Distribution Agreement

Using Lagged Outpatient Visits to Improve Forecasts of

Patient Arrivals at an Inpatient Hospital

By

Gabriel R. Vece

Mater of Science in Public Health

Biostatistics and Bioinformatics

Howard Chang, PhD

(Thesis Advisor)

Eugene Huang, PhD

(Reader)

Using Lagged Outpatient Visits to Improve Forecasts of

Patient Arrivals at an Inpatient Hospital

By

Gabriel R. Vece

B.S

Kennesaw State University

2008

Thesis Advisor: Howard Chang, PhD

Reader: Eugene Huang, PhD

An abstract of a Thesis

Submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health in Biostatistics

2017

Abstract

Using Lagged Outpatient Visits to Improve Forecasts of

Patient Arrivals at an Inpatient Hospital

By: Gabriel R. Vece

Introduction: Recent increases in healthcare expenditures have incentivized hospitals to reduce labor costs without adversely affecting patient outcomes, requiring administrators to forecast when and where patients will arrive for care. Many forecasting approaches involve subjective judgement and, among those that employ statistical methods, finding sources of data that can help predict trends in arrivals presents a significant difficulty. Meanwhile, the "gatekeeper" structure of many large healthcare systems is such that patients are incentivized to visit outpatient clinics, specialists, or primary care providers before coming to hospital, which may imply that surges in hospital visits could be preceded by similar surges in outpatient visits. The goal of this paper will be to investigate the improvement, if any, gained by the inclusion of lagged outpatient visits in a model for forecasting daily inpatient arrivals, modeling the outpatient visits as distributed-lag predictors in a Poisson regression model.

Methods: Several canonical time-series models were fit to the data to establish the performance some common forecasting models. Afterwards, series of outpatient visits to neurologists and to neurosurgeons were modelled as distributed-polynomials using the Aikake Information Criteria (AIC) to select the optimal values of several key parameters, such as lag length and polynomial degree. The polynomials were included in a Poisson regression for forecasting inpatient arrivals, with the model performance assessed by its root mean square error (RMSE) and mean absolute error (MAE) compared to the canonical time series model and a Poisson model excluding the distributed-lag terms.

Results: Although the Poisson model including the distributed-lag terms failed to outperform the naïve or univariate time-series models at making short-term (7-day) forecasts, it achieved better performance with a longer forecast window (30 days). However, this improvement was also seen in the Poisson model excluding the lagged covariates, suggesting that the outpatient series contributed little, if any, added predictive power. Furthermore, a sensitivity analysis showed that the improvements did not hold when the models were fitted at various seasonal subsets of the dataset.

Conclusion: Although the data used in this study constitute only the patterns observed at a particular hospital system at a particular point in time, the results suggest that the outpatient series were unable to significantly improve the model's forecasts. In practice, forecasters may benefit from the use of other multivariate modeling approaches or from more thorough searches for useful predictors.

Using Lagged Outpatient Visits to Improve Forecasts of
Patient Arrivals at an Inpatient Hospital

By

Gabriel R. Vece

B.S
Kennesaw State University
2008

Thesis Advisor: Howard Chang, PhD
Reader: Eugene Huang, PhD

A Thesis
Submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health in Biostatistics
2017

Acknowledgements

I would like to thank Dr. Howard Chang for his invaluable support throughout this project. He must have seen something in this project which I—at the time—could not, as he volunteered his time and expertise to help me think and work through all of the obstacles I faced during the writing of this thesis. This project would not have been possible without Dr. Robert Palmer and Nikolay Braykov, who provided me with the resources and knowledge I needed to learn how to work with the (often very alien) claims data used herein. I would like to thank Dr. Eugene Huang, who was actually the first person whose advice I sought concerning this thesis project and who volunteered to become my reader, offering insightful commentary and advice that has been immensely beneficial. I would also like to thank the faculty of the Department of Biostatistics and Bioinformatics at the Rollins School of Public Health, who have shared with me their collective knowledge, support, and guidance during the last two years.

Finally, I would like to thank my friends and family for all the times they've been there, with a special acknowledgement to Kirstie and Deauxbac.

To Deauxbac: "ma-goot."

To Kirstie: I love you so much and you have made all of this possible; I hope you will always know how much that means to me.

Table of Contents

1. Introduction

Growth in U.S healthcare systems has led to dramatic and unprecedented changes in the ways that health services are distributed, accessed, and consumed by patients. A report by the Kaiser Family Foundation (2012) concluded that national healthcare expenditures rose from $4878 per person (overall, 13.8% of gross domestic product[[GDP]) to $8402 per person (17.9% of GDP) from 2000 to 2010, with costs expected to continue rising in subsequent years. Furthermore, this growth rate is accelerating, and is expected to exceed the rate of GDP growth, becoming increasingly unsustainable (Follette & Sheiner, 2005). Of these costs, inpatient care expenses represent the largest proportion, accounting for 29.3% of expenditures in 2009 (Kashihara & Carper, 2012); as a result, many strategies that are targeted at reducing the burden of healthcare costs wisely focus on inpatient expenses as the primary driver of overall costs. However, inpatient care settings themselves are complex facilities, often serving a large and diverse region or pool of patients through a similarly large network of professionals and resources; as a result, addressing the largest drivers of costs within inpatient facilities is often a nontrivial endeavor.

Within the inpatient care setting, many solutions to reduce costs are focused on either directly or indirectly control labor costs. In 2016, labor costs represented the single largest hospital expense, accounting for nearly half of total operating costs (Glied, Ma, & Solis-Roman, 2016). However, simply cutting staff or reducing hours worked has been shown to have detrimental effects on patient outcomes, so more sophisticated solutions are necessary. For example Weiss, Yakusheva, and Bobay (2011) showed that lower nurse staffing rates significantly reduced patients' likelihoods of being readmitted or having an emergency department admission within 30 days of discharge.

For this reason, resource planners must know when it will be necessary to have a fully-staffed hospital floor and when fewer resources will be sufficient. As a result, knowing when—and where—to expect peak volumes is very valuable. In many cases, demand forecasts such as these are informal in nature, constituting a "best guess" by an experienced manager or resource planner. In many other cases—particularly in emergency departments and large hospital institutions—statistical models spanning a wide variety of modelling approaches have been used to varying degrees of success (Wiler, Griffey & Olsen, 2011).

1.1. Problem Statement

An implicit assumption in many common statistical models using linear regression to predict hospital inpatient arrivals is that, after accounting for one or more known effects, residual variation represents an independent random process. In many hospitals, patients may arrive for inpatient care only after first visiting a specialist or primary care physician at an outpatient setting, as part of a sequence of episodes of care (EOC), such as a series of follow-up visits after diagnosing a serious disorder. Under ideal circumstances, these visits are scheduled in advance and are therefore available to resource planners. However, identifying and locating all the pertinent and useful sources of information is often difficult logistically and, when this information is not available, incorporating this temporal sequence of patient encounters can be very difficult. Hence, in practice, planners typically use simpler, naïve modeling approaches that fail to consider “upstream” patient encounters, leading to potentially biased and uncertain forecasts.

1.2. Purpose Statement

The purpose of this investigation is twofold: 1) to examine and compare the performance of a number of univariate models for forecasting daily inpatient volumes at a Children’s Hospital in Atlanta, Georgia (henceforth known as “Test Hospital”), and 2) to evaluate the utility of lagged outpatient arrivals in terms of forecasting inpatient arrivals. Primarily, the goal of this investigation is to evaluate the improvement, if any, achieved by incorporating historical outpatient arrival data as predictors in models for forecasting inpatient arrival volumes. We hypothesize that outpatient centers, by serving overlapping patient populations and by acting as “upstream reservoirs” for patients, will be able to significantly improve forecasters’ abilities to anticipate inpatient arrivals.

1.3. Significance Statement

Firstly, this paper critically evaluates potentially useful sources of information that can be readily incorporated into hospital staffing models, improving forecast accuracy and reducing costs. Secondly, as hospital systems grow, there is an increasing need to make use of the large amounts of data available, and the analytic procedures presented here could suggest innovative ways to integrate separate sources of healthcare data.

2. Background/Literature Review

2.1. Staffing Models

The period 2000–2010 saw unprecedented increases in the costs of healthcare that have led to much reexamination of the ways in which healthcare is offered and delivered in order to emplace more efficient systems. In response to increasing inpatient costs, many hospitals are incentivized to do so through "value-based" reimbursement structures that encourage the elimination of wasteful spending and the provision of evidence-based and effective care. Since labor costs constitute the largest single contributor to hospital costs[4], it is naturally one of the first expenses to which managers turn when considering strategies for reducing costs. However, since unilaterally reducing the number of medical staff (such as nurses) has negative consequences for patient outcomes (Weiss, Yakusheva & Bobay, 2011), managers must respond by anticipating needs on a shift-by-shift basis and allocate their staff accordingly.

In practice, forecasts for patient volumes within hospital units are frequently made based on expert testimony from experienced resource managers, since building statistical models can present difficulties associated with obtaining (and cleaning) the required data and with having the expertise needed to build such a model. In addition, the data sometimes require specialized approaches for dealing with time-series data that managers and planners may not be trained to use. For example, day-to-day patient volumes tend to exhibit seasonal trends (particularly in emergency departments; Asplin, Flottemesch, & Gordon, 2006) that pertain to certain department-specific trends in care, such as high volumes occurring in the early months of each year due to flu. Ensuring that a model can capture such a trend requires a combination of both clinical and statistical expertise.

Furthermore, serial observations with trends often leads to autocorrelation that violates the basic assumptions regression modeling, and diagnosing the severity of these violations to recommend corrective measures is needed. Finally, the processes that govern patients' movements into and out of the healthcare system are often difficult to forecast, even when there is a great deal of current information present.

Much healthcare data being collected—which frequently tracks patient-level information—may inform about the "trajectory" of the current patient arrivals (presenting symptoms, prognoses, etc.) but does little to predict what a completely different set of patients might present with on a different day, since individual clinical data do not necessarily reflect the population-level trends that govern when and where patients move from one episode of care to another. For these reasons time series analysis techniques, which estimate trends over time based on historical data (that are easily extrapolated into forecasts) have become a widely used tool for forming predictive models that can forecast patient volumes. However, while time-series models can easily handle autocorrelated observations, model interpretation and the incorporation of external information can often be more difficult (Wargon 2009).

To date, much has been done to describe methods used to form patient volume forecasts using both regression and time-series methods. Jones et al. (2008) performed a detailed review of several univariate (using only historical data) time-series approaches to modelling patient volumes, including neural networks, exponential smoothing, autoregressive integrated moving average (ARIMA; Box, Jenkins, & Reinsel, 2008) models, and several regression methods using calendar-based binary covariates (e.g., holidays, seasons, etc.). Despite the added complexity of the time-series methods, only marginal improvements in out-of-sample forecasts were achieved compared to the regression approaches; however since the data were from only three different hospital facilities, these results may not generalize well. Kam, Sung and Park (2010) utilized a similar approach of time-series models with calendar time as covariates but also demonstrated a multivariate ARIMA model incorporating weather patterns (rain, snow, temperature, relative humidity, and air velocity) which outperformed the univariate time-series models in terms of out-of-sample performance based on mean average percent error (MAPE). Similar efforts have tended to arrive at similar conclusions, and these investigations have collectively identified univariate time-series methods as powerful alternatives for forecasting models, although models with multiple covariates can be promising if the added complexity is acceptable and justifiable (Jones et al., 2009; Bergs, Heerinckx & Verelst, 2014; Sun, Heng, Seow & Seow, 2009).

One limitation of the most common time-series approaches (e.g., ARIMA, Holt-Winters Smoothing) is that they exclude information related to potential predictors of patient volumes. Although resource planners may be enabled to better anticipate volume patterns through forecasts from univariate models, having no awareness of the relevant predictors of these patterns means that they have no basis by which appropriate interventions may be constructed to modify the patterns. Since being able to modify patterns in patient volumes is potentially quite valuable, forecasters hoping to better anticipate patient trends have a strong incentive to turn to models that permit the inclusion of possible covariates. For example, Kam, Sung and Park’s (2010) model using weather variables represents a successful attempt to identify predictors that can improve forecasting models; however, even in this case, the weather is a similarly uncontrollable process and does not present a solution for resource planners to influence the trends affecting their departments.

One area of study which has appeared to hold some promise is related to the flow of patients through the healthcare system. It is well-known that networks of healthcare providers are intentionally designed to afford some level of control over the way in which individuals access and navigate hospitals, specialists, and clinics (e.g., “gatekeeping”; Forrest, 2003). Furthermore, research has increasingly demonstrated that these movements are, in many cases, predictable and useful with respect to resource planning. Broyles, Cochran, and Montgomery (2010) used a discrete-time Markov Chain approach to describe patient influx and exodus as arrival and service rates, using maximum likelihood to estimate transition probabilities between different inpatient units at an Arizona hospital. This procedure allowed the analytic derivation of the probability distribution of unit volumes at an hourly level, permitting resources planners to determine the probability of being understaffed given a particular number of onsite staff. Additionally, this model was found to have less predictive variance than the more typical seasonal ARIMA models. In a similar effort, Littig and Isken (2007) build a predictive occupancy model using multinomial logistic regression to estimate the probability of patient movements from unit-to-unit (or unit-to-exit) while also incorporating arrival processes such as scheduled surgeries and randomly-modelled emergency department arrivals with similar success.

These approaches to modelling patient movements, however, depend on discrete perspectives of time which are often very small (such as shift-by-shift or hourly) in order to capture individual movements, which may be less useful for forecasting further into the future, such as predictions for a subsequent week or month. At this level, patient pools increasingly become the unit of concern for which one can visualize, on a particular day, a surge in outpatient appointments among which only a few may require follow-up care at an inpatient unit. From a forecasting point of view, it may be possible to build models that incorporate this temporal sequence of patient encounters that improve our ability to anticipate inpatient visits. In this framework one might, for example, infer that recent outpatient activity would allow one to anticipate changes in the number of future inpatient arrivals, constituted by the patients whose needs required resources only available at an inpatient unit. For this present study, I propose that recent outpatient activity can have usable predictive power with respect to forecasting future inpatient arrivals, and that models incorporating this outpatient information will have improved forecast accuracy compared to similar models that do not.

2.2. Distributed Lag Models

In modeling situations where autocorrelation is suspected to be present, it is a common practice to include lagged values of one or more predictors as covariates in a linear regression; however, a number of methodological issues must be considered. Primarily, the presence of autocorrelation among values of the outcome variable implies that a fundamental assumption of regression (independence of observations) has been violated and that, as a result, the variances of the estimated parameters will be incorrect. In addressing this problem, one must specify a particular lag structure—namely, which specific lags to include in the model— with the understanding that this proposed structure may underspecify or overspecify the true structure. Generally, models that include too few lags suffer from bias due to *omitted variable bias* (Clarke 2005) whereas models that include too many lags often can easily encounter issues of multicollinearity that can lead to inflated variances on the parameters.

This second issue of multicollinearity among the predictors is especially problematic when serial autocorrelation is present: for a given association between some value observed at time t and another value observed at time $t - k$, nontrivial associations with other nearby values (say, at time $t - k - 1$ or $t - k + 1$) are also likely to be present, which encourages the inclusion of additional lagged covariates in the model. As a result, a large set of candidate lags may be suspected for inclusion into the model, with little hypothetical foundation by which to judge their relevance. One approach to avoid these issues involves the selection of a lag range and a mathematical structure by which to model the lag-response relationship. Constraining the lag-response relationship to some particular form—such as a polynomial function of the lag—allows one to flexibly model a complex dynamic relationship using a much smaller number of model terms.

Distributed lag (DL) models are a class of statistical model that capture this type of behavior. Within the context of time-series observations DL models permit the modeling of the effect of a predictor variable—observed over a period of time—that exerts a dynamic influence on the current and subsequent values of an outcome variable, which is observed over the same period of time. These models have been used in a variety of settings, such as for estimating the longitudinal effects of prolonged uranium exposure on mortality (Gasparrini 2014) and the effects of socioeconomic factors on productivity in the long and short run (Wang et al., 2016).

Typically, these models have been a mainstay of econometric investigations but have been shown to be of similar use in modelling relationships with hospital admissions:

For example, Shrestha (2007) used a DL model to demonstrate the time-dependent relationship between air pollution (PM_{10}) and subsequent hospital admissions. The value of DL models in this present investigation lies in the fact that, when an outcome series can be shown to be associated with distant lagged values, this relationship may be intuitively utilized to make a forecasting equation based on the lagged values. In this way, DL models represent a flexible and convenient framework with which external data can be included in a more traditional regression model while also permitting the creation of forecasts.

This present investigation proposes the use of DL models to capture the relationship between outpatient specialist arrivals and inpatient admissions at the Neurosciences department of Test Hospital, a large pediatric hospital that is part of a large network incorporating dozens of outpatient specialist clinics. The primary focus will be assessing the predictive power obtainable by including lagged specialist visits into forecasting models, to address the question of how forecasts may be impacted by explicitly building models that consider patient arrivals at outpatient clinics. In doing so, the forecasts from univariate time series models, regression models, and multivariate models with and without the inclusion of the outpatient data will be evaluated.

3. Methodology

3.1. Dataset

The data used for this present study were obtained from hospital account record within a large network of hospitals and clinics during the period from January 6, 2013 to

January 31, 2016; included in the network was one major pediatric hospital (Test Hospital) and 100 different pediatric outpatient sites, which spanned 40 different unique provider specialties. When this project was initiated, our goal was to use the outpatient information to make patient arrival forecasts at only one department (Neurosciences) of Test Hospital, so I decided to consider only outpatient visits to a related specialist: a neurologist or a neurosurgeon. For each specialist, patient visits were aggregated across locations into daily counts to produce two separate time series of daily arrivals (those to a neurologist and those to a neurosurgeon) in addition to the time series of daily inpatient arrivals at the neurosciences department of Test Hospital. With respect to forecasting, models were trained on the data prior to 2016, reserving the first 30 days (approximately the month of January) of 2016 as a test set from which to evaluate the forecasts. During the period covered by the training data, 2,796 inpatient encounters among 2,277 different patients were recorded in addition to 35,332 visits to a neurologist (among 6,821 patients) and 26,076 visits to a neurosurgeon (among 4,152 patients).

One notable feature of the data is that, while the inpatient department received patients on every day of the week, all of the neurological and neurosurgical outpatient clinics in the dataset operated on a 5-day work week, only receiving patients from Monday through Friday. As a result, observations pertaining to Saturdays and Sundays always had exactly zero arrivals throughout the study period. This meant that, in terms of weekly arrivals, the inpatient arrivals had a period of seven days while the outpatient arrivals had a period of only five days. To adjust for this offset, outpatient observations from Saturdays and Sundays were imputed using a simple 7-day moving average with exponentially increasing (with further lag) weights.

3.2. Statistical Approach

The primary goal of this present study was to evaluate the predictive content of the two outpatient arrival series with respect to making forecasts of inpatient arrivals. Initially, several univariate and naïve models common to the time-series literature were used to make forecasts, using ARIMA, exponential smoothing, autoregressive neural networks, and a Poisson generalized linear regression having day-of-the-week and holiday covariates. Naïve models such as random walk, seasonal naïve, and baseline mean forecasts were also used to establish the relative performance of more complex models. Finally, a Poisson model incorporating the same day-of-week and holiday variables was modified to include distributed lag terms to capture the relationships between the inpatient arrivals and lagged outpatient visits to neurologists or neurosurgeons. This structure of the two distributed lag terms was chosen based on the model having the minimum AIC among all other possible candidate models. Ultimately, the comparison of the forecast accuracy for short (7-day) and long (30-day) forecasts between this model and the other models were evaluated based on the root mean squared error (RMSE) and mean absolute error (MAE) of the forecasted values and the actual values.

One important feature of DL models is their straightforward incorporation into a generalized regression model. A simple polynomial DL model of order *q* over the lag range 1 to *n*, where *t* denotes the time index, has the form:

$$E(y_t) = b_0 + \sum_{k=1}^{n} b_k x_{t-k} + \epsilon_t$$

$$b_k = k^0\alpha_0 + k\alpha_1 + k^2\alpha_2 \cdots + k^q\alpha_q$$

In the above model, only the weights b_0, α_0, α_1, … α_q need to be estimated. This model has the equivalent representation:

$$E(y_t) = b_0 + \alpha_0 \sum_{k=0}^{n} x_{t-k} + \alpha_1 \sum_{k=0}^{n} k x_{t-k} + \alpha_2 \sum_{k=1}^{n} k^2 x_{t-k} \cdots + \alpha_q \sum_{k=0}^{n} k^q x_{t-k} + \epsilon_t$$

from which it becomes apparent that distributed-lag terms can be added to a regression model as a simple linear combination of terms. Note that, in this model, the effect at lag zero is ignored (it represents a same-day effect and is therefore not useful in a forecasting context). In addition, it is important to note that, for a *q*-degree polynomial lag-response structure, only $q + 1$ coefficients need to be estimated, allowing even dozens of lag terms to be used with relatively flexible restrictions on their structure.

For both the neurologist arrivals and the neurosurgeon arrivals, the maximum lag length of the range to be considered was fixed at 30 days. In other words, I chose to model the relationship between the inpatient arrivals and the history of outpatient arrivals up to and including observations from 30 days prior, with the minimum lag allowed to vary in order to create a set of competing candidate models. Ultimately, the primary goal of making accurate forecasts led to this decision, since the smallest lag in the chosen window would determine when and how the forecasts could be used. For example, for a model forecasting series *Y* based on the 10-, 11-, and 12-day lagged observations of some exogenous series *X*, these forecasts could be available no sooner than 10 days prior to the date being forecasted since the 10-day lagged observation of *X* would simply not be available before then. Therefore, it became of practical benefit to focus on models that emphasized the effect of historically older data and the model selection process employed considered the *maximum* lag to be fixed in order to use the older data as much as possible. Meanwhile, the *minimum* lag and polynomial were selected to achieve a minimal AIC, arriving at a final model. Once the number of lags to be included in the distributed-lag terms were chosen, the terms were added to a model including day-of-the-week and holiday factors and were estimated in R (version 3.3.2; R Core Team, 2016).

Finally, since the available data pertains only to a particular hospital system at a particular time, I sought to test the robustness of the multivariate model by refitting it to various time points. For example, since the models would be trained on data spanning a nearly 3-year period and evaluated on relatively short 7- and 30-day forecasts, retraining the data on less data or at different points may avoid bias due to the model estimates being trained over long-term cycles that are irrelevant with respect to the forecasts. For this reason, both Poisson models (including and excluding the outpatient data) were refitted to all possible annual and quarterly windows within the available data period and the results were averaged to imply a mean level of performance. To illustrate this process in the case of using an annual window, both models were trained on data spanning a 365-day period that began on a different dates (such as 1/1/2015–1/1/2016, or 2/15/2013–2/15/2014), with the models being compared on the basis of 7- and 30-day forecasts extending just beyond the training data. This sensitivity analysis was intended to demonstrate how the model performed over different time periods and on different amounts of data allow more general statements about the relative performance of the models to be made.

4. Results

4.1. Exploratory Data Analysis

Before attempting to fit any models, characterizing the three time series being investigated was necessary, as these properties could inform the model-fitting process and suggest the structure of the time-varying relationships between the three series. Due to the nature of the time series data being investigated, some serial dependence was expected from the observations; In particular, dealing with time series data in a regression context puts the model at risk of identifying spurious relationships unless certain precautions are taken (Granger & Newbold, 1974). Initially, Kwiatkowski-Philips-Schmidt-Shin (KPSS; 1992) tests were used to check the stationarity of the data.

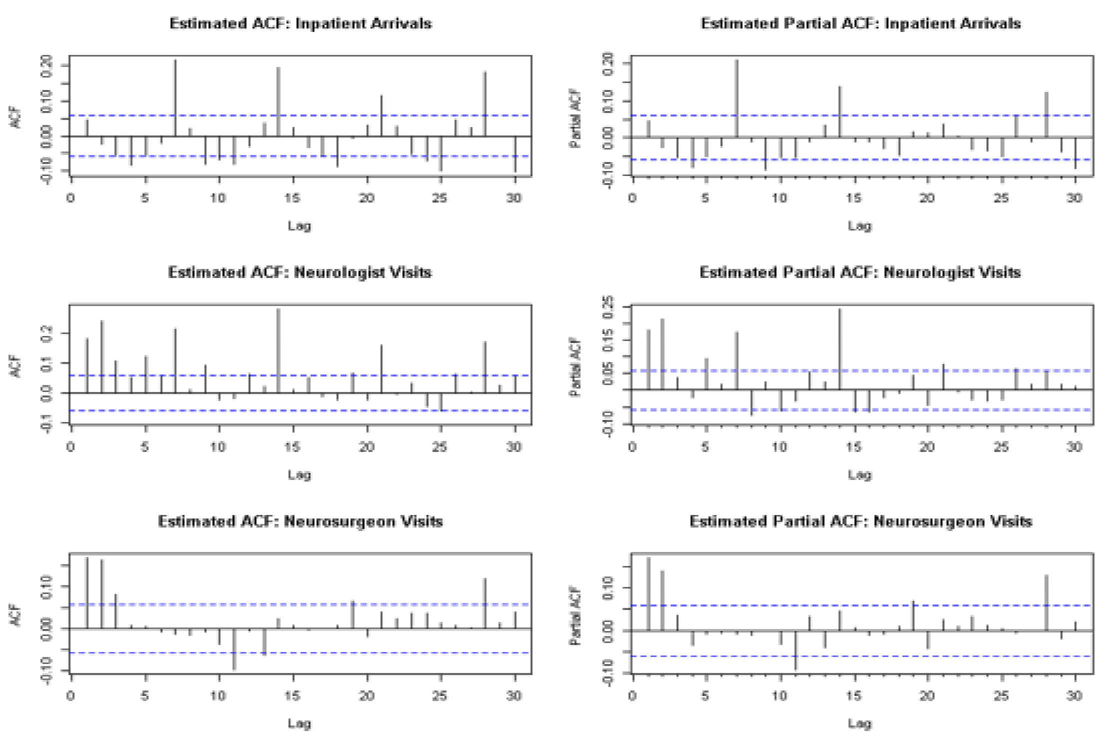

Figure 1. Autocorrelation plots for the three series of inpatient arrivals, neurologist visits, and neurosurgeon visits. Of note are the frequent significant associations among lags at multiples of seven, implying weekly periodicity (seasonality).

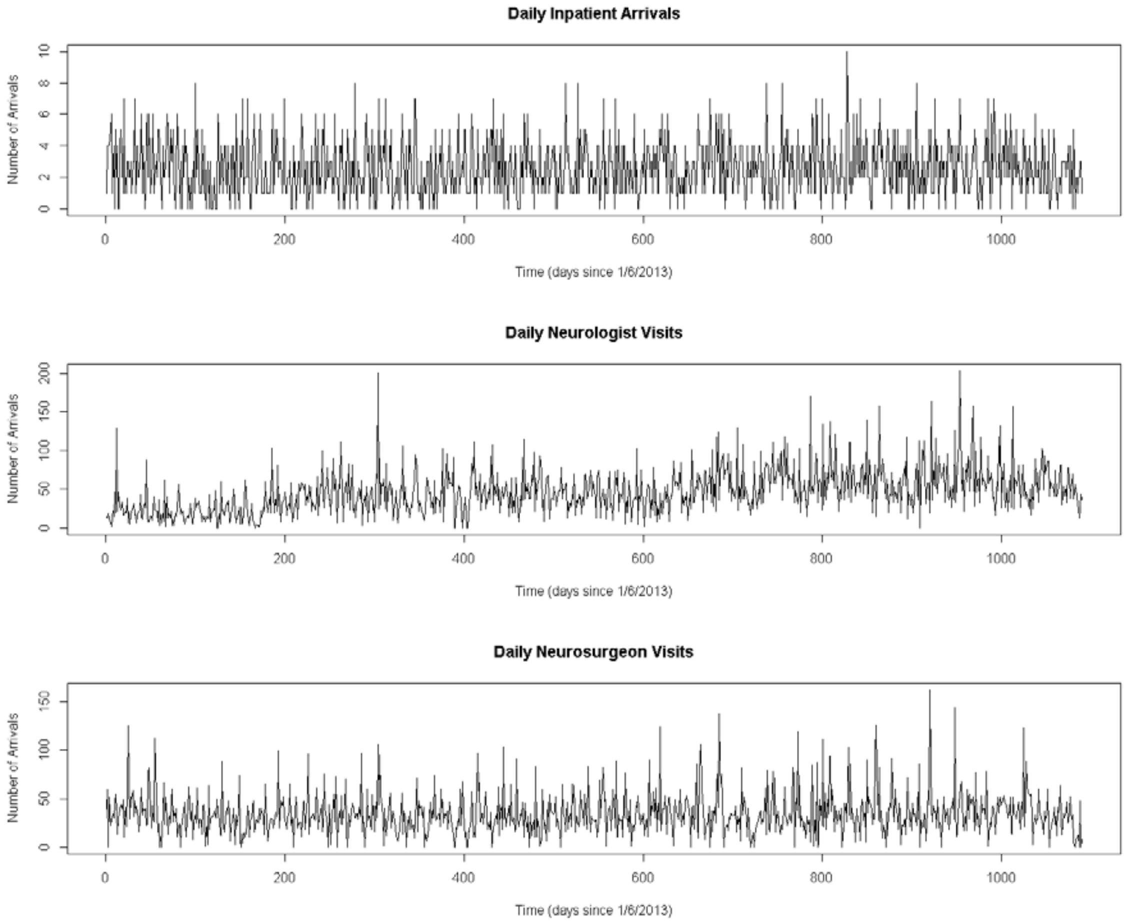

Figure 2. Time plots of the daily inpatient arrivals at Test Hospital, neurologist clinics, and neurosurgeon clinics from January 2013 to December 2015.

Both the inpatient arrivals and the neurosurgeon arrivals were found to be stationary over the study period and thusly remained untransformed, entering the model in their original state. However, the series of neurologist arrivals showed an increasing trend over time (and therefore failed to be stationary), so subsequent model fitting included the observation time as a predictor in order to account for the trend. Subsequent KPSS indicated that the detrended series showed no significant evidence of nonstationarity. In examining the autocorrelation structure within and between the three series, weekly patterns were apparent with each series showing significant and persistent autocorrelations repeating every 7th day. This appeared to be

the major source of autocorrelation in the inpatient arrivals series; in the outpatient arrivals, some significant autocorrelations were observed for very small lags (i.e., very recent observations at lags 1 and/or 2), although this relationship quickly decayed after two to three lags. In general, far fewer inpatient arrivals were observed over the data period (median = 2 per day, range = [0, 10]) compared to outpatient arrivals (neurologists: median = 26 per day, range = [0, 203]; neurosurgeons: median = 21 per day, range = [0, 162]) since the outpatient arrivals were collected and aggregated over many different clinics (11 different neurological clinics and 6 different neurosurgical clinics).

4.2. Univariate Models

In order to determine the added value of incorporating the outpatient time series data, the effectiveness of several univariate models of increasing complexity was first evaluated. The simplest (naïve) models generally involve no covariates, instead merely extending recent information into the forecasting period. For example, the seasonal naïve model simply assumes that the next *k* forecasted values will precisely equal the values observed during the most recent season (the previous week, for this study), repeating the most recent observations, if necessary. Slightly more complex, the seasonally adjusted random walk first estimates the seasonal cycle (through LOESS regression) and then removes it from the data, yielding a deseasonalized series of observations which is forecasted using

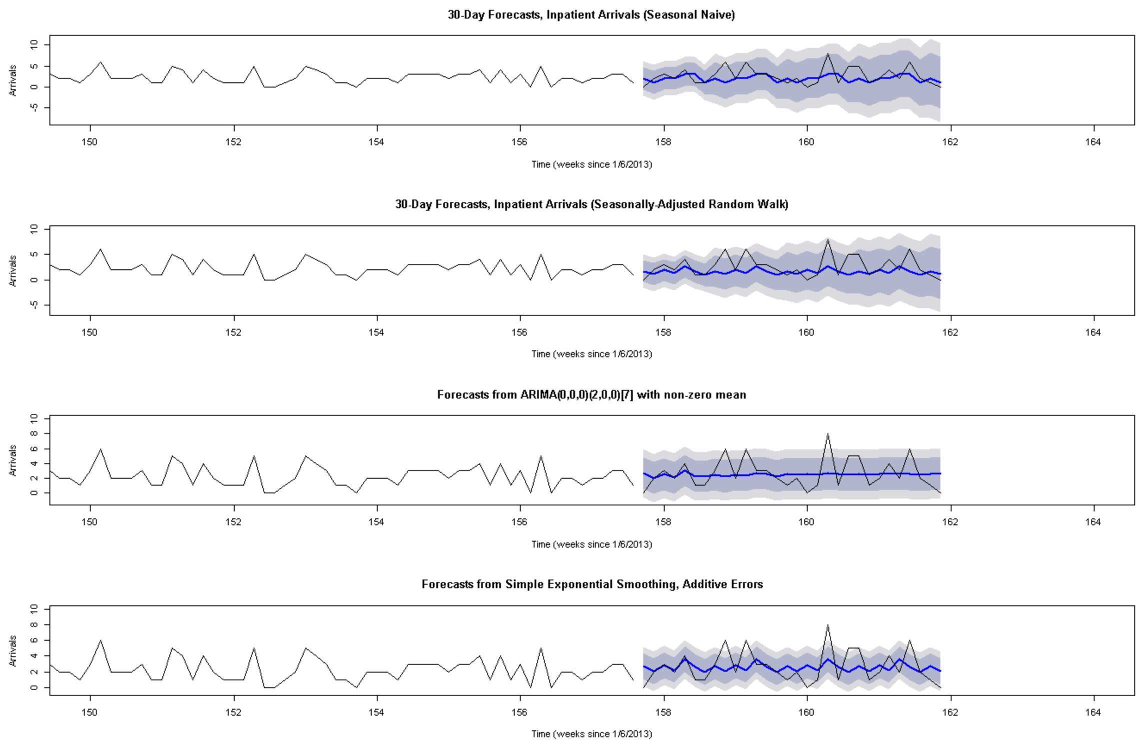

Figure 3. Forecasts from two naïve models (seasonal naïve and seasonal random walk) and two optimal time-series models (ARIMA and exponential smoothing). Forecasts (blue, with 95% and 80% prediction intervals) begin just prior to week 158 (1/1/2016) are provided to demonstrate common approaches, and the accuracy achievable (compared to actuals, in black) even without incorporating covariates into the model.

a random walk from the most recent observation before adding the seasonal cycle back into the forecasted values. Plots of the forecasts from these models are presented in Figure 2. Afterwards, more sophisticated canonical time series models (exponential smoothing, SARIMA, and neural network) were fitted using the *forecast* package available in R (version 8.0; Hyndman, 2017), which allows for convenient selection of model tuning parameters and assumptions based on having the minimum AIC among other possibilities. In general, the simpler models performed well where short-term (7-

day) forecasts were concerned, but were outperformed by the more complex models at 30-day forecasts, although this is not true in all cases. One important feature of the models is that the naïve models have the property of extending static trends outward indefinitely, while the more complex models do not. For example, a stationary ARIMA model can be shown to eventually converge to the mean of the series (which can be observed in Figure 3); this implies that, although the dependence structure of the series means that recent observations will resemble future observations to some degree, this relationship decays into the long term and is less informative for forecasting.

4.3. Multivariate Model

Model selection was carried out over all models in which the two separate outpatient series, indexed by *j*, were represented as distributed-lag polynomials of degree k_j over the minimum lag n_j through 30, in which the values of k_j (ranging from 1 to 20) and n_j (ranging from 1 to 30) were chosen as the values which minimized the model AIC. After testing all the possible models, the neurologist arrivals and the neurosurgeon arrivals were modelled separate cubic and quartic polynomials over lags 10–30 and lags 4–30, respectively. The model equation can be written as:

Model 1:

$$log\left[E(y_t)\right] = \beta_0 + \sum_{k=10}^{30} a_k X_{1,t-k} + \sum_{p=4}^{30} b_p X_{2,t-p} + \beta_1(DOW) + \beta_2(Holiday)$$

In Model 1, changes in the outpatient series (X_1 = neurologist, X_2 = neurosurgeon) are allowed to have a dynamic relationship with the inpatient arrivals over the defined lags, but the estimated magnitudes of this relationship were typically small (a summary of the model's fitted estimated parameters is provided in Table 1). An important assumption of Model 1 is that, after accounting for the covariates, the inpatient visits on different days were considered independent; a residual plot confirmed that accounting for day-of the week and lagged effects removed the autocorrelation observed in the original series (Figure 4) with only a few slight correlations remaining.

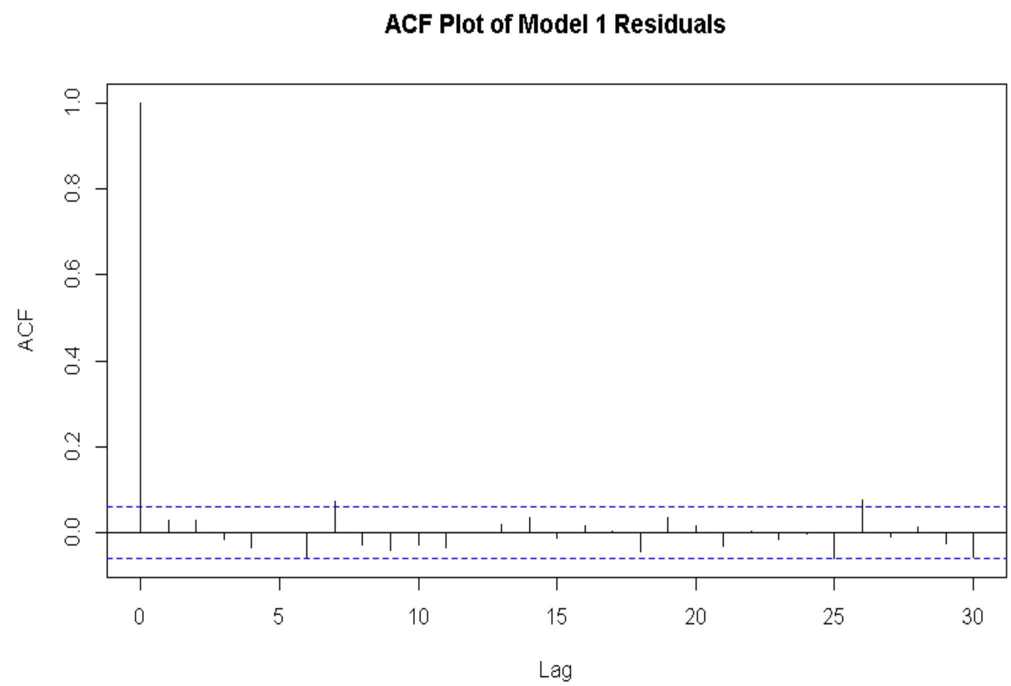

Figure 4. Autocorrelation plot of residuals of model 1 (accounting for holiday, day-of-the-week, and lagged outpatient visits).

In interpreting the distributed lag terms, it is important to note that these terms are defined over the predictor-lag-response space, for which it is perhaps easiest to consider the lag-response dimension given some fixed change in the value of the predictor. For example, the effect of a 10-unit change in daily neurologist arrivals was significant only at short (e.g., 10 days prior) and long (e.g., 30 days prior) lags. On the other hand , an equivalent unit change in the neurosurgeon arrivals implied significant positive effects at a range of early lags (approximately 7-13 days) and again, briefly, at lag 24 (risk ratio = 1.01 [Wald 95% CI: 1.0008–1.0123]), with negative associations again occurring at short and long lags.

Using this model to build 7- and 30-day forecasts, the model's forecasting performance was typically poorer than the naïve or more complex univariate time series models in the short (7-day) term but more accurate as the forecast window became longer

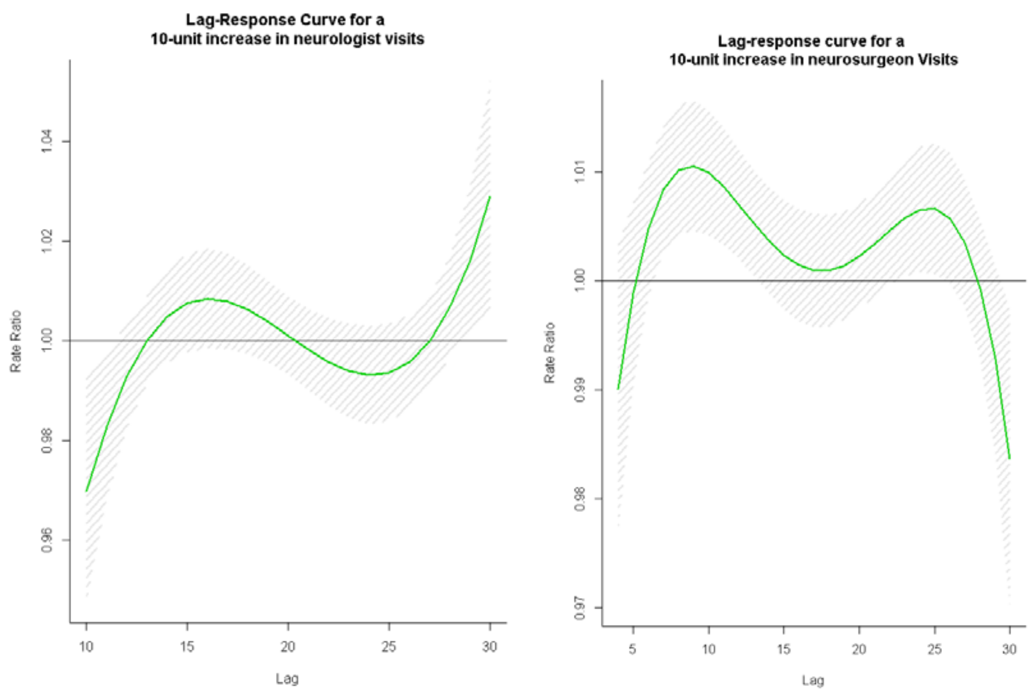

Figure 5. Plot of the lag-specific effects of a 10-visit increase in daily arrivals at outpatient neurological and neurosurgical clinics. Separate polynomials were fitted in the model, with the neurological visits represented as a cubic polynomial over lags 10–30, and the neurosurgical visits represented by a 4^{th}-order polynomial over lags 4–30.

(30-days). However, it is not apparent that this improvement is due strictly to the inclusion of the lagged outpatient information; the univariate Poisson model using only day-of-the-week and holiday factors achieved similar performance, implying that the improvement from incorporating the lagged terms is small at best.

Effect	Model with Outpatient Series			Model without Outpatient Series		
	Estimate (Rate Ratio)	SE	p	Estimate	SE	p
Monday	0.309	0.0683	<<0.0001	0.325	0.0655	<<0.0001
Tuesday	-0.434	0.0813	0.21519	0.122	0.0682	0.0734
Wednesday	-0.399	0.0815	0.12720	0.142	0.0679	0.0360
Thursday	-0.0417	0.0738	0.57227	-0.0313	0.0710	0.6589
Saturday	0.0885	0.0714	<<0.0001	-0.461	0.0799	<<0.0001
Sunday	0.108	0.0710	<<0.0001	-0.446	0.0795	<<0.0001
Holiday	-0.588	0.144	<<0.0001	-0.537	0.138	<<0.0001
Lag Term: Neurologist, degree 0	-0.0207	0.00733	0.00475			
Lag Term: Neurologist, degree 1	0.101	0.0361	0.00510			
Lag Term: Neurologist, degree 2	-0.157	0.0564	0.00530			
Lag Term: Neurologist, degree 3	0.0781	0.0281	0.00536			
Lag Term: Neurosurgeon, degree 0	-0.00859	0.00322	0.00758			
Lag Term: Neurosurgeon, degree 1	0.0882	0.0304	0.00372			
Lag Term: Neurosurgeon, degree 2	-0.279	0.0953	0.00342			
Lag Term: Neurosurgeon, degree 3	0.354	0.121	0.00337			
Lag Term: Neurosurgeon, degree 4	-0.156	0.0532	0.00329			

Table 1. Estimated coefficients and standard errors from the distributed lag model with day-of-the-week and holiday factors (left). Also shown (right) are the parameters from the model excluding the distributed lag terms.

Since only one particular data period was observed in this study (1/6/2013–12/31/2015), general statements about the relative performance of the models being examined were difficult to make. In analyzing time series data, the presence of serial dependence makes it unlikely that each observation may be considered independent, and frequently the entire series must be considered as a single realization of some underlying process. In addition, the forecast period itself spanned a particular period of time (approximately January 2016) and may be subject to some bias due to its particular "location" in seasonal cycles or the presence of holidays. To address this, the two primary models (Poisson without distributed-lag covariates and Poisson with distributed-lag covariates) were refit to all possible annual and quarterly windows, summarizing the mean RMSE, and MAE of the 7- and 30-day forecasts.

	7 – Day Forecasts		30-Day Forecasts	
Forecast Method	**RMSE**	**MAE**	**RMSE**	**MAE**
Mean Forecast	1.43	1.24	2.01	1.61
Seasonal Naïve Forecast	1.25	1.00	2.10	1.53
Seasonally Adjusted Naïve Random Walk	**1.06**	0.92	2.23	1.60
SARIMA(0, 0, 0)(2,0,0)$_7$	1.27	0.95	2.00	1.53
Seasonally Adjusted Exponential Smoothing (Additive Errors)	1.25	**0.87**	1.94	1.52
Neural Network: NNAR(29, 1, 15)$_7$	1.73	1.47	1.96	1.62
Poisson Regression (Without Outpatient Predictor Series)	1.44	1.35	1.86	**1.44**
Poisson Regression (With Outpatient Predictor Series)	1.35	1.31	**1.83**	1.45

Table 2. Performance of short- and long-term forecasts (compared to actual observations) from all tested models, including univariate time series models, and Poisson regression with and without the lagged outpatient data. The multivariate Poisson model performed the best in terms of 30-day forecasts with the exception of MAE error, although the Poisson model excluding the outpatient data performed nearly equivalently.

Although the results using the entire dataset showed slightly improved performance using the Poisson model with the outpatient information (compared to the model using only day-of-the-week and holiday variables), refitting the model using fewer data points (e.g., fitting to a single quarter or year of data) resulted in poorer performance using the outpatient information even after considering all the possible time periods.

		7 – Day Forecasts		**30-Day Forecasts**	
Training Data	**Model**	**Mean RMSE [5%, 95% percentiles]**	**Mean MAE [5%, 95% percentiles]**	**Mean RMSE [5%, 95% percentiles]**	**Mean MAE [5%, 95% percentiles]**
All Possible Quarterly (120-day) Periods, n = 970	Includes Outpatient Data	2.37 [1.11, 5.81]	2.01 [0.91, 4.90]	2.24 [1.42, 4.55]	2.14 [1.51, 4.16]
	Excludes Outpatient Data	1.56 [0.88, 2.32]	1.27 [0.70, 1.89]	1.61 [1.29, 1.89]	1.27 [1.02, 1.51]
All Possible Yearly (365-day) Periods, n = 725	Includes Outpatient Data	1.56 [1.15, 1.97]	1.32 [0.96, 1.67]	1.58 [1.27, 1.91]	1.65 [1.47, 1.82]
	Excludes Outpatient Data	1.48 [0.82, 2.22]	1.21 [0.66, 1.79]	1.55 [1.22, 1.86]	1.22 [0.96, 1.46]
All Possible Two-Year (730-day) Periods, n = 360	Includes Outpatient Data	1.49 [1.11, 1.85]	1.28 [1.00, 1.57]	1.57 [1.30, 1.88]	1.66 [1.52, 1.79]
	Excludes Outpatient Data	1.49 [0.85, 2.27]	1.20 [0.67, 1.77]	1.55 [1.24, 1.85]	1.21 [0.97, 1.46]

Table 3. Summarized RMSE and MAE of 7- and 30-day forecasts from the Poisson-Distributed-Lag model incorporating the outpatient information compared to the same Poisson model excluding the outpatient covariates. The models were fitted to every possible fixed-length time period in the time frame represented by the data to permit informed generalizations about the predictive power of the outpatient information at different time points.

5. Discussion

5.1. Implications

The goal of this investigation was to determine the improvement, if any, obtained by incorporating lagged outpatient data into a forecasting model for predicting the number of daily arrivals at an inpatient of Test Hospital. The model selection process, which was performed by finding the combination of lagged terms and parameters having the minimum AIC, yielded a Poisson model incorporating both recent and older (up to 30 days) outpatient information, with several points at which the lagged outpatient information showed a significant effect on the estimated rate parameter. Short-term forecasts from the chosen model were outperformed by nearly all the common univariate time-series approaches employed (random walk, ARIMA, exponential smoothing, neural network) but this trend was reversed when longer-term forecasts were considered.

Furthermore, refitting the model to other time periods within the available dataset showed that the observed performance of the two Poisson models did not necessarily represent the conceivable range of performance when fit to different data. To investigate this form of robustness, the models were fit to differently sized periods of data ranging from 120 days (quarterly) to 730 days (two years; the original data spanned nearly 3 years—1090 days) that were allowed "slide" across all possible starting dates. Based on these simulations, the simpler model tended on average to outperform the model containing the outpatient information.

Although these differences tended to become smaller as more data was used to train the models, it would not be appropriate to interpret this as necessarily a sign of improved precision; after all, the 360 different datasets used to train and compute the performance of the two-year models were mostly trained using a shared set of common data points (e.g., the model trained on data from January 2013 to January 2015 used nearly the same data as the model taking data from February 2013 to February 2015). The results of this sensitivity analysis permitted the observed performance of the models (using all of the data) to be placed in some context, allowing one to infer what the results may have been had the models been trained on different data and perhaps decide whether the models' performances are generalizable.

Taken together, these results imply that the inclusion of lagged outpatient information may not be unilaterally helpful in building models to accurately forecast inpatient arrivals. Although the presence of "gatekeeping" structures and pathways present in many health systems suggest that various points of care could—ostensibly—inform one another, this relationship was not observed in the current data. This may be due to several reasons.

Firstly, the outpatient data for this present study was obtained by aggregating across all clinics pertaining to one of two related specialties: neurologists and neurosurgeons. In doing so, any between-clinic heterogeneity is ignored and the focus comes to be on trends between population movements related to visiting the hospital after having seen a specialist rather than on individual clinic-to-hospital transitions. More importantly, visits to outpatient clinics that do in fact function as "upstream indicators" of future hospital visits will be combined with clinics that may not have this function. For example, specialists that focus on patients with more severe diagnoses may frequently handle patients who are likely to be admitted to a hospital, while other specialists that see patients with fewer needs can provide care at the clinic, preventing a hospital admission altogether. Differences in the types of patients seen at particular clinics or in their size or makeup may be critical in forming the sort of lagged relationship necessary to successfully make forecasts as has been attempted in this present study, and ignoring this heterogeneity may have led to overestimation of the variance on the model parameters and subsequent failure to detect significant trends.

Secondly, a major concern in making forecasts that is unique to time series data is in ensuring that the process being forecasted does not already respond to other, possibly similar forecasts (Hyndman & Athanasopoulos, 2013). In the present context, a major reason that a Poisson model with day-of-the-week and holiday variables was used as a "baseline" model from which to compare the model incorporating the lagged covariates is because similar models are sometimes used by nurse managers and resource planners at Test Hospital. In essence, this creates problems in that the observed data originate from a structured process which is being actively anticipated and perhaps guided by forecasters. Hospital planners who are already using forecasts have the ability to modify the true random process of arrivals, allocating resources in such a way to avoid volatile activity or unplanned admissions. For example, the observed data may not reflect a random process but rather one in which hospital resource planners—who, having advance knowledge (accurate or not) through the use of forecasts—have attempted to alleviate unwanted surges in arrivals and otherwise ensure that things "run smoothly."

In a more general sense, the type of structure presumed to describe that of Test Hospital and its associated outpatient clinics—in which outpatient clinics "feed into" inpatient clinics in a downstream fashion— implies that patients who visit an outpatient specialist present with symptoms or needs that can only be addressed by admission to the hospital, to which they are subsequently sent. In reality, it may be more plausible that outpatient specialists function as "diverters" more so than as "senders," and function principally to address patient needs before they become sufficiently exacerbated as to require an inpatient admission. In this way, a properly functioning outpatient clinic may serve to reduce, as much as possible, the number of patients that subsequently become admitted to the hospital and have the reverse effect that is hypothesized in this present study. Although this in itself would not bias the estimated effect found by regressing on this data, it does call attention to the role of forward and reverse causality in the context of modeling lagged relationships. Likely, the causal effect—if one exists—of patient surges at outpatient clinics on subsequent inpatient admissions depends on various patient-level characteristics (e.g., particular illnesses with seasonal prevalence or more/less severe needs) that cannot be observed by merely examining total counts.

5.2. Recommendations

From a practical standpoint, although the results of this current study imply that lagged outpatient information may not improve forecasts of inpatient arrivals, the data represent only a particular hospital system observed at a particular window of time. Furthermore, although the use of distributed-lag polynomials to model the outpatient arrivals provides a convenient and flexible framework for dynamically modelling predictors that are observed as time series data, other approaches may yield different results. Vector autoregressive (VAR; Jones et al., 2009) models have been used with success to produce accurate forecasts based on interrelated time series data. However, the relative difficulty of fitting such models may make them intractable to all but the most statistically inclined resource planners, whereas the relatively simple regression model employed in this study may be more desirable among forecasters.

Finally, since the minimum lag modelled in the distributed-lag polynomials carries implications for the model, this may limit its practical applications. For example, in the model proposed in this study, lagged neurosurgeon visits were modeled over lags 4–30, which means that 4-day-old data is needed in order to compute forecasts. Although it is conceivable to use previously forecasted values as predictions of the lagged values needed to make a forecast further into the future (indeed, this is the approach taken by many other techniques, such as ARIMA or Holt-Winters Smoothing), this necessarily injects additional variability into the forecasted values and becomes increasingly unstable. For this reason, the proposed model may have little practical utility, as the forecasts can be made available no sooner than 4 days prior to the actual dates being forecasted.

In conclusion, the present study shows that, even in a flexible distributed-lag framework, for which many model selection parameters can be tuned, there may be insufficient predictive information in outpatient arrival data to justify their inclusion into inpatient forecasting models. In addition, the steps taken to ensure optimal representation of the distributed lag polynomials (such as selecting the lag length and degree) can impose restrictions on the ways in which the forecasts may be generated and used, which may be undesirable. Although there still exists some hypothetical promise in the use of lagged outpatient data to inform future or present inpatient observations, newer and more creative methods of incorporating this information should be sought.

References

Asplin, B. R., Flottemesch, T. J., & Gordon, B. D. (2006). Developing models for patient flow and daily surge capacity research. *Academic Emergency Medicine, 13*(11), 1109–1113.

Bergs, J., Heerinckx, P., & Verelst, S. (2014). Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis. *International Emergency Nursing, 22,* 112–115.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. New Jersey: John Wiley & Sons, Inc.

Broyles, J. R., Cochran, J. K., & Montgomery, D. C. (2010). A statistical Markov Chain approximation of transient inpatient inventory. *European Journal of Operational Research, 207,* 1645–1657.

Clarke, K. A. (2005). The phantom menace: Omitted variable bias in econometric research, presented at International Studies Association Conference 2005 Annual Meeting, Istanbul. Storrs, CT: International Studies Association.

Follette, G., & Sheiner, L. (2005). The Sustainability of Health Spending Growth. Federal Reserve Board, Division of Research & Statistics and Monetary Affairs. Social Science Research Network. Retrieved from https://www.federalreserve.gov/pubs/feds/2005/200560pap.pdf

Forrest, C. B. (2003). Primary care gatekeeping and referrals: effective filter or failed experiment? *British Medical Journal, 326,* 692–695.

Gasparrini, A. (2014). Modelling exposure-lag-response associations with distributed lag nonlinear models. *Statistics in Medicine, 33,* 881–899.

Glied, S., Ma, S., & Solis-Roman, C. (2016). Where the Money Goes: The Evolving Expenses of the U.S. Health Care System. *Health Affairs*, *35*(7), 1197–1203.

Granger, C. W. J., & Newbold, P. (1975). Spurious regressions in econometrics. *Journal of Econometrics, 2*(2), 111–120.

Hyndman, R.J. & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. Available at http://otexts.org/fpp/. Accessed on March 10, 2017.

Hyndman, R. J. (2017). *Forecast: Forecasting functions for time series and linear models*. R package version 8.0, available at http://github.com/robjhyndman/forecast.

Jones, S. S., Thomas, A., Evans, R. S., Welch, S.J., Haug, P. J., & Snow, G. L. (2008). Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine, 15*(2), 159–170.

Jones, S. S., Evans, R. S., Allen, T. L., Thomas, A., Haug, P. J., & Snow, G. L. (2009). A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of Biomedical Informatics, 42,* 123–139.

Kam, H. J., Sung, J. O., & Park, R. W. (2010). Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthcare Informatics Research, 16*(3), 158–165.

Kashihara, D., & Carper, K. (2012). Statistical Brief #355: National Health Care Expenses in the U.S. Civilian Noninstitutionalized Population, 2009. Agency for Healthcare Research and Quality, Rockville, MD. Retrieved from http://www.meps.ahrq.gov/mepsweb/data_files/publications/st355/stat355.shtml

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*(1–3), 159–178.

Littig, S. J., & Isken, M. W. (2007). Short term hospital occupancy prediction. *Health Care Management Science, 10,* 47–66.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Shrestha, S. L. (2007). Time series modelling of respiratory hospital admissions and geometrically weighted distributed lag effects from abient particulate air pollution within Kathmandu Valley, Nepal. *Environmental Modeling & Assessment*, *12,* 239–251.

Sun, Y. Heng, B. H., Seow, Y. T., & Seow, E. (2009) Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine, 9*(1).

The Henry J. Kaiser Family Foundation (2012). Health care costs: A primer. Retrived from https://kaiserfamilyfoundation.files.wordpress.com/2013.01.7670-03.pdf

Wang, T., Yu, Y., Zhou, W., Liu, B., Chen, D., & Zhu, B. (2016). Dynamics of material productivity and socioeconomic factor based on auto-regressive distributed lag model in China. *Journal of Cleaner Production, 137,* 752–761.

Wargon, M. Guidet, B., Hoang, T. D., & Hejblum, G. (2009). A systematic review of models for forecasting the number of emergency department visits. *Emergency Medicine Journal, 26,* 395–399.

Weiss, M.E., Yakusheva, O., & Bobay, K.L. (2011). Quality and Cost Analysis of Nurse Staffing, Discharge Preparation, and Postdischarge Utilization. *Health Services Research*, *46*(5), 1473–1494.

Wiler, J. L., Griffey, R. T., & Olsen, T. (2011). Review of modeling approaches for emergency department flow and crowding research. *Academic Emergency Medicine*, *18*(12), pp. 1371–1379.