A CONCEPTUAL FRAMEWORK FOR MODELING LONGITUDINAL HEALTHCARE ENCOUNTER DATA

Hari Balasubramanian
Nora Murphy
Michael Rossi

University of Massachusetts, Amherst
Department of Mechanical and Industrial Engineering
160 Governor’s Drive
Amherst, MA 01003, USA

ABSTRACT

We discuss a framework for analyzing data concerning healthcare encounters at the individual level. These encounters can be of various types – outpatient, emergency room, inpatient, pharmaceutical etc., each corresponding to one or more diagnoses. Each encounter happens on a certain day (or a certain hour) and when such data is collected over a period of time, it creates an evolving point process unique to each individual. The point process provides information about the intensity and diversity of encounters – how frequent and how fragmented the care is across multiple settings. When such longitudinal “point process” data is available for a cohort of individuals, it is possible to analyze the aggregate burden of managing the cohort’s care in a particular time period. We provide examples where such data could be used and discuss the stochastic methods that are best suited for generating insights.

1 INTRODUCTION

The two main questions we address in this paper are: (1) How should time-varying event data at the individual level be used for decision-making? (2) What stochastic methodologies are best suited to address such data both at the individual and aggregate level? While the questions are broad and applicable in many domains, to keep the discussion tangible, we will consider them in the context of care coordination. Our primary goal in this paper is to suggest a framework, a point from which future modeling efforts can proceed.

We start with a specific example from an article in the August 2014 issue of the New England Journal of Medicine (Press 2014). In the article, Dr. Matthew Press, a primary care/internal medicine physician at the Massachusetts General Hospital (Boston, MA), describes the challenges in coordinating the care of just one patient. The patient, Mr. K, a 70-year old man, had thus far had only minor medical problems. Dr. Press writes in the article:

“M. K’s care was fairly straightforward— I was the only doctor he saw regularly — until the day he came into my office with flank pain and fever. A CT scan of his abdomen revealed a kidney stone — and a 5-cm mass in his liver, which a subsequent MRI indicated was probably a cholangiocarcinoma…Over the 80 days between when I informed Mr. K. about the MRI result and when his tumor was resected, 11 other clinicians became involved in his care, and he had 5 procedures and 11 office visits (none of them with me). As the complexity of his care increased, the tasks involved in coordinating it multiplied…In total, I communicated with the other clinicians 40 times (32 e-mails and 8 phone calls) and with Mr. K. or his wife 12 times. At least 1 communication occurred on 26 of the 80 days, and on the busiest day (day 32), 6 communications occurred.” (Excerpt from Press 2014)
The article provides an animation of how the care evolved for Mr. K. The animation shows the patient gradually getting linked to a network of clinicians in the 80-day time period: a urologist, a surgeon, a hematologist, a neurologist, a lab technician, a gastroenterologist, an interventional radiologist, an oncologist, a cardiologist, a pathologist, and a social worker. As the patient’s primary care physician and responsible for his safety and overall care, Dr. Press had to closely track what happened to Dr. K.

We could also visualize Mr. K’s healthcare encounters as an *longitudinally evolving point process*, described by a simple dot plot in Figure 1. The dot plot is not the actual process because we do not have the timeline that Press (2014) developed; rather it is a conceptualization of what happened. Mr. K had the occasional healthcare visit with Dr. Press, but experiences a sudden burst of encounters. Each encounter generates a point on the timeline. The different shapes indicate different encounter types with various specialists. The diamond shape might indicate a visit with a gastroenterologist, the circular shape a visit with the lab technician, the square shape a visit with the PCP, and so on. The situation was resolved in the 80-day period after which Mr. K now returns to his normal, pre-escalation visit intensity.

![Figure 1: A conceptual illustration of a longitudinally evolving point process.](image.png)

We define a point-process formally as follows. Consider a discretized timeline, where the equal size intervals are indexed by $t$. In this paper, we use a day as the smallest interval. The length of our timeline, $T$, is set to a reasonably long time period, say 730 days (2 years). So $t$ runs from 1 to $T$. Let the set $E$, indexed by $e$, represent the types of events that an individual can experience. For example, a doctor’s office visit event needs to be distinguished from an emergency room event. A point process for an individual is fully described by a binary parameter $B_{e,t}$ (for all $e$ and $t$) where $B_{e,t} = 1$ if event of type $e$ happened in time interval $t$ and 0 otherwise.

The advantage of creating a point process for each individual is that the individual’s history is visualized, allowing the clinicians to infer the sequence of events from the patient’s perspective, the time elapsed between successive events, and thereby provide customized care and treatment. Each point in the timeline potentially represents value-added information that requires the attention of one or more providers. To avoid the possibility of duplication and medical errors, each specialist/clinician involved in the patient’s care will have to invest time in understanding the information generated by each encounter. For example, the patient might have been prescribed medications that conflict with each other, raising the possibility of a complication or adverse event. The more diverse the event types in a particular time period, the more time a provider has to spend in coordinating care. In the above example, Dr. Press had communicate with the patient and numerous other specialists by phone and email (32 emails and 8 phone calls), in addition to spending time evaluating diagnostic results.

Thus the greater the frequency and diversity of encounters (i.e. if more specialists/clinicians are involved), the greater the chance of safety risks, and the greater the capacity burden on the clinicians in the health system.

Longitudinal point process data for each patient makes it possible to quantify the frequency and diversity of encounters over time. It can provide a more holistic perspective unavailable in encounter based or provider based data sets. For example, if we had a dataset that pertains only to the surgical procedures, we would not know of what other conditions or events happened to the patients who underwent the procedures. The analysis of longitudinal data makes possible to test if recurring patterns can be identified for at risk, expensive, or high utilizing patients. If such patterns do exist, then
interventions can be tried earlier to improve patient care. Longitudinal data can be used to inform clinical decisions for individual patients: for example, observing the history of events in real time, a clinician may decide to call a specialist or change the patient’s medications. When longitudinal data is assembled for a cohort of patients, the demand distribution for different events in the health system (ER, primary care and specialty care, inpatient) in a given time period can be estimated. This can help with capacity planning at the level of the entire health system.

The literature most relevant to this paper is the stochastic theory on point processes. The Poisson process is the most basic type of point process and widely used. Palm-Khintchine’s theorem (Heyman and Sobel, 2003) states that when sufficiently many independent point processes (not necessarily Poisson) are aggregated, the count of events in a particular time interval (obtained by superposition) begins to approach a discrete Poisson distribution. This property is relevant when we are concerned with a cohort of patients. Also relevant is the medical decision making literature in which disease progression or transition between a patient’s health states are modeled as a Markov process. See Oghuzan et al. (2009) for a tutorial and Maillart et al. (2008) for an application to breast cancer. While this literature too extensive to review here, to the best of our knowledge, a data-driven Markov modeling approach, which we propose in the paper, has not been used to analyze transitions between events in the health system and their implications for care coordination. The framework in Deo et al. (2013) comes closest in that it models disease progression and time between chronic care appointments for individual patients within a Markov Decision Process framework and a capacity constraint.

2 ESTIMATING LONGITUDINAL HISTORIES FROM THE MEDICAL EXPENDITURE PANEL SURVEY

We now discuss how longitudinal histories on system-wide encounters can be constructed from a national dataset called the Medical Expenditure Panel Survey.

The Medical Expenditure Panel Survey (MEPS) is an extensive survey of the United States population and it is the most comprehensive source of information for health care utilization and costs. The survey began in 1996 and is often used as the primary source of data for extrapolating health care statistics for the entire country. A representative sample of the population is surveyed and information is collected regarding “demographic characteristics, health conditions, health status, use of medical services, charges and sources of payments, access to care, satisfaction with care, health insurance coverage, income, and employment” (see meps.ahrq.gov). Each individual remains in the survey for two years and all of their health events, are tracked and recorded during this period. This makes MEPS unique in that it is the only source of publicly available longitudinal patient data.

As an example, we consider the case of a 50-year old female from the MEPS 2011 and MEPS 2012 datasets. This patient had no emergency room events or hospitalizations in 2011. However she visited a doctor’s office 42 times and had 198 prescription refills in that year which cost $199,672; this appears to have been the primary contributor to her healthcare expenditures in 2011. Further, the office visits were not to a single provider, but to 8 different providers as seen the breakdown of the 42. In 2012, she had 57 doctor’s office visits and 4 outpatient department visits (these usually refer to procedures) and 89 prescription refills.
The MEPS dataset contains the exact date of each event. Thus we can create a timeline of healthcare encounters for the above individual for a 730-day (2-year) period. Approximately 16000 individuals are newly surveyed each year and the majority of them also stay on in the survey the following year. Thus it is possible to create 2-year timelines for each such individual. The vast majority of patient timelines are empty or sparse, since many patients are healthy. But a small percent of patients, such as the patient in Figure 2, generate a large number of encounters and costs. The point processes of these individuals therefore can be analyzed, their frequency and diversity can be quantified.

While MEPS can used to infer general patterns based on individuals surveyed in different parts of the US, it cannot be used to analyze a cohort of patients living in a particular city or small area. Large integrated health systems such as the VA, Mayo Clinic, Massachusetts General Hospital and academic medical centers, which provide the full range of health services (outpatient, emergency, inpatient, surgical) have electronic records that allow longitudinal histories to be created in real-time.

3 A DISCRETE TIME MARKOV CHAIN APPROXIMATION

Once longitudinal histories become available, what methodologies are best suited to analyze them? Markov models are the most obvious choice to quantify a sequence of changes in time. Discrete, continuous time and semi-Markov process are all capable of quantifying such changes. In this paper, due to the discretized nature of our timelines (granularity of one day), we illustrate the methodology with Discrete Time Markov Chains (DTMCs). DTMCs typically assume time-homogeneous event rates and allow only for one step memory. As we saw in the introduction, point process data for an individual is unlikely to be time-homogeneous: the rate at which events happen is likely to change. Further, the process is not memoryless since an event triggers a sequence of other events in the immediate future. Nevertheless, the structure of a DTMC is reasonable first approximation to quantify the rate and diversity of care transitions.

Consider a framework in which the data available in MEPS (Section 2) is used to create a Discrete Time Markov Chain (DTMC) of transitions between different types of health encounters. For ease of exposition, we consider only office based visits (to primary care physicians and specialists), outpatient visits (for procedures), emergency department (ED) events, inpatient hospitalizations and home health
Balasubramanian, Murphy, and Rossi

More detail could easily be included in this framework, for example pharmaceutical events, and separate states for office-visits to each specialty.

![Discrete Time Markov Chain Diagram](image)

Figure 3: States and Possible Transitions in the Discrete Time Markov Chain.

The discrete epoch when transitions happen can be assumed to be one day, however this can be shortened to half a day or even an hour if the exact hour the event happened is known. Sometimes multiple events can happen on the same day so a more granular transition epoch may be desirable. There are three main segments in the transition diagram: start/end waiting states, active states, and intermediate wait states. Each patient begins their survey period in the start/end state and then if they had a health event in their 2-year period, the patient will transition to the beginning wait state. If the patient transitions to the beginning wait state, they will wait there until their first event and then they will transition to that activity state. There are six activity states, one for each event type that MEPS tracks and a patient can only spend one day in an activity state. From the activity state, the patient will transition to the corresponding post_activity state if the previous event was not the last for the survey period. Then from a post_activity state, the patient can transition to any of the other activity states. This functionality holds true for all activity types besides inpatient. If a patient has an inpatient event, they will first transition to inpatient_in then to inpatient for the duration of the hospital stay and then to inpatient_out for the discharge and from inpatient_out the patient will transition to post_inpatient. If any event was the last health event for the patients 2-year study period, then they will transition from the activity state to the end wait state and they will stay there for remaining duration and then transition back to the start/end state. It is easier to understand figure 1 with the following rules:

- Each blue state is a state in which the patient is not actively utilizing the health system while the red states represent states in which the patient is using healthcare.
- Patients can only stay in oval shaped states for one day and they can stay in rectangular states for an indefinite period of time.
- Patients can only transition from an activity state to the corresponding wait state, but they can transition to any activity state from a wait state (excluding inpatient events).
Balasubramanian, Murphy, and Rossi

Thus each individual surveyed in MEPS will follow a certain trajectory or a realization through this state diagram. We now provide a numeric illustration of how the DTMC would be parameterized for relatively healthy patient with the following point process in a 2-year period:


The above point process can be interpreted as follows. The first day that an event can happen is Jan 1\textsuperscript{st}. In approximately 8 months (241 days after Jan 1\textsuperscript{st}: we assume that an epoch is 1 day), the above person visited a doctor/PCP’s office for an annual checkup. In the Markov model, this would be equivalent to spending 241 days in the Beginning Wait state and then transitioning to the Office Based state where the patient stays for one day. Then the person transitions to Post_Office state and stays there for 300 days, before having another annual exam in the second year. Finally the patient transitions to the End Wait state until the second year ends. In total there are 365*2 = 730 transitions. The transition count matrix which captures the number of times the individual moved from one state to another is shown in Table 1. Note that since the individual did not have any events other than two office based visits, the remaining states are unvisited and are therefore not part of the matrix. From the count of transitions, the conditional transition probabilities, that is the probability of going to state $j$, given that the person is currently in state $i$, can be calculated. Let $S$ represent the set of states and $C_{ij}$ represent the number of transitions from state $i$ to state $j$. Then the transition probability estimate is $\frac{C_{ij}}{\sum_{j \in S} C_{ij}}$.

For example the probability estimate of returning to an office visit, given that the individual is currently in the Post_Office state is 1/(1+299) = 1/300. The Begin Wait, Post_Office and End Wait states together capture days when the patient did not have any events. Although, the example is simple, it is not stylized. A large percent of the US population does not use healthcare very much and their except for annual physical exams and their transition matrix will have a structure very similar to Table 1.

Table 1: DTMC matrix showing the count of transitions from one state to another.

<table>
<thead>
<tr>
<th></th>
<th>Start/End</th>
<th>Begin Wait</th>
<th>Office Based Visit</th>
<th>Post_Office</th>
<th>End Wait</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start/End</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Begin Wait</td>
<td>0</td>
<td>241</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Office Based Visit</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Post_Office</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>299</td>
<td>0</td>
</tr>
<tr>
<td>End Wait</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>185</td>
</tr>
</tbody>
</table>

For individuals who have multiple chronic conditions, we can expect to see a diverse set of encounters covering a larger set of states. The point process for one such patient may look as follows:


We see that the above individual had two emergency events in the very first month, the latter of which leads to an 8-day hospitalization. Ten days after the hospitalization, several office based visits happen with some number of days elapsing between successive visits. This goes on throughout the year. The 2\textsuperscript{nd} year ends with a 4-day inpatient hospitalization.
Notice that if we created a conditional transition probability matrix uniquely representing this individual (i.e. the probability of going to state $j$, given that the person is currently in state $i$) then we would get a very different matrix compared to that of the first person (with only the two office visits). In particular, for the second individual, the large majority of the ‘Post’ states (Post_ER, Post_Inpatient, Post_Office) would be represented (indicating that the encounter types are diverse).

4 NUMERIC RESULTS

In the previous section, we demonstrated how each individual’s longitudinal history or point process could be quantified into a DTMC. We now discuss results in which DTMC is created in aggregate for two groups of individuals in the Medical Expenditure Panel Survey. Group A consists of all individuals ($n=28145$) surveyed in MEPS 2011 who had at least one healthcare encounter in 2 years. Group B consists of all individuals ($n=8956$) surveyed in MEPS 2011 with age greater than 50. Group B is a subset of Group A.

For each group, we create a transition matrix similar to Table 1, by adding up the counts across all individuals in the group (the total number of transitions for each group is in the millions). We then calculate the empirical probability estimate of being in state $i$ as follows.

Let $C$ be the sum total of all transitions between any states across all individuals in the group. Then the empirical probability estimate of being in state $i$ is therefore

$$\pi_i = \frac{\sum_{j \in S} C_{ij}}{C}.$$ 

In Table 2, we compare how the probability estimate of being in state $i$ differs between the two groups. The intermediate states such as Begin, Wait, End Wait and all the Post states, especially Post_Office state, understandably have larger probabilities. This is because more often than not, a patient has no events on a particular day. However, the average Group B patient spends less time in Begin, Wait and End Wait states (0.16 and 0.15 respectively) compared to the average Group A patient (0.21 and 0.25). This is because more events happen to Group B patients, reflected in the fact that the average Group B patient has a 0.56 probability of being in the Post_Office state compared to 0.44 for a Group A patient. Once a Group B patient has an office visit, she transitions to the Post-Office Visit state and is likely to return to that state, through another office visit (which is also why the $\pi_i$ value for Office Visit is higher for a Group B patient). The probability of having an inpatient hospitalization on any given day for a Group B patient is 0.00040, higher than 0.00026 for a Group A patient. These trends make sense since a Group B patient is older and therefore more likely to sicker than the average Group A patient.

Next, we look at specific transition probability estimates for the two groups. We focus only on the most important transitions in care, listed in Table 3. For example, having reached the Post_Inpatient state, what is the probability that a patient is likely to (a) continue to stay on in the Post_Inpatient state, (b) have an office visit, or (c) be hospitalized again (i.e. transition to Inpatient_In)? The time period after a patient is discharged from the hospital is considered a particularly crucial one. If adequate office-based care is not provided soon enough, the patient may be hospitalized again. Similar questions could asked about the Post_ER state. In general, Table 3 reveals that the probability of having an office visit from Post_Inpatient, Post_ER, Post_Office states is higher for Group B than it is for Group A.

5 OTHER APPLICATIONS

As illustrated in the examples in Section 5 (the patient with two office visits in two years and the patient with a more complicated sequence of events) each individual’s point process in the health system can be coded as a ‘string’ with a particular structure. Successive events types are always linked by the number of days elapsed between the events. Strings can be various lengths depending on the number of encounters. The publicly available MEPS dataset can be used to assemble thousands of strings, one for each surveyed individual. Below we list the meaningful insights that could emerge from analyzing such trajectories:
Balasubramanian, Murphy, and Rossi

1. Classification, Comparison and Prediction: Machine learning techniques could be used to categorize the strings into broadly similar groups. The similarity would be based on how closely matched the conditional probability matrices are (the matrix of raw counts could also be used).

Table 2: Empirical probability estimate of being in a particular state, by group.

<table>
<thead>
<tr>
<th>State</th>
<th>$\pi_i$ values</th>
<th>Group B</th>
<th>Group A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Begin. Wait</td>
<td>0.16772</td>
<td>0.25620</td>
<td></td>
</tr>
<tr>
<td>End Wait</td>
<td>0.15177</td>
<td>0.21653</td>
<td></td>
</tr>
<tr>
<td>Office Based Visit</td>
<td>0.01946</td>
<td>0.01235</td>
<td></td>
</tr>
<tr>
<td>Post_Office</td>
<td>0.56464</td>
<td>0.44265</td>
<td></td>
</tr>
<tr>
<td>Home_Health_Event</td>
<td>0.00089</td>
<td>0.00038</td>
<td></td>
</tr>
<tr>
<td>Post_Home</td>
<td>0.01771</td>
<td>0.00796</td>
<td></td>
</tr>
<tr>
<td>Inpatient_In</td>
<td>0.00040</td>
<td>0.00026</td>
<td></td>
</tr>
<tr>
<td>Inpatient</td>
<td>0.00270</td>
<td>0.00168</td>
<td></td>
</tr>
<tr>
<td>Inpatient_out</td>
<td>0.00040</td>
<td>0.00026</td>
<td></td>
</tr>
<tr>
<td>Post_Inpatient</td>
<td>0.01056</td>
<td>0.00838</td>
<td></td>
</tr>
<tr>
<td>Outpatient</td>
<td>0.00193</td>
<td>0.00104</td>
<td></td>
</tr>
<tr>
<td>Post_Outpatient</td>
<td>0.04521</td>
<td>0.02845</td>
<td></td>
</tr>
<tr>
<td>Emergency</td>
<td>0.00058</td>
<td>0.00057</td>
<td></td>
</tr>
<tr>
<td>Post_ER</td>
<td>0.01472</td>
<td>0.02198</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Selected transition probability estimates.

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Group B</th>
<th>Group A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post_Inpatient</td>
<td>Office Based Visit</td>
<td>0.0225</td>
<td>0.0184</td>
</tr>
<tr>
<td>Post_Inpatient</td>
<td>Post_Inpatient</td>
<td>0.9651</td>
<td>0.9730</td>
</tr>
<tr>
<td>Post_Inpatient</td>
<td>Inpatient_In</td>
<td>0.0032</td>
<td>0.0022</td>
</tr>
<tr>
<td>Post_ER</td>
<td>Office Based Visit</td>
<td>0.0191</td>
<td>0.0126</td>
</tr>
<tr>
<td>Post_ER</td>
<td>Post_ER</td>
<td>0.9634</td>
<td>0.97816</td>
</tr>
<tr>
<td>Post_ER</td>
<td>Inpatient_In</td>
<td>0.0097</td>
<td>0.0004</td>
</tr>
<tr>
<td>Post_Office</td>
<td>Office Based Visit</td>
<td>0.0293</td>
<td>0.0229</td>
</tr>
<tr>
<td>Post_Office</td>
<td>Post_Office</td>
<td>0.9675</td>
<td>0.9746</td>
</tr>
<tr>
<td>Post_Office</td>
<td>Inpatient_In</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td>Post_Office</td>
<td>ER</td>
<td>0.0006</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

Using the diagnosis codes, which are also available in MEPS, strings could be aggregated for say patients who have only diabetes; or patients who diabetes and asthma; or some other combination of chronic conditions of interest. The conditional transitional probability matrix could be trained for specific combinations of diseases. The average behavior of a cohort, the variance within a cohort as well as the differences between two cohorts (with different combinations of chronic conditions) can be measured. Once parameterized, the Markov chains can be used to simulate future realizations, and the accuracy of these predictions could be evaluated prospectively. The
classic outputs of a Markov chain such as first passage probabilities and durations, probability of transitioning to state \( j \) given that the individual is in state \( i \) in \( t \) epochs/days, mean recurrence times, all have useful interpretations. For example, if a patient has just been hospitalized, the probability of readmission in \( t \) days can be estimated.

2. **Periods of Escalation**: How often are escalations in encounter rates (higher than average encounter rates) observed in a timeline, and how long do such periods of escalation last? Escalations are important to quantify since medical errors are most likely to happen when care is fragmented across multiple settings. Such periods can be disconcerting for the patient, and it is important to have a nurse or a health coach help the patient navigate the healthcare landscape.

3. **Aggregate-level Capacity Planning**: As discussed earlier, when sufficiently many independent point processes are aggregated, Palm-Khintchine’s theorem (Heyman and Sobel, 2003) states that the count of events in a particular time interval begins to approach a Poisson distribution. The event counts in this context refer to visits in a particular time period: weekly number of primary care visits; weekly number of visits to particular specialties; weekly number of ER visits and hospitalizations. The distributions of these events can therefore be estimated for capacity planning purposes. This methodology is particularly relevant to an integrated health system or an accountable care organization (i.e. a consolidation of various providers) providing a range of health services to a population.

4. **Evaluating an Intervention**: If a patient or a group of such high-utilizing patients are chosen for an intervention, then it is possible to evaluate whether the conditional transition probability matrix has changed significantly after the intervention began. For example, many practices are now employing a team of nurses and social workers to proactively improve the health of patients with a history of multiple hospitalizations and chronic conditions. The nurses and social workers help the patients be self-reliant, address social needs such as housing and insurance, reconcile their medications, establish a connection with a primary care doctor, and facilitate connections with specialists if needed.

If these prevention-focused interventions work, then we should see a change in the point-processes before and after the intervention. Prior to the intervention, patients exhibit a high rate of ED visits and hospitalizations. In the Markov model outlined above, they would largely be in the post_ER or post_Inpatient states with a relatively high conditional probability of having another ER or inpatient event. If the intervention works, we expect to see a decrease in time spent in the post_ER or post_Inpatient states, and an increase in Office Based visits and Post_Office states. Even if a hospitalization does happen, the conditional probability of readmission might decrease compared to the period prior to the intervention, while the conditional probability of having an office-based visit and staying in post_office state will be higher. It is also possible to add special new ‘intervention’ states dynamically, to reflect the situation that the patient now has an increased intensity of nurse and social worker visits/contact. To be truly rigorous, the results need to be compared to a control group that has not had an intervention. If a randomized control trial is not feasible, it is possible to instead analyze those patients whose point processes have the closest match to the intervention group.

6 **CONCLUSION**

Simulation studies in the health applications realm have primarily focused on two areas: capacity planning for a particular health service (outpatient and surgical scheduling, emergency department or inpatient flow), and medical decision making under uncertain disease progression (for example, breast
Balasubramanian, Murphy, and Rossi

and prostate cancer screening). These areas remain fertile grounds for stochastic modeling and simulation efforts. However, few studies have focused on modeling individual trajectories in the health system and quantifying patterns of care in a population using longitudinal data. In this paper, we outline a data-driven conceptual framework to quantify the frequency and diversity of healthcare encounters experienced by an individual in the health system. We discuss how longitudinal histories can be estimated from a publicly available national dataset, the Medical Expenditure Panel Survey (MEPS). Next, we demonstrate how a discrete time Markov chain can be used to model the sequence of events and time elapsed between events.

The proposed framework still needs careful evaluation and validation before any conclusions can be drawn. In addition to encounters, the availability of email interactions between patient and provider, regularly measured clinical lab values or other quantifiable indicators of the patient’s health, would further enhance the inferences that can be drawn from an individual’s timeline. Beyond Markov models, there are opportunities to use machine learning and other statistical techniques for ranking, classification, and prediction.

ACKNOWLEDGMENTS

This research was funded partially by the National Science Foundation Grant 1254519. The views expressed in this paper are of the authors alone and not of the National Science Foundation.

REFERENCES


AUTHOR BIOGRAPHIES

**HARI BALASUBRAMANIAN** is an Associate Professor of Industrial Engineering at the University of Massachusetts, Amherst. He holds a PhD in Industrial Engineering and Operations Research from Arizona State University (2006) and was a post-doctoral research associate at Mayo Clinic in Rochester, Minnesota from 2006-2008. His research interests lie in operations research applied to healthcare delivery. His email address is hbalasubraman@ecs.umass.edu.

**NORA MURPHY** obtained her B.S in Industrial Engineering from the University of Massachusetts, Amherst. Her email address is noramurphy@umass.edu.

**MICHAEL ROSSI** is currently a doctoral student in Industrial Engineering at the University of Massachusetts, Amherst. His email address is mrossi09@gmail.com.