Using Kalman Filtering to Forecast Disease Trajectory for Patients With Normal Tension Glaucoma

GIAN-GABRIEL P. GARCIA, KOJI NITTA, MARIEL S. LAVIERI, CHRIS ANDREWS, XIANG LIU, ELIZABETH LOBAZA, MARK P. VAN OYEN, KAZUHISA SUGIYAMA, AND JOSHUA D. STEIN

• PURPOSE: To determine whether a machine learning technique called Kalman filtering (KF) can accurately forecast future values of mean deviation (MD), pattern standard deviation, and intraocular pressure for patients with normal tension glaucoma (NTG).

• DESIGN: Development and testing of a forecasting model for glaucoma progression.

• METHODS: We parameterized and validated a KF (KF-NTG) to forecast MD, pattern standard deviation, and intraocular pressure at 24 months into the future using 263 eyes of 263 Japanese patients with NTG. We determined the proportion of patients with MD forecasts within 0.5, 1.0, and 2.5 dBs of the actual values and calculated the root mean squared error (RMSE) for each forecast. We compared KF-NTG with a previously published KF model calibrated using patients with high-tension open-angle glaucoma (KF-HTG) and to 3 conventional forecasting algorithms.

• RESULTS: The 263 patients with NTG had mean ± standard deviation age of 63.4 ± 10.5 years. KF-NTG forecasted MD values 24 months ahead within 0.5, 1.0, and 2.5 dBs of the actual value for 78 eyes (32.2%), 122 eyes (50.4%), and 211 eyes (87.2%), respectively. The proportion of eyes with MD values forecasted within 2.5 dB of the actual value for the KF-NTG (87.2%) were similar to KF-HTG (86.0%) and the null model (86.4%), and much better than the 2 linear regression–based models (72.7-74.0%; $P < .001$). When forecasting MD, KF-NTG (RMSE = 2.71) and KF-HTG (RMSE = 2.68) achieved lower RMSE than the other 3 forecasting models (RMSE = 2.81-3.90), indicating better performance.

• CONCLUSION: As observed previously for patients with HTG, KF can also effectively forecast disease trajectory for many patients with NTG. (Am J Ophthalmol 2019;199:111–119. © 2018 Elsevier Inc. All rights reserved.)

N ORMAL TENSION GLAUCOMA (NTG) IS A COMMON form of open-angle glaucoma (OAG) that is characterized by irreversible structural damage to the optic nerve and the loss of nerve fiber layer tissue resulting in vision loss despite no record of intraocular pressure (IOP) >21 mm Hg. The prevalence of NTG varies throughout the world, though it is known to be common among Asians and Asian Americans.[1] Compared with other types of OAG, patients with NTG tend to experience dense visual field (VF) loss that can occur close to central fixation and earlier in the disease course. Given that central and paracentral visual field loss can make it difficult for patients to perform activities such as driving, reading, and other daily tasks, a personalized decision aid that can accurately forecast disease trajectory for patients with NTG and identify those who are at increased risk for progression before such damage occurs could greatly enhance patient outcomes.

Kalman filtering (KF) is an approach that has been used by engineers at the National Aeronautics and Space Administration to help guide Apollo missions to the moon.[2] Commonly used for decades by the aerospace industry to help guide flights, researchers have only recently applied this technique to forecast disease trajectory for patients with chronic diseases.[3–5] KF considers the underlying disease dynamics for a population of patients with the condition of interest along with the unique dynamics of a particular patient to generate personalized forecasts that can be dynamically updated each time a patient undergoes additional diagnostic testing.[6] We tested this approach on patients with high-tension OAG (HTG) and found that it was capable of effectively forecasting disease progression in that patient population.[7]

The purpose of the present study is to assess whether the same KF model that we developed for patients with HTG (KF-HTG) can effectively forecast disease trajectory for persons with NTG. We also sought to determine whether developing a new KF that is calibrated, parameterized,

and validated with data specifically from a population of patients with NTG (KF-NTG) can generate improved forecasts for patients with NTG compared with our existing KF-HTG model. Finally, we compared these 2 KF models with 3 conventional methods of forecasting glaucoma progression to determine which algorithm can most accurately predict disease trajectory dynamics for the most patients with NTG included in our study sample.

METHODS

THIS RETROSPECTIVE SECONDARY DATA ANALYSIS WAS approved by the University of Michigan and Fukui-ken Saiseikai Hospital institutional review boards and faithfully adhered to the tenets of the Declaration of Helsinki.

• DATA SOURCES: We were interested in forecasting the glaucoma trajectory for a convenience sample of 263 Japanese patients with NTG under the care of one of the study authors (K.N.) at the Fukui-ken Saiseikai Hospital in Fukui, Japan. Each patient had 4 to 6 years of longitudinal monitoring with standard automated perimetry and IOP measurements between June 1, 2009, and May 31, 2015. Data from clinic visits were used to develop our KF-NTG model. We compared the performance of KF-NTG to KF we previously developed using data on patients with HTG from the Advanced Glaucoma Intervention Study (AGIS) and the Collaborative Initial Glaucoma Treatment Study (CIGTS) randomized clinical trials, which we hereafter refer to as KF-HTG.[7] The AGIS involved 591 adults with advanced OAG who were enrolled at 11 clinical centers between 1988 and 1992 and who were followed until 1996.[8] In AGIS, participants were randomly assigned to receive either argon laser trabeculoplasty or trabeculectomy. VF and IOP measurements were obtained at baseline and every 6 months for up to 8 years to monitor for disease progression. There were 607 persons with newly diagnosed early to moderate OAG who were enrolled in CIGTS across 14 clinical centers between 1993 and 1997.[9,10] In the CIGTS, participants were randomly assigned to treatment consisting of either IOP-lowering medications or trabeculectomy and monitored at baseline and every 6 months for up to 11 years to check for disease progression.

• SAMPLE SELECTION: Following established guidelines for diagnosing NTG as endorsed by the Japan Glaucoma Society[11] and the European Glaucoma Society,[12] we identified a convenience sample of 263 eyes of 263 patients who had been diagnosed with NTG and who had ≥4 years of follow-up. Specifically, these patients had open angles on gonioscopy, changes to the optic nerve consistent with glaucomatous optic neuropathy as judged by a glaucoma subspecialist (K.N.), glaucomatous VF loss on perimetry as judged by a glaucoma subspecialist (K.N.), and ≥3 baseline untreated IOP measurements ≤21 mm Hg and no baseline readings >21 mm Hg. Using criteria in line with the Hodapp-Anderson-Parrish Glaucoma Grading Scale,[13] patients with mild (baseline MD > −6 dB), moderate (baseline MD between −6 dB and −12 dB), and severe (baseline MD < −12 dB) NTG were included in the study. Hereafter, we refer to this cohort as the NTG group.

All participants in AGIS and CIGTS had primary or secondary forms of OAG with elevated IOPs at baseline in ≥1 eye. In AGIS, all participants had IOPs of ≥22 mm Hg at baseline or before entry into the trial and glaucomatous VF loss.[8] In the CIGTS, trial participants were eligible if they had IOPs ≥21 mm Hg without treatment along with evidence of glaucomatous damage on standard automated perimetry.[9] None of the participants from the AGIS or CIGTS trials had NTG. Hereafter, we refer to an aggregate dataset of patients from these 2 clinical trials as the HTG group.

• INCLUSION AND EXCLUSION CRITERIA: To be included in our analyses, we required all persons in the NTG and HTG groups to contribute at least 4 Swedish Interactive Threshold Algorithm Standard 24-2 Humphrey VFs (HVFs) (Carl Zeiss Meditec, Dublin, CA, USA) and 4 IOP measurements on separate dates over the course of a period of ≥2 years' duration. Abrupt changes in glaucoma progression dynamics may be challenging to accurately forecast, and therefore we excluded all persons who underwent trabeculectomy or other incisional glaucoma surgeries. If a patient underwent one of these surgeries during the follow-up period, we still included him or her but censored all IOP and HVF measurements beginning at the time at which they underwent surgery. For the NTG and HTG groups, we randomly chose 1 eye if both eyes were eligible.

• FORECASTING FUTURE VALUES OF MEAN DEVIATION, PATTERN STANDARD DEVIATION, AND IOP: *Data elements.* For persons with NTG and HTG, we obtained each patient's demographic information (ie, age, sex, and race/ethnicity) along with all mean deviation (MD) and pattern standard deviation (PSD) measurements from HVFs and IOP measurements from the baseline readings until the end of the follow-up period. These data elements were used to parameterize and train our models. Our KF models required relatively evenly spaced testing time intervals. Following the protocols of the AGIS and CIGTS trials and clinical practice, most of the testing of the patients with NTG and HTG was done in 6-month intervals. Occasionally, linear interpolation was required to obtain readings that were evenly spaced every 6 months, eg, if a patient happened to miss a study or clinic visit. We computed rates of change (ie, velocity and acceleration) at each visit for MD, PSD, and IOP.

To compute velocity, we took the slope of 3 measurement readings at consecutive examinations; to compute acceleration, we took the slope between velocities at 2 consecutive examinations. We compared the NTG and HTG groups using a 2-sample Student t test for continuous variables (eg, age, MD, PSD, and IOP) and a Pearson χ^2 test for categorical variables (eg, sex and race/ethnicity).

KF. KF is a statistical model that incorporates population-level disease progression dynamics along with individual patient readings to forecast future levels of a given variable for each patient.[6] For these analyses, we applied KF to forecast future values of MD, PSD, and IOP for persons with NTG. As new readings are obtained for a patient, the model dynamically updates and self-corrects based on past prediction errors, improving its ability to accurately forecast future values of MD, PSD, and IOP over time. We created a KF model to predict future measurements of MD, PSD, and IOP based on their current values, velocities, and accelerations. Before our KF can begin forecasting, a minimum of 3 measurements over a 12-month period are required to ensure that velocity and acceleration could be computed. Based on recommendations of earlier research,[14,15] we used the first 6 measurements of IOP, MD, and PSD to train our KF before using it to forecast future values. As an ancillary analysis, we assessed whether the forecasts generated from our KFs were similar when using only the first 3 measurements of IOP, MD, and PSD before the model started forecasting future values instead of requiring 6 baseline measurements.

• ANALYSES: *Model parameterization.* We created 2 different parameterizations of the KF. The first was identical to our previously validated and published algorithm[7] using data from the AGIS and CIGTS trials (KF-HTG) to determine how well it can forecast glaucoma progression for persons with NTG. For comparison, we created a second parameterization of the KF using data exclusively from the NTG cohort (KF-NTG) to determine whether it would forecast more accurately compared with KF-HTG. We compared these KF models with one another and with 3 conventional forecasting models: a null model and 2 linear regression models (detailed below).

Null model. We compared KF-HTG and KF-NTG with a null model, which assumes that every 24 months MD decreases by 0.70 dB, PSD increases by 0.76 dB, and IOP remains unchanged for all patients with NTG. For example, a patient whose current MD reading is −1.2 dB would be predicted to have an MD reading of −1.9 dB 24 months into the future. These rates of change are consistent with mean rates reported in previous studies of

patients with NTG who have not undergone incisional intraocular surgery.[16–22]

Linear regression models. Previous research has found that linear estimates for rates of progression predict future VF loss more accurately for patients with NTG compared with more complex, nonlinear models,[23] and therefore we also compared KF-HTG and KF-NTG with 2 linear regression models created using data from the NTG cohort. These linear regression models are similar to those used for forecasting with the VF index algorithm in the Humphrey perimeter.[15] Our linear regression prediction models are of the form:

$$y = b_0 + b_1 t,$$

where b_0 is the intercept term, b_1 is the slope, t is time in months since the last measurement, and y is the forecasted parameter (eg, future values of MD). Both b_0 and b_1 are estimated from the 6 most recent measurements for the parameter of interest. The first linear regression model, LR1, is a simple linear regression model. The second linear regression model, LR2, is a common econometric forecasting model that has the same slope as LR1 but passes through the most recent observation.[24]

To illustrate the difference between LR1 and LR2, consider a patient whose 6 previous MD readings were −1.0, −1.1, −1.3, −1.4, −1.2, and −1.2 dB. Using standard least squares regression formulas,[25] the LR1 model for predicting the future value of MD in t months is:

$$MD \text{ at visit } t = -1.3 - 0.00667t \text{ dB}.$$

LR2 replaces the intercept term in the model (ie, −1.3) with the most recent reading (ie, −1.2). Therefore, LR2 predicts future values of MD using the equation:

$$MD \text{ at visit } t = -1.2 - 0.00667t \text{ dB}.$$

Using these equations, LR1 would predict MD to be −1.46 dB while LR2 would predict MD to be −1.36 dB 24 months into the future.

For all 5 models, we were interested in how well they forecasted values of MD, PSD, and IOP at 24 months into the future for the cohort of patients with NTG. Forecasting 24 months ahead required ≥10 HVF and IOP measurements, where the first 6 measurements are used for learning and the tenth measurement (ie, 24 months after the learning period) is used to compare against the model forecasts.

• PERFORMANCE MEASURES: We used all eligible patients with NTG to test our models. We assessed model performance by evaluating prediction errors 24 months into the future, where prediction error is defined as the difference between the future measurement prediction and the actual value from the clinic visit. We first analyzed the distribution of MD prediction errors by determining the proportion of prediction errors between clinically relevant thresholds,

ie, ±0.5, ±1.0, and ±2.5 dBs from the actual value obtained from the clinic visit. We then compared the distribution of MD prediction errors across these thresholds between KF-NTG and all other models using the Bhapkar test for equality of marginal distributions.[26]

Next, we evaluated how well each of the 5 models forecasted future values of MD, PSD, and IOP by determining the root mean squared error (RMSE) of each prediction. RMSE is determined by squaring all prediction errors, taking the mean, and then taking the square root of this mean. Lower RMSE implies less prediction error, ie, the closer the RMSE is to 0, the more aligned the predicted value is with the actual value obtained from the clinic visit. Another way to compare the model forecasts is by calculating the mean absolute error (MAE). Much like RMSE, MAE measures the average magnitude of forecasting errors and is computed by taking the absolute value of all prediction errors and then taking the mean. MAE has a similar interpretation to RMSE (ie, lower MAE implies lower magnitude of prediction error). Compared with using just 1 of these measures, evaluating our models based on both RMSE and MAE provides a more holistic perspective on our model performance. For the KF-NTG model, we used leave-one-out cross validation to obtain nearly unbiased estimates of the prediction error.[25] That is, for every patient with NTG, a separate KF model was created using data from all other patients except for the particular patient for whom we were generating a forecast.

• OUTLIER ANALYSIS: For each of the 5 models, we identified patients whose MD prediction errors 24 months into the future were >2.5 dB off from the actual value obtained from the trial or clinic visit (ie, outliers). We computed the proportion of persons who were outliers in our KF-NTG model who were also outliers in the other models. We also analyzed whether patients for whom the KF-NTG model predicted poorly (the outliers) differed from those for whom the model predicted well in terms of their demographics and baseline and follow-up readings for MD, PSD, and IOP using analysis of variance or Pearson χ^2 tests, as appropriate.

RESULTS

• STUDY SAMPLE: The cohort of patients with NTG used to develop the KF-NTG model included 263 eyes of 263 patients. These included eyes with mild, moderate, and severe disease (Supplemental Figure 1; Supplemental Material available at AJO.com). There was a mean ± standard deviation (SD) of 5.9 ± 0.5 years of longitudinal follow-up on these patients. The mean ± SD age at baseline of the group was 63.4 ± 10.5 years. There were 159 males (60.5%) in the group and all patients were Asian. Our previously created KF-HTG model was developed using data from 601 eyes of 601 participants from the

AGIS and CIGTS trials. Three hundred twenty-four (53.9%) of the patients in this group were recruited from the AGIS and 277 (46.1%) were recruited from the CIGTS. The mean ± SD follow-up duration was 6.3 ± 2.8 years for the HTG cohort. The mean ± SD age of the HTG group was 66.1 ± 10.9 years. In this group, 291 persons (48.4%) were male and there were 275 whites (45.8%), 304 blacks (50.6%), 9 Asians (1.5%), and 13 (2.2%) who were classified as some other race or ethnicity. The NTG group had a significantly greater decline in MD ($P = .01$) and rise in PSD ($P < .001$) compared with the HTG group (Table 1).

• KF FORECASTING OF GLAUCOMA PROGRESSION DYNAMICS: Using the first 6 observations to teach the KF the progression dynamics for each person with NTG, we forecasted MD, PSD, and IOP values 24 months into the future for all 263 patients with NTG. There were 242 (92.0%) patients with enough measurements to forecast 24 months into the future. Our KF-NTG model forecasted MD within 0.5 dB of the actual value for 78 eyes (32.2%), within 1 dB for 122 eyes (50.4%), and within 2.5 dB for 211 eyes (87.2%) with NTG. The forecasted values differed by >2.5 dB from the actual values for 31 eyes (12.8%). By comparison, when we applied KF-HTG to our cohort with NTG, KF-HTG forecasted MD values 24 months into the future within 0.5 dB of the actual value for 72 eyes (29.8%), within 1 dB for 127 eyes (52.5%), and within 2.5 dB for 208 eyes (86.0%). The MD predictions for KF-HTG differed by >2.5 dB from the actual value for 34 eyes (14.0%) (Table 2). Figure 1 shows the forecast of a sample patient with NTG using KF-NTG and how it compared with actual MD readings obtained in the clinic.

• DISTRIBUTION OF MD PREDICTION ERRORS: We next evaluated the distribution of prediction errors for KF-NTG and KF-HTG when forecasting future values of MD for patients with NTG. Figure 2 presents violin plots of the distribution of MD prediction errors for each model. Violin plots that are thinner and longer show greater variability in prediction errors relative to plots that are shorter and fatter. Table 2 provides a quantitative summary of these MD prediction errors with reference to the clinically relevant threshold values. The KF models and the null model outperformed LR1 and LR2. The proportion of eyes with MD values within 2.5 dB of the actual value was 86.0% for KF-HTG, 87.2% for KF-NTG, and 86.4% for the null model while it was 72.7% to 74.0% for the other 2 models. The distribution of prediction errors for the LR1 and LR2 models were each significantly different from that of KF-NTG ($P < .001$ for both). We also analyzed the performance of each model when the NTG group was stratified into mild NTG (MD > −6 dB, n = 141) and moderate/severe NTG (MD ≤ −6 dB, n = 101) subgroups. We found that KF-NTG, KF-HTG, and the null model perform similarly well for forecasting MD, with 89.4% to 91.5% of

TABLE 1. Description of the Study Sample

	Persons With NTG	Persons With HTG	P Value[a]
Eyes, N	263	601	
Patients, N	263	601	
Sex, n (%)			<.001
Male	159 (60.5)	291 (48.4)	
Female	104 (39.5)	310 (51.6)	
Race, n (%)			<.001
White	0	275 (45.8)	
Black	0	304 (50.6)	
Asian	263 (100)	9 (1.5)	
Other	0	13 (2.2)	
Baseline glaucoma severity,[b] n (%)			.02
Mild	148 (59.7)	273 (47.6)	
Moderate	57 (23.0)	169 (29.4)	
Severe	43 (17.3)	132 (23.0)	
Total no. of visits	9412	8228	
VFs per patient, mean (SD)	12.6 (4.4)	13.7 (5.5)	.01
IOPs per patient, mean (SD)	30.0 (16.8)	13.7 (5.5)	<.001
Follow-up length (y), mean (SD)	5.9 (0.5)	6.3 (2.8)	<.001
Age at baseline (y), mean (SD)	63.4 (10.5)	66.1 (10.9)	<.001
Initial MD (dB), mean (SD)	−6.2 (5.9)	−7.5 (5.6)	<.001
Initial PSD (dB), mean (SD)	7.8 (4.6)	6.4 (3.7)	<.001
Initial IOP (mm Hg), mean (SD)	13.7 (2.9)	17.7 (4.2)	<.001
MD change,[c] mean (SD)	−0.8 (2.2)	−0.3 (3.2)	.01
PSD change,[c] mean (SD)	0.7 (1.7)	0.1 (1.8)	<.001
IOP change,[c] mean (SD)	0.0 (1.9)	−0.3 (4.2)	.38

dB = decibel; HTG = high-tension glaucoma; IOP = intraocular pressure; MD = mean deviation; NTG = normal tension glaucoma; PSD = pattern standard deviation; SD = standard deviation; VF = visual field.

[a]P values for sex, race, and initial glaucoma severity were computed using the Pearson χ^2 test for independent samples. All other P values were computed using a 2-sample Student t test.

[b]Mild = initial MD > −6 dB, moderate = initial MD between −6 and −12 dB, severe = initial MD < −12 dB.

[c]Change computed as: reading at 24 months − initial reading.

forecasts within 2.5 dB of the actual value for the mild NTG subgroup and 81.2% across all 3 models for the moderate/severe NTG subgroup. The LR1 and LR2 had worse performance for patients with mild and moderate/severe NTG compared with the other 3 models (Supplemental Table 1; Supplemental Material available at AJO.com).

• COMPARISON OF OVERALL PREDICTIVE CAPABILITIES OF MODELS: Table 3 shows the RMSE performance for each model when forecasting MD, PSD, and IOP 24 months into the future. Compared with the LR1 model (which is similar to the forecasting method used in previous studies[23] and with Zeiss' VF index[15]), both KF models show improved performance. When forecasting MD, KF-NTG (RMSE = 2.71) and KF-HTG (RMSE = 2.68) outperformed the other 3 models (RMSE = 2.81-3.90). Similar findings were attained when forecasting future values of PSD whereby the 2 KF models had similar or lower RMSEs compared with

the other 3 models. When forecasting IOP 24 months into the future, KF-NTG (RMSE = 2.10) and KF-HTG (RMSE = 2.11) achieved the lowest RMSE of the 5 algorithms (Table 3).

We also analyzed RMSE for each model when the NTG group was stratified into those with mild NTG and patients with moderate/severe NTG at baseline. For the mild NTG subgroup, the KF-NTG, KF-HTG, and null models had lower RMSE than the LR1 and LR2 models when forecasting MD (RMSE = 2.68-2.74), PSD (RMSE = 1.86-1.94), and IOP (RMSE = 1.85-1.92). For the moderate/severe NTG subgroup, KF-NTG and KF-HTG had lower RMSE than the null, LR1, and LR2 models when forecasting MD (RMSE = 2.59-2.66), PSD (RMSE = 1.68-1.71), and IOP (RMSE = 2.40-2.41) (Supplemental Table 2; Supplemental Material available at AJO.com).

As an ancillary analysis, our KF-NTG achieved a lower RMSE when using 6 baseline observations compared with requiring only using 3 baseline observations, though the

Amount of Error in Forecast (dB)	KF-NTG Eyes, n (%)	KF-HTG Eyes, n (%)	Null Eyes, n (%)	LR1[a] Eyes, n (%)	LR2[a] Eyes, n (%)
≤0.5	78 (32.2)	72 (29.8)	70 (28.9)	38 (15.7)	44 (18.2)
≤1.0	122 (50.4)	127 (52.5)	136 (56.2)	81 (33.5)	85 (35.1)
≤2.5	211 (87.2)	208 (86.0)	209 (86.4)	176 (72.7)	179 (74.0)
>2.5	31 (12.8)	34 (14.0)	33 (13.6)	66 (27.3)	63 (26.0)

dB = decibel; KF-HTG = Kalman filter built using a sample of patients with high-tension glaucoma; KF-NTG = Kalman filter built using a sample of patients with normal tension glaucoma; LR = linear regression; MD = mean deviation.

[a]This model had a significantly different distribution of prediction errors compared with the KF-NTG model when predicting MD 24 months into the future at a significance level of .01 based on the Bhapkar test for equality of marginal distributions. Two hundred forty-two of 263 (92.0%) patients with normal tension glaucoma had enough measurements for this analysis.
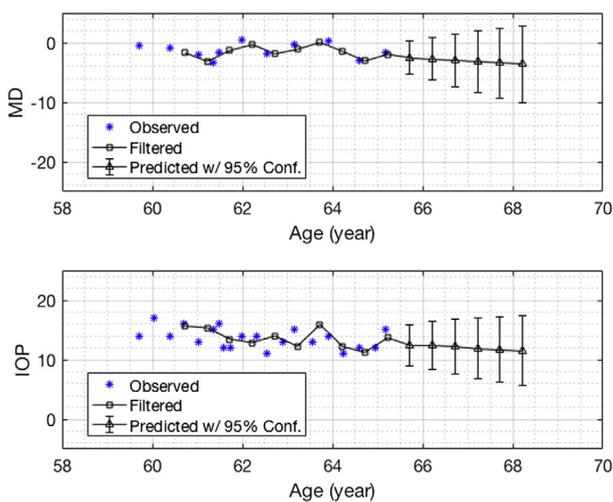

FIGURE 1. Kalman filter forecast of sample patient with normal tension glaucoma using the Kalman filter model trained on patients with normal tension glaucoma algorithm. Observed values are those obtained when the patient took the test in the clinic. Filtered readings are the estimated values by our Kalman filter model. Predicted measurements are future forecasted values predicted by our Kalman filter model. IOP = intraocular pressure; MD = mean deviation.
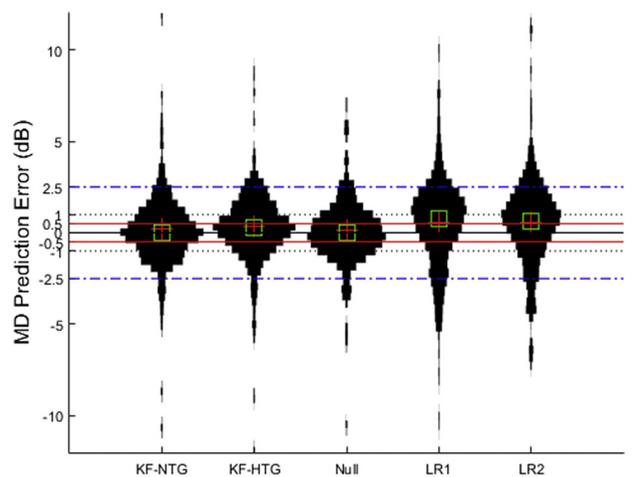

FIGURE 2. Violin plots showing the proportion of patients in each of the 5 forecasting models with prediction errors of mean deviation (MD) within 0.5, 1.0, and 2.5 dB from the actual value at 24 months into the future. The models compared are a Kalman filter model trained on patients with normal tension glaucoma (KF-NTG), a Kalman filter model trained on patients with high tension glaucoma (KF-HTG), a null model that assumes a constant 24-month change of −0.70 dB for MD, 0.76 dB for PSD, and no change for intraocular pressure (null), a simple linear regression model (LR1), and a modified linear regression model (LR2). Plots that are thinner and taller have more variability in their prediction errors relative to others that are shorter and wider. dB = decibel; PSD = pattern standard deviation.

improvement in predictive accuracy was minimal (data not shown). We also performed another ancillary analysis in which we evaluated the magnitude of forecasting errors for all 5 models using MAE. The results of this MAE analysis were similar to our RMSE analysis wherein the KF models outperformed the other models in forecasting MD, PSD, and IOP. However, using MAE, the null model performed relatively similar to the KF models in forecasting MD (data not shown).

• DESCRIPTION OF OUTLIERS: Most patients who were outliers in the KF-NTG model were also outliers for the other 4 models. For example, there were 31 patients (12.8%) whose predicted MD were ±2.5 dBs than the actual value. Of those 31 patients, 22 (71.0%) were also outliers for the KF-HTG model, 20 (64.5%) for the null model, and 26 (83.9%) for both the LR1 and LR2 models. We compared patients whose future MD readings were poorly predicted by KF-NTG with those patients who were well-predicted and found no significant differences in sex or age between outliers and nonoutliers. (Supplemental Table 3; Supplemental Material available at AJO.com).

TABLE 3. Comparison of the Root Mean Square Error of the 5 Models at Forecasting Key Glaucoma Metrics at 24 Months into the Future

Metric	KF-NTG		KF-HTG		Null		LR1	LR2	
	RMSE[a]	Improvement, %	RMSE	Improvement, %	RMSE	Improvement, %	RMSE	RMSE	Improvement, %
MD	2.71	30.6	2.68	31.2	2.81	28.0	3.90	3.71	4.8
PSD	1.85	31.6	1.85	31.7	2.01	25.5	2.70	2.56	5.3
IOP	2.10	31.0	2.11	31.0	2.34	23.5	3.05	3.22	−5.6

KF-HTG = Kalman filter built using a sample of patients with high-tension glaucoma; KF-NTG = Kalman filter built using a sample of patients with normal tension glaucoma; LR = linear regression; RMSE = root mean square error.

RMSE values closer to 0 indicate predictions closer to the actual values obtained in the trial/clinic. Improvement (%) was measured with respect to the LR1 model and computed as $(RMSE_{LR1} - RMSE_M)/(RMSE_{LR1})$, where $RMSE_M$ is the RMSE belonging to the KF-NTG, KF-HTG, null, or LR2 model. Positive improvement indicates improved performance compared with the LR1 model.

[a]RMSE for KF-NTG was computed using leave-one-out cross-validation.

DISCUSSION

KF IS AN INNOVATIVE MACHINE LEARNING METHODOLOGY that integrates data from large populations of patients with a given condition of interest along with past measurements from the actual patient of interest to help clinicians generate personalized forecasts of the disease trajectory for patients with chronic diseases. Research from our group has shown that these algorithms work well at forecasting future values of MD, PSD, and IOP for patients with HTG.[7] The results of the present study show that the KF algorithm we developed to forecast future values of key parameters for patients with HTG based on data from the AGIS and the CIGTS seems to work well when tested on a group of patients from Japan with NTG. While one expects to see that a KF developed specifically from a population of patients with NTG (KF-NTG) would outperform the KF-HTG at forecasting disease trajectory for patients with NTG, the additional improvement in RMSE we observed in the present study was rather modest. This result suggests that KF-HTG appears to perform well for most patients with NTG. Another important finding of these analyses it that both KF models appeared to forecast as well as—if not better than—existing forecasting techniques.

There is debate in the literature regarding whether NTG is a fundamentally different entity than HTG[16,27] or if they are both a continuum of the same disease entity,[28–31] but it has been established that patients with NTG tend to experience progressive VF loss at lower levels of IOP relative to those with HTG. In addition, their VF loss tends to be denser and closer to central fixation earlier in the disease course. Therefore, we were unsure, a priori, whether the model we had previously developed using a population of patients with HTG would forecast well when tested on persons with NTG. Our results seem to indicate that regardless of differences in the demographic characteristics of the patients with NTG versus those with HTG or their baseline levels of MD, PSD, or IOP, both KF models seem to perform well at forecasting future values of MD, PSD, and IOP for patients with NTG. We suspect that KF-HTG fared well in this new population because the methodology by which KF generates the forecasts involves learning about the unique disease dynamics of each individual patient and integrating this into the forecasts. Therefore, when we test our KF-HTG model on persons with NTG, we are noticing that KF-HTG appears to quickly learn that these patients tend to progress at lower IOP levels than the underlying population with HTG for which our model was derived. KF-HTG adjusts its forecasts accordingly such that, over time, they are less influenced by the underlying population and more by the actual patients in the testing set.

Comparing our 2 KF models with 3 other models (ie, a null model that assumes a constant rate of −0.70 dB for MD, 0.76 dB for PSD, and no change for IOP over 24 months and 2 linear regression models that assume changes to a patient's progression dynamics occur in a linear manner over time), we find that the KF outperformed these simpler models when forecasting 24 months into the future. In fact, compared with LR1 and LR2, which are similar to the forecasting method used in previous studies[23] and the Humphrey perimeter's VF index,[15] the KF models show considerable improvement in forecasting accuracy. Furthermore, we note that while the simpler null model we tested seems to perform relatively similarly compared with the KF models on selected forecasts, it relies on the assumption that the patient is known to behave like the population of patients with NTG reported in the literature from which the model was derived. Furthermore, the null model may not work as well on patients whose VFs progress much more slowly or quickly compared with the MD and PSD progression rates we used based on the published literature. This point raises concern, because previous studies have shown that patients with NTG generally experience great variability in VF progression.[17] Overall, we suspect that these findings stems from the fact that KF models 1) base their predictions on both the patient and underlying

population, 2) dynamically reassess the disease trajectory (by considering velocity and acceleration of the model parameters) with each new measurement, and 3) are designed to identify and reduce noise that is inherent in measurements of IOP and VF parameters.

With both KF-HTG and KF-NTG, there were only a small subset (12.8-14.0%) of patients for whom the forecasted MD values were >2.5 dB from the actual values. We studied these patients in detail to analyze which features made them difficult to forecast. These patients did not differ from well-predicted patients in terms of patient age and sex. Moreover, most patients whose MD values fell outside 2.5 dB of the actual value for the 2 KF models also could not be forecasted well with the other 3 models either. These patients may be poor VF takers or have underlying comorbid medical conditions that affected their testing performance for which our KF simply could not take into account when generating its forecasts.

As an ancillary analysis, we tested our algorithms using only 3 baseline observations for a given patient instead of 6 baseline observations and found that while the KFs tended to forecast better using more baseline observations, they still seemed to function quite well with only 3 observations. This finding is important for clinical practice because clinicians caring for patients often lack 6 sets of readings on many of their patients. Since our tool appears to function well with only 3 sets of baseline readings, eye care providers should be able to start using the tool sooner, knowing that as they obtain more additional measurements the KF forecasts will improve over time.

From a practical perspective, the KF models we are developing have the potential to greatly enhance clinical decision making. We have already developed a tool that uses KF-HTG to aid clinicians with determining personalized and dynamically updated assessments of how soon to perform additional diagnostic testing[7] and a personalized menu of target IOPs[32] for patients with HTG. It is possible to enhance the KF-NTG such that it has similar functionalities.

• STRENGTHS AND LIMITATIONS: Our sophisticated forecasting approach using KF was able to accurately forecast 24 months into the future the level of MD to within 0.5 dB of the actual value for approximately one-third of the patients with NTG we studied and within 1.0 dB of the actual value for roughly half of all the patients with NTG. We were also able to demonstrate the flexibility of the KF approach such that a model that was trained on patients with HTG could quickly adapt itself to a new cohort of patients with NTG. Our study is not without limitations. We were limited by the sample size of patients with NTG who met our study inclusion criteria. With larger study populations, KF can better learn the underlying disease trajectory for patients with conditions such as NTG. We would expect KF-NTG to achieve greater accuracy if we parameterized, calibrated, and validated it on a larger sample of patients with NTG. Second, we were limited by the amount of follow-up data for each patient. For example, extending our analysis to forecast 36 months into the future would require at least 12 HVFs and IOPs—a criterion which was met by only 27 patients. Third, all the patients with NTG we analyzed were from Japan. Additional research is needed to determine whether the accuracy of our predictions are directly generalizable to patients with NTG residing in other countries—though we see no reason why this approach will not work for other patients with NTG once the KF learns whether the progression dynamics are similar or dissimilar to those in this particular patient cohort. Regardless, extending our analysis to a larger cohort of patients with NTG (ie, not just those with NTG from Japan) would make our models more generalizable to all patients with NTG. Fourth, our models have yet to consider other factors, such as the results of structural tests like optical coherence tomography, pachymetry, and the presence or absence of disc hemorrhages. Because the AGIS and CIGTS studies did not routinely collect these data, we could not incorporate them into our earlier KF-HTG model. To facilitate comparisons between KF-HTG and KF-NTG, we opted not to include these variables into the KF-NTG we developed. We are presently building an enhanced KF-NTG that incorporates these and other variables, such as pointwise regression analysis of individual test points on the pattern deviation plot to assess for focal changes on the VF over time and hope that it will further improve the accuracy of our predictions.

In conclusion, KF is a promising approach that can be used to help learn the future disease trajectory of groups of patients with glaucoma and be used to generate personalized forecasts. The present sets of analyses highlight the adaptability of the KF approach such that it can be parameterized, calibrated, and validated on patients with HTG. When applied to a new cohort of patients with NTG, the model is able to adapt and learn that many of these patients are at risk of experiencing disease progression at levels of IOP that did not result in progression for those with HTG. As we continue to refine our KF models by incorporating additional variables, and by performing additional validation studies, we hope to soon be able to make them accessible to the clinical community so they can assist clinicians with making more informed decisions on how to manage patients with sight-threatening conditions such as NTG.

REFERENCES

1. Stein JD, Kim DS, Niziol LM, et al. Differences in rates of glaucoma among Asian Americans and other racial groups, and among various Asian ethnic groups. *Ophthalmology* 2011;118(6):1031–1037.
2. Lefferts EJ, Markley FL, Shuster MD. Kalman filtering for spacecraft attitude estimation. *J Guid Control Dyn* 1982;5(5):417–429.
3. Eberle C, Ament C. The unscented Kalman filter estimates the plasma insulin from glucose measurement. *Biosystems* 2011;103(1):67–72.
4. Lavieri MS, Puterman ML, Tyldesley S, Morris WJ. When to treat prostate cancer patients based on their PSA dynamics. *IIE Trans Healthc Syst Eng* 2012;2(1):62–77.
5. Wang Q, Molenaar P, Harsh S, et al. Personalized state-space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose, and meal Intake. *J Diabetes Sci Technol* 2014;8(2):331–345.
6. Catlin DE. The Discrete Kalman Filter. In: Estimation, Control, and the Discrete Kalman Filter. 71st ed. New York, NY: Springer Science & Business Media; 2012:133–163.
7. Schell GJ, Lavieri MS, Helm JE, et al. Using filtered forecasting techniques to determine personalized monitoring schedules for patients with open-angle glaucoma. *Ophthalmology* 2014;121(8):1539–1546.
8. The AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS): 1. Study design and methods and baseline characteristics of study patients. *Control Clin Trials* 1994; 15(4):299–325.
9. Musch DC, Lichter PR, Guire KE, Standardi CL, the CIGTS Study Group. The Collaborative Initial Glaucoma Treatment Study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology* 1999;106(4):653–662.
10. Musch DC, Gillespie BW, Lichter PR, Niziol LM, Janz NK. Visual field progression in the Collaborative Initial Glaucoma Treatment Study. The impact of treatment and other baseline factors. *Ophthalmology* 2009;116(2):200–207.e1.
11. The Japan Glaucoma Society Guidelines for Glaucoma, 3rd ed [in Japanese]. *Nihon Ganka Gakkai Zasshi* 2012;116(1):3–46.
12. European Glaucoma Society Terminology and Guidelines for Glaucoma, 4th ed - Chapter 2: Classification and terminology. Supported by the EGS Foundation: Part 1: Foreword; Introduction; Glossary; Chapter 2 Classification and Terminology. *Br J Ophthalmol* 2017;101:73–127.
13. Hodapp E, Parrish RK, Anderson DR. Clinical Decisions in Glaucoma. London, UK: Mosby Inc; 1993.
14. Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. *Invest Ophthalmol Vis Sci* 2002;43(5):1400–1407.
15. Bengtsson B, Patella VM, Heijl A. Prediction of glaucomatous visual field loss by extrapolation of linear trends. *Arch Ophthalmol* 2009;127(12):1610–1614.
16. Heijl A, Bengtsson B, Hyman L, Leske MC. Natural history of open-angle glaucoma. *Ophthalmology* 2009;116(12): 2271–2276.
17. Anderson DR, Drance SM, Schulzer M, Collaborative Normal-Tension Glaucoma Study Group. Natural history of normal-tension glaucoma. *Ophthalmology* 2001;108(2):247–253.
18. Nitta K, Wajima R, Tachibana G, et al. Prediction of visual field progression in patients with primary open-angle glaucoma, mainly including normal tension glaucoma. *Sci Rep* 2017;7(1):15048.
19. Sung MS, Kang YS, Heo H, Park SW. Optic disc rotation as a clue for predicting visual field progression in myopic normal-tension glaucoma. *Ophthalmology* 2016;123(7): 1484–1493.
20. Lee J, Kong M, Kim J, Kee C. Comparison of visual field progression between relatively low and high intraocular pressure groups in normal tension glaucoma patients. *J Glaucoma* 2014;23(8):553–560.
21. Ahrlich KG, De Moraes CG, Teng CC, et al. Visual field progression differences between normal-tension and exfoliative high-tension glaucoma. *Invest Ophthalmol Vis Sci* 2010; 51(3):1458–1463.
22. Drance S, Anderson DR, Schulzer M. Risk factors for progression of visual field abnormalities in normal-tension glaucoma. *Am J Ophthalmol* 2001;131(6):699–708.
23. McNaught AI, Hitchings RA, Crabb DP, Fitzke FW. Modelling series of visual fields to detect progression in normal-tension glaucoma. *Graefes Arch Clin Exp Ophthalmol* 1995; 233(12):750–755.
24. Krajewski L, Ritzman L, Malhotra M. Forecasting demand. In: Operations Management: Processes and Supply Chains. 10th ed. Boston: Pearson; 2013:490–500.
25. Hastie T, Tibshirani R, Friedman J. Model assessment and selection. In: The Elements of Statistical Learning. 2nd ed. New York, NY: Springer; 2009:241–249.
26. Bhapkar VP. A note on the equivalence of two test criteria for hypotheses in categorical data. *J Am Stat Assoc* 1966;61(313): 228–235.
27. Thonginnetra O, Greenstein VC, Chu D, Liebmann JM, Ritch R, Hood DC. Normal versus high tension glaucoma. *J Glaucoma* 2010;19(3):151–157.
28. Sommer A. Ocular hypertension and normal-tension glaucoma. *Arch Ophthalmol* 2011;129(6):785.
29. Shields MB. Normal-tension glaucoma: is it different from primary open-angle glaucoma? *Curr Opin Ophthalmol* 2008; 19(2):85–88.
30. Shields MB, Spaeth GL. The glaucomatous process and the evolving definition of glaucoma. *J Glaucoma* 2012;21(3): 141–143.
31. Mroczkowska S, Benavente-Perez A, Negi A, Sung V, Patel SR, Gherghel D. Primary open-angle glaucoma vs normal-tension glaucoma. *JAMA Ophthalmol* 2013; 131(1):36.
32. Kazemian P, Lavieri MS, Van Oyen MP, Andrews C, Stein JD. Personalized prediction of glaucoma progression under different target intraocular pressure levels using filtered forecasting methods. *Ophthalmology* 2017;125(4): 569–577.