s_{j-1}}\}$. We then develop a recursive mechanism to show that the inequality of Eq. 35 continues to hold as we expand outwards symmetrically according to the parentheses, encompassing larger groupings of terms.

$$\mathbf{T}^{s_j-s}\mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}}\mathbf{T}^{s_j-s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} \succeq_a \mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} \succeq_a \mathbf{C}^{\pi_m}_{s_j,s_{j-1}}. \tag{38}$$

The first inequality follows from the property that $\mathbf{T}$ is a progressing transformation and that $\mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}}$ is clearly positive semi-definite. The second inequality follows from the a-monotone property of the covariance estimate adjustments. To see this, consider a subsystem, with an initial observation at s_{j-1} and a final observation at s_j. The subset of policy π_m that intersects with this interval, $\pi_2 = \pi_m \bigcap [s_{j-1}, s_j]$ consists of only the initial and final observations, $\pi_2 = \{s_{j-1}, s_j\}$. On the other hand $\pi_3 = \pi_{m+1} \bigcap [s_{j-1}, s_j] = \pi_3 = \{s_{j-1}, s, s_j\}$. Since the two subsets of observations are being considered on the same interval and $\pi_2 \subset \pi_3$, we can invoke the a-monotonicity of the covariance estimate updates to show that the inequality holds. Eq. 38 can be rewritten as

$$\mathbf{T}^{s_j - s} \mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}} \mathbf{T}^{s_j - s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} - \mathbf{C}^{\pi_m}_{s_j,s_{j-1}} \succeq_a 0. \tag{39}$$

We now use the relationship in Eq. 39 to show that the $\succeq_a$ inequality also holds for the terms enclosed in the square brackets [and] by subtracting those terms in Eq. 37 (RHS) from the terms in Eq. 36 (LHS) and showing that the result is positive. First, Eq. 39 satisfies the positivity conditions in the definition of a progressing transformation, and since $\mathbf{T}$ is a progressing transformation, it follows that

$$\begin{aligned} &\mathbf{T}^{s_{j+1} - s_j} \left\{ \mathbf{T}^{s_j - s} \mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}} \mathbf{T}^{s_j - s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} - \mathbf{C}^{\pi_m}_{s_j,s_{j-1}} \right\} \mathbf{T}^{s_{j+1} - s_j{}'} \succeq_a \\ &\left\{ \mathbf{T}^{s_j - s} \mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}} \mathbf{T}^{s_j - s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} - \mathbf{C}^{\pi_m}_{s_j,s_{j-1}} \right\} \succeq_a 0. \end{aligned} \tag{40}$$

Now we can show that $\succeq_a$ holds for the terms of enclosed in the square brackets [and] Eq. 36 (LHS) and Eq. 37 (RHS) by showing that the subtraction of the square bracket terms of Eq. 37 (RHS) from those of Eq. 36 (LHS) is non-negative.

$$\begin{aligned} &\left[\mathbf{T}^{s_{j+1} - s_j} \left\{ \mathbf{T}^{s_j - s} \mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}} \mathbf{T}^{s_j - s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} - \mathbf{C}^{\pi_m}_{s_j,s_{j-1}} \right\} \mathbf{T}^{s_{j+1} - s_j{}'} + \mathbf{C}^{\pi_{m+1}}_{s_{j+1},s_j} - \mathbf{C}^{\pi_m}_{s_{j+1},s_j} \right] \succeq_a \\ &\left\{ \mathbf{T}^{s_j - s} \mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}} \mathbf{T}^{s_j - s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} - \mathbf{C}^{\pi_m}_{s_j,s_{j-1}} \right\} + \mathbf{C}^{\pi_{m+1}}_{s_{j+1},s_j} - \mathbf{C}^{\pi_m}_{s_{j+1},s_j} \succeq_a \\ &\mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} - \mathbf{C}^{\pi_m}_{s_j,s_{j-1}} + \mathbf{C}^{\pi_{m+1}}_{s_{j+1},s_j} - \mathbf{C}^{\pi_m}_{s_{j+1},s_j} \succeq_a 0, \end{aligned} \tag{41}$$

where the first inequality follows from Eq.'s 39 and 40. The second inequality follows from Eq. 38. The final inequality comparing a 3 observations with 2 observations follows from a-monotonicity of the covariance estimate updates using the subsystem that intersects the two policies, π_m and π_{m+1}, with the interval $[s_{j-1}, s_{j+1}]$ and using the same arguments that were previously used to show Eq. 38. It is clear that Eq. 41 is equivalent to

$$\begin{aligned} &\left[\mathbf{T}^{s_{j+1} - s_j} \left\{ \mathbf{T}^{s_j - s} \mathbf{C}^{\pi_{m+1}}_{s,s_{j-1}} \mathbf{T}^{s_j - s'} + \mathbf{C}^{\pi_{m+1}}_{s_j,s} \right\} \mathbf{T}^{s_{j+1} - s_j{}'} + \mathbf{C}^{\pi_{m+1}}_{s_{j+1},s_j} \right] \succeq_a \\ &\left[\mathbf{T}^{s_{j+1} - s_j} \left\{ \mathbf{C}^{\pi_m}_{s_j,s_{j-1}} \right\} \mathbf{T}^{s_{j+1} - s_j{}'} + \mathbf{C}^{\pi_m}_{s_{j+1},s_j} \right] \end{aligned} \tag{42}$$

We have now shown the first two steps of showing that the $\succeq_a$ relationship between Eq. 36 and Eq. 37 holds for increasingly large groups of terms. These arguments can be continued on each successive symmetric superset of terms (enclosed in parentheses) until the relationship Eq. 36 $\succeq_a$ Eq. 37 is established. The result could also easily be shown using induction on the number of terms in Eq. 36. Since the result did not depend on the particular interval into which the extra observation was inserted, it is clear that the result is general. The result also holds for the end points of the observation interval, $[1, s_1]$ and $[s_{m-1}, t]$, directly from the arguments above.

The general result for feasible $n > m+1$ can be shown by building up a series of policies $\pi_m \subset \pi_{m+1} \subset \cdots \subset \pi_n$. Using the arguments developed to show that $\Sigma_{t|t}^{\pi_m} \succeq_a \Sigma_{t|t}^{\pi_{m+1}}$, we can iteratively show that, for any feasible m,

$$\Sigma_{t|t}^{\pi_m} \succeq_a \Sigma_{t|t}^{\pi_{m+1}} \succeq_a \cdots \succeq_a \Sigma_{t|t}^{\pi_{n-1}} \succeq_a \Sigma_{t|t}^{\pi_n}. \tag{43}$$

Thus the result has been shown for arbitrary feasible m and $n > m$. □

Lemma 2 both yields Property 2 – more patient observations correlates with more certainty about whether the patient has progressed.

Property 3 (No. 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, which is 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, $F_{\rho,\tau}(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}^{\pi_m}) \leq F_{\rho,\tau}(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}^{\pi_n})$, where $F_{\rho,\tau}(\cdot,\cdot)$ is given by Eq. 20.*

Proof. To prove the theorem, we first show that for all $\ell = 0, 1, 2, \ldots$,

$$h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}^{\pi_m}, \ell) = \mathbf{a}'\mathbf{T}^\ell \hat{\alpha}_{t|t} + \sqrt{\chi^2(1-\rho, n)\mathbf{a}' \left(\mathbf{T}^\ell \hat{\Sigma}_{t|t}^{\pi_m} \mathbf{T}^{\ell\prime} + \sum_{j=0}^{\ell-1} \mathbf{T}^j Q \mathbf{T}^{j\prime} \right) \mathbf{a}} \geq$$

$$\mathbf{a}'\mathbf{T}^\ell \hat{\alpha}_{t|t} + \sqrt{\chi^2(1-\rho, n)\mathbf{a}' \left(\mathbf{T}^\ell \hat{\Sigma}_{t|t}^{\pi_n} \mathbf{T}^{\ell\prime} + \sum_{j=0}^{\ell-1} \mathbf{T}^j Q \mathbf{T}^{j\prime} \right) \mathbf{a}} = h_\rho(\hat{\alpha}_{t|t}, \hat{\Sigma}_{t|t}^{\pi_n}, \ell), \tag{44}$$

where $n > m$ and $\pi_m \subset \pi_n$ represent a policy with m observations and n observations on $[1,t]$ respectively. The first term on the RHS and LHS is the same, so it is only necessary to compare the terms under the square root. This is equivalent to showing that

$$\mathbf{a}'\left(\mathbf{T}^{\ell}\hat{\Sigma}_{t|t}^{\pi_m}\mathbf{T}^{\ell'}+\sum_{j=0}^{\ell-1}\mathbf{T}^{j}Q\mathbf{T}^{j'}\right)\mathbf{a}\geq\mathbf{a}'\left(\mathbf{T}^{\ell}\hat{\Sigma}_{t|t}^{\pi_n}\mathbf{T}^{\ell'}+\sum_{j=0}^{\ell-1}\mathbf{T}^{j}Q\mathbf{T}^{j'}\right)\mathbf{a}. \tag{45}$$

The $\mathbf{Q}$ terms cancel out. By Lemma 2 $\hat{\Sigma}_{t|t}^{\pi_m}-\hat{\Sigma}_{t|t}^{\pi_n}\succeq_a 0$, so it is possible to invoke the property of progressing transformation $\mathbf{T}$ to obtain the result

$$\mathbf{a}'\mathbf{T}^{\ell}\left(\hat{\Sigma}_{t|t}^{\pi_m}-\hat{\Sigma}_{t|t}^{\pi_n}\right)\mathbf{T}^{\ell'}\mathbf{a}\geq\mathbf{a}'\left(\hat{\Sigma}_{t|t}^{\pi_m}-\hat{\Sigma}_{t|t}^{\pi_n}\right)\mathbf{a}\geq 0 \tag{46}$$

Eq. 44 follows from Eq. 45, which follows directly from Eq. 46. Since $h_\rho(\hat{\alpha}_{t|t},\hat{\Sigma}_{t|t}^{\pi_m},\ell)\geq h_\rho(\hat{\alpha}_{t|t},\hat{\Sigma}_{t|t}^{\pi_n},\ell)$ for all $\ell=0,1,2,\ldots$, then clearly for a given progression threshold τ, the relationship holds for the time to next test optimization:

$$F_{\rho,\tau}(\hat{\alpha}_{t|t},\hat{\Sigma}_{t|t}^{\pi_n})=\min_{\ell\in\mathbb{Z}^+}\ell\text{ s.t. }\{h_\rho(\hat{\alpha}_{t|t},\hat{\Sigma}_{t|t}^{\pi_n},\ell)\geq\tau\}\geq F_{\rho,\tau}(\hat{\alpha}_{t|t},\hat{\Sigma}_{t|t}^{\pi_m})=\min_{\ell\in\mathbb{Z}^+}\ell\text{ s.t. }\{h_\rho(\hat{\alpha}_{t|t},\hat{\Sigma}_{t|t}^{\pi_m},\ell)\geq\tau\}$$

□

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." This property manifests itself clearly in the case study results of Section 5. 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.*

Proof. For a given progression threshold τ, by combining Theorem 1, Eq. 20 and the Kalman filter dynamics (Eq.'s 10 and 11, the time to next test becomes

$$F_{\rho,\tau}(\hat{\alpha}_i,\hat{\Sigma}_i)=\min_{\ell\in\mathbb{Z}^+}\ell\text{ s.t. }h_\rho(\hat{\alpha}_i,\hat{\Sigma}_i,\ell)=\mathbf{a}'\mathbf{T}^{\ell}\hat{\alpha}_i+\sqrt{\chi^2(1-\rho,n)\mathbf{a}'\left(\mathbf{T}^{\ell}\hat{\Sigma}_i\mathbf{T}^{\ell'}+\sum_{j=0}^{\ell-1}\mathbf{T}^{j}Q\mathbf{T}^{j'}\right)\mathbf{a}}\geq\tau. \tag{47}$$

By assumption $\mathbf{a}(\alpha_1-\alpha_2)>0$, and because $\mathbf{T}$ is a progressing transformation we have that $\mathbf{a}\mathbf{T}^{\ell}(\alpha_1-\alpha_2)\geq\mathbf{a}(\alpha_1-\alpha_2)>0$, and $\mathbf{T}^{\ell}\hat{\Sigma}_1\mathbf{T}^{\ell'}\succeq_a\mathbf{T}^{\ell}\hat{\Sigma}_2\mathbf{T}^{\ell'}$. Therefore, $h_\rho(\hat{\alpha}_1,\hat{\Sigma}_1,\ell)\geq h_\rho(\hat{\alpha}_2,\hat{\Sigma}_2,\ell)$. It follows that the optimal ℓ will be no greater for patient 1 than for patient 2. □

In the next section, we present a case study where these structural properties will be seen as applied to actual patients from a glaucoma clinical trial. We also show the value of using the TNT algorithm over current practice in glaucoma monitoring.

5. CIGTS Case Study Results

In this section we present an experiment based on historical data taken from the Collaborative Initial Glaucoma Treatment Study (CIGTS). In the experiment, we test our POMP TNT algorithm against the current practice of fixed intervals between tests, using three different fixed interval lengths: Short (1 year between tests), Medium (1.5 years between tests) and Long (2 years between tests). We begin by describing the available data and the design of the experiment. Then we present the aggregate results comparing the POMP TNT algorithm with fixed interval algorithms. Finally, patient treatment insights are explored using examples from the patient population.

5.1. Data and Design of Experiment

The CIGTS was a glaucoma clinical trial that followed 607 patients with newly-diagnosed glaucoma for up to 10 years. During the course of the trial, VF and IOP readings were taken every 6 months. During the trial some of the patients required further intervention due to inadequate control of glaucoma. The trial protocol required that further intervention to be Argon Laser Trabeculoplasty (ALT). We divided the data into training data (70% of the data) and test data (30% of the data). The training and test sets were selected randomly, but we maintained the original ratio between progressing and non-progressing patients in both the training and test sets.

Once the Kalman filter was trained, we ran the POMP TNT algorithm on all the test patients. We allowed the algorithm to run, scheduling tests for each patient, until either the trial finished for the patient or progression was detected in the patient. Progression was defined as either a patient who met the progression criteria in Section 3.3 or a patient who had ALT treatment.

To calibrate the POMP TNT algorithm, we tested different parameters for (1) the size of the prediction region and (2) the cutoff for determining whether progression has occurred using the logistic regression from Eq. 12. The results presented in Section 5.2 are from the algorithm with parameters $\rho = 0.75$ for prediction region size and progression threshold $\tau = 0.75$. In practice, the glaucoma specialist would adjust the levels dynamically, but fix them for simplicity.

5.2. Results and Insights

In this section we first present aggregate measures of the effectiveness of the POMP TNT algorithm. Next, we present some examples of how the algorithm behaves for different types of patients to support the insights into effective testing policies from Section 4.2. Finally, we discuss how the results from our POMP TNT algorithm support the hypothesis in the glaucoma community that variable testing intervals may perform better than fixed intervals and demonstrate how our algorithm provides a means to determine the length of the variable intervals.

Three aggregate effectiveness measures were considered when comparing the POMP TNT with fixed interval testing: (1) percent of cases that called for a test on the period in which a patient actually did progress (Accuracy, higher is better) (2) average number of periods that a patient's progression went undetected (Diagnostic Delay, lower is better) (3) average number of tests per patient (Number of Tests, lower is better). The results are presented in Table 4.

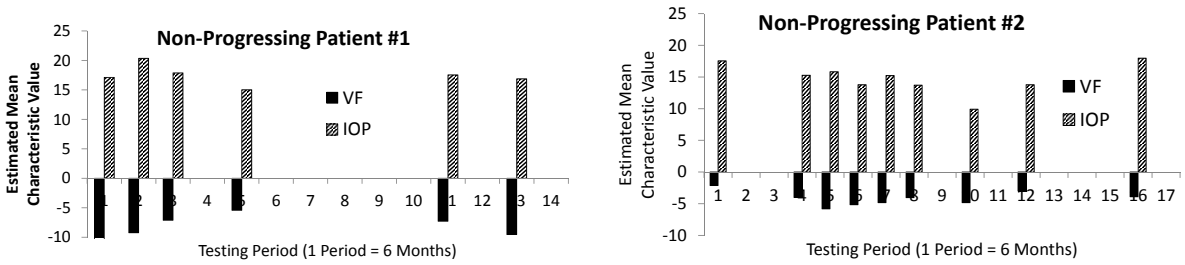
	POMP TNT	1 year fixed	1.5 year fixed	2 year fixed
Accuracy	86%	65%	54%	48%
Diagnostic Delay	0.46	0.36	0.70	1.05
Number of Tests	2.79	4.22	2.81	2.61

Table 4 Comparison of POMP TNT with Fixed Interval Testing.

The POMP TNT algorithm is decidedly more accurate than any of the fixed interval testing regimens and nearly as good or better in terms of diagnostic delay and number of tests. The comparison with the Short (1 year) fixed interval is particularly interesting because 1 year testing intervals would be considered too frequent for a significant proportion of glaucoma patients, and yet the POMP TNT algorithm is able to achieve nearly the same level of detection (diagnostic delay differs by only 18 days on average) or better (32% increase in accuracy) with 34% fewer tests. Similarly, POMP TNT registers nearly *twice the accuracy* of the 2 year fixed interval with less than *half the diagnostic delay* using nearly the same number of tests.

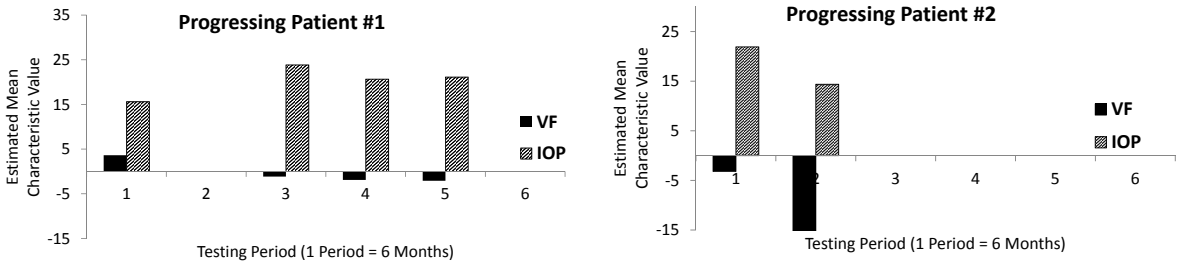
Next we present examples of how POMP TNT works on different patients to demonstrate the difference between the POMP algorithm and traditional fixed interval algorithms. For the purposes of exposition, we chose two progressing patients and two non-progressing patients. Figure 3 shows the values of VF and IOP measured at the testing points prescribed by POMP TNT. As a reminder, a high IOP is a risk factor for progression and a low VF score is an indicator of vision loss. The x-axis is divided into intervals with each period representing six months (based on the CIGTS test intervals). Periods with no VF and IOP values represent periods in which no measurement should be taken according to POMP TNT. Figure 4 represents two patients that progressed during the CIGTS trial, with the final measurement representing the period in which they progressed.

These graphs show how the POMP TNT algorithm intelligently uses the available information to make decisions based on a variety of complex factors. Non-progressing patient 1's glaucoma at the first reading appears to be fairly advanced as indicated by the low VF score, which would be of concern to a physician. This is consistent with Theorem 4. Because the patient's first reading is significantly in the direction of the progression threshold, the algorithm will test again in the next period. The next several tests, however, yield better VF scores indicating that the low VF in the



(a) Non-progressing patient #1 (b) Non-progressing patient #2

Figure 3 Two examples of POMP TNT applied to non-progressing patients.



(a) Progressing patient #1 (b) Progressing patient #2

Figure 4 Two examples of POMP TNT applied to progressing patients.

first reading was most likely a result of measurement noise. Once the algorithm gains confidence that the true state is farther from progression than initially believed, it starts to increase the length of time in between tests. This is consistent with the approach many glaucoma specialists take with respect to treatment scheduling. That is, when a patient's vision loss is significant already, the physician will be more cautious to preserve the remaining vision and avert complete blindness.

Non-progressing patient 2 illustrates the effect of the quantity of patient data on the POMP TNT approach. For this patient, the initial VF reading close to zero and an IOP score less than 20 would indicate only mild glaucoma. Consistent with the insights of Theorem 4, the model recommends a wait period of approximately 1.5 years between patient visits. When the VF score at the next test drops significantly, the algorithm signals a danger of disease progression, and the recommendation is now for more frequent monitoring. With subsequent tests, however, the patient's VF score appears to stabilize, and the readings from periods 4–8 reduce the variance around the state estimate. As in Theorem 3 and Lemma 2, the model triggers an increase in testing interval.

For progressing patient 1 of Fig. 4, the initial VF and IOP reading indicate only mild glaucoma. Based on Theorem 4 and the distance from progression, the second test is scheduled for a year (2 periods) later. At the second test, the patient's VF score has dropped and the IOP score has risen significantly with respect to the patient's baseline score (both occurrences signal danger of

progression). This causes the algorithm to schedule another test in the next period (period 4). The period 4 reading confirms the drop in VF with another low reading; however, the new reading is not low enough to signal progression yet, so the algorithm schedules another test for period 5. In period 5, the ProP function indicates that the patient has progressed so the algorithm terminates.

Progressing patient 2 has an initial VF reading that indicates fairly advanced glaucoma and they have a risk factor of a high IOP level. Thus, the second test is scheduled for the next period (6 months) later. The second reading yields a significantly lower VF score, which indicates that we have detected progression by our definition and hence the algorithm terminates.

6. 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 disease trajectory. Then the Probability of Progression (ProP) function was optimized over the Gaussian density of the Kalman filter to determine the Time to Next Test (TNT).

Beyond the ability to handle multidimensional state spaces, a key benefit of this approach is that the model output gives the full distribution, a multivariate Gaussian, on the patient's current state. 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 physician interaction by setting simple model parameters, enabling clinicians to complement their medical knowledge with the advanced statistical predictions. This approach will benefit both eye care professionals who treat glaucoma patients and potentially internists and other medical professionals who treat other chronic diseases.

Our validation study was based on data from the Collaborative Initial Glaucoma Treatment Study (CIGTS) 10 year clinical trial. It demonstrated that POMP TNT was able to outperform fixed interval regimens in terms of accuracy, diagnostic delay and efficiency (i.e. number of tests per patient). 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. This new monitoring approach significantly advances the methodology and practice of monitoring chronic diseases and glaucoma in particular.

References

[1] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. The optimal timing of living-donor liver transplantation. *Management Science*, 50(10):1420–1430, 2004.

[2] American Academy of Ophthalmology Glaucoma Panel. Preferred practice pattern guidelines. www.aao.org/ppp, 2010.

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

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

[5] D.P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1. Athena Scientific, Belmont, MA, third edition, 2000.

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

[7] Centers for Disease Control and Prevention CDC. Chronic disease prevention and health promotion. http://www.cdc.gov/chronicdisease/, February 2011.

[8] V. Chew. Confidence, prediction, and tolerance regions for the multivariate normal distribution. *Journal of the American Statistical Association*, 61(315):605–617, 1966.

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

[10] N.T. Choplin and R.P. Edwards. *Visual field testing with the Humphrey Field analyzer: a text and clinical atlas*. Slack, Thorofare, NJ, 1999.

[11] B.T. Denton, M. Kurt, N.D. Shah, S.C. Bryant, and S.A. Smith. Optimizing the start time of statin therapy for patients with diabetes. *Medical Decision Making*, 29(3):351, 2009.

[12] F.J. Fowler, J.E. Wennberg, R.P. Timothy, M.J. Barry, A.G. Mulley, and D. Hanley. Symptom status and quality of life following prostatectomy. *JAMA: The Journal of the American Medical Association*, 259(20):3018, 1988.

[13] N.M. Jansonius. Progression detection in glaucoma can be made more efficient by using a variable

interval between successive visual field tests. *Graefe's Archive for Clin Exp Ophtalmol*, 245(11):1647–1651, 2007.

[14] R.E. Kalman et al. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[15] Z. Kander. Inspection policies for deteriorating equipment characterized by N quality levels. *Nav Res Logist Q*, 25(2):243–255, 1978.

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

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

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

[19] C.P. Lee, G.M. Chertow, and S.A. Zenios. Optimal initiation and management of dialysis therapy. *Oper Res*, 56(6):1428–1449, 2008.

[20] P.P. Lee, J.G. Walt, J.J. Doyle, S.V. Kotak, S.J. Evans, D.L. Budenz, P.P. Chen, A.L. Coleman, R.M. Feldman, H.D. Jampel, et al. A multicenter, retrospective pilot study of resource use and costs associated with severity of disease in glaucoma. *Arch Ophthalmol*, 124(1):12, 2006.

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

[22] L. Meier III, J. Peschon, and R. Dressler. Optimal control of measurement subsystems. *IEEE T Automat Contr*, 12(5):528–536, 1967.

[23] R.C. Morey. Some stochastic properties of a compound-renewal damage model. *Oper Res*, 14(5):902–908, 1966.

[24] AG Munford and AK Shahani. A nearly optimal inspection policy. *Oper Res Q*, 23(3):373–379, 1972.

[25] K. Murphy. Kalman filter toolbox for Matlab. *Comp Sci and AI Lab., MIT, Cambridge, MA*, 1998.

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

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

[28] National Eye Institute NEI. Facts about glaucoma . http://www.nei.nih.gov/health/glaucoma/glaucoma_facts.asp, February 2011.

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

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

[31] Y. Oshman. Optimal sensor selection strategy for discrete-time state estimators. *IEEE T Aero Elec Sys*, 30(2):307–314, 1994.

[32] W.P. Pierskalla and J.A. Voelker. A survey of maintenance models: the control and surveillance of deteriorating systems. *Nav Res Logist Q*, 23(3):353–388, 1976.

[33] M.S. Rauner, W.J. Gutjahr, K. Heidenberger, J. Wagner, and J. Pasia. Dynamic policy modeling for chronic diseases: Metaheuristic-based Identification of pareto-optimal screening strategies. *Oper Res*, 58(5):1269–1286, 2010.

[34] G. Schell, M.S. Lavieri, J. Stein, and D. Musch. Filtering repeated measures data for improved disease progression identication. Working Paper, University of Michigan, 2011.

[35] S.M. Shechter, O. Alagoz, and M.S. Roberts. Irreversible treatment decisions under consideration of the research and development pipeline for new therapies. *IIE Transactions*, 42(9):632–642, 2010.

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

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

[38] W. Wu and A. Arapostathis. Optimal sensor querying: General markovian and lqg models with controlled observations. *IEEE T Automat Contr*, 53(6):1392–1405, 2008.

[39] R.H. Yeh. Optimal inspection and replacement policies for multi-state deteriorating systems. *Eur J Oper Res*, 96(2):248–259, 1997.