Published in final edited form as:

Med Decis Making. 2018 January ; 38(1): 107–119. doi:10.1177/0272989X17731753.

Uncovering Longitudinal Healthcare Behaviors for Millions of Medicaid Enrollees: A Multi-State Comparison of Pediatric Asthma Utilization

Ross Hilton, Ph.D.¹, Yuchen Zheng¹, Anne Fitzpatrick, Ph.D., RN, CPNP, MSCR², and Nicoleta Serban, PhD¹

¹H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology

²Department of Pediatrics, School of Medicine, Emory University

Abstract

Background—This study introduces a framework for analyzing and visualizing healthcare utilization for millions of children, with a focus on pediatric asthma, one of the major chronic respiratory conditions.

Methods—The data source is the 2005–2012 Medicaid Analytic Extract claims for 10 southeast states. The study population consists of Medicaid-enrolled children with persistent asthma. We translate multiyear, individual-level medical claims into sequences of discrete utilization events, which are modeled using Markov renewal processes and model-based clustering. Network analysis is used to visualize utilization profiles. The method is general, allowing the study of other chronic conditions.

Results—The study population consists of 1.5 million children with persistent asthma. All states have profiles with high probability of asthma controller medication, as large as 60.6% to 90.2% of the state study population. The probability of consecutive asthma controller prescriptions ranges between 0.75 and 0.95. All states have utilization profiles with uncontrolled asthma with 4.5% to 22.9% of the state study population. The probability for controller medication is larger than for short-term medication after a physician visit but not after an emergency department (ED) visit or hospitalization. Transitions from ED or hospitalization generally have a lower probability into physician office (between 0.11 and 0.38) than into ED or hospitalization (between 0.20 and 0.59).

Conclusions—In most profiles, children who take asthma controller medication do so regularly. Follow-up physician office visits after an ED encounter or hospitalization are observed at a low rate across all states. Finally, all states have a proportion of children who have uncontrolled asthma, meaning they do not take controller medication while they have severe outcomes.

Address correspondence to: Nicoleta Serban, Georgia Institute of Technology, H. Milton Stewart School of Industrial and Systems Engineering, 755 Ferst Dr. NW, Atlanta, GA 30332, [nserban@isye.gatech.edu], 404-385-7255.

Financial Disclosure Statement: The authors have no financial relationships relevant to this article to disclose.

Conflict of Interest Statement: The authors have no conflicts of interest to disclose.

Keywords

discrete event sequence profiling; longitudinal utilization behaviors; Medicaid claims data; pediatric asthma

INTRODUCTION

Data in healthcare are generated at every patient's encounter with the healthcare system, at every implementation of medical processes, with every decision made by healthcare organizations, and with every policy implementation in the healthcare ecosystem, resulting in billions of data points every day. Every patient in any medical setting generates an invaluable data point that can contribute to understanding what works, for who and where.

One health-related information technology (IT) that has provided substantive opportunities to study healthcare across large populations and many years is the medical claims system. Information coded in claims data is standardized to a great extent [1], hence making such data amenable to large scale studies. Developing methods to translate medical claims data into meaningful information is the first crucial step in medical decision making. Further development of adaptive and scalable data mining and statistical methods provide the means for analyzing these data.

In this study, we propose a method that translates medical claims data into individual-level utilization sequences, longitudinally over multiple years, and that models heterogeneity in individual-level healthcare utilization. The modeling approach is motivated by the need of drawing inferences on longitudinal utilization for pediatric asthma healthcare towards advancing health policy and medical decision making for pediatric asthma. Asthma is the most common respiratory chronic disease in children [2]. More than 10 million children have had asthma in their lifetime [3], with 42.9% classified as uncontrolled [4]. While asthma cannot be cured, with the appropriate medication and treatment plan, its symptoms can be controlled [5, 6]. Controlling asthma is important for children since it can prevent damage to growing lungs and can also improve their quality of life and potentially reduce the cost of care by preempting severe health outcomes [7–10]. The U.S. Centers for Disease Control and Prevention has identified pediatric asthma as a priority condition for intervention [11].

Healthcare utilization has been the topic in many healthcare studies, with most studies explaining the frequency of utilization with respect to patient characteristics and other determinants of utilization for various conditions [12–22] among others. Most studies rely on survey data [16, 23–25] and claims data [26–32], with few using electronic health records [33]. Our study is novel in that it provides a framework for studying longitudinal healthcare utilization from patient-level data, summarizing a large set of medical visits into a series of probabilistic inferences and visual displays that are easier to understand and manage by decision makers. More specifically, healthcare utilization for asthma care is limited to one state study. In most existing studies, asthma healthcare with respect to different care types and the impact on controlling asthma is evaluated overall the study population [34–38].

In particular, this study uncovers multi-year longitudinal utilization for pediatric asthma care using individual-level claims data for 10 states in the south for the Medicaid system. We focus on these states due to the poor health outcomes there [39]; some southeastern states also have among the highest expenditures, such as Georgia [40], of all states [11]. States in the south are similar in asthma trigger environmental factors but they also have different Medicaid systems; if differences are identified across the 10 states, the differences could be directly related to policy.

Medicaid is the largest insurance program for children in the U.S., with more than 27 million children enrolled nationally. Understanding healthcare utilization for treating asthma in the Medicaid-enrolled child population is important in assessing whether state Medicaid programs provide appropriate asthma healthcare, with comparisons across states [41].

METHODS

The method translates medical claims data into individual-level utilization sequences, longitudinally over multiple years, and that models heterogeneity in individual-level healthcare utilization using mixture Markov renewal processes. The method combines the benefits of network analysis and model-based clustering for discrete event sequences to also provide visual summaries of underlying utilization profiles. The method has the ability to scale to massive data while producing meaningful stochastic networks that can then be used in decision support through visualization and simulation.

Data Sources

The main data source is the Medicaid Analytical Extract (MAX) medical claims data acquired from the Centers of Medicare and Medicaid Services (CMS), consisting of identifiable individual-level claims data for all Medicaid-enrolled beneficiaries. The 10 states in the south are Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas. We provided a comparison in terms of asthma triggers and Medicaid managed care systems for the 10 states in the Web-Appendix A.

The Institutional Review Board (IRB) approval for this research was obtained under the protocol number H11287. The IRB protocol includes strict data safeguards for protected health information according to the data management plan with CMS. We provide the main data elements we extracted from the MAX files in the Web-Appendix B.

Study Population

The study population consisted of Medicaid-enrolled children ages 4–18 with an asthma-related diagnosis. (We excluded the age group 0–3 from this study because of the difficulty and inaccuracy of diagnosing asthma at this age.) Consistent with standard modifications [42] to the Healthcare Effectiveness Data and Information Set (HEDIS) measures defined by the National Committee for Quality Assurance (NCQA) [43], we defined patients with *persistent asthma* as those who meet one of the criteria:

1. At least two visits to the physician's office with a primary asthma diagnosis;
2. At least two asthma controlled medication prescriptions; or

3. One emergency room visit or hospitalization with a diagnosis of asthma in addition to at least another visit and/or medication prescription for asthma care.

While visits to physician's office, emergency room and hospitalization are commonly used to identify persistent asthma patients [36], regular intake of controller medication is also considered an important measure and a good predictor for emergency room visit and/or hospitalization [34–38]. These modified filtering criteria help to avoid including patients with incorrect diagnosis of asthma.

To obtain the population of children with persistent asthma, as defined above, we filtered the claims based on the ICD-9 diagnosis codes 493.XX, date of birth and utilization of outpatient, inpatient, emergency department, and medication to obtain the study population.

To capture longitudinal utilization behaviors, we only considered those patients that are enrolled in Medicaid for at least four of the eight years between 2005 and 2012.

Translating Claims into Individual-level Utilization Sequences

The MAX claims are structured into inpatient care (IP), other care including outpatient services (OT), patient summary (PS) and prescription claim summary (RX) files. Included for each claim are data entries specifying the date of service, the Medicaid Statistical Information System identification (MSIS ID) of each Medicaid enrollee, the International Classification of Diseases, Ninth Revision (ICD-9) codes for diagnosis or services provided, and the type and place of services rendered. We used the IP and OT files to determine the visits to a specific provider type, and the RX file to determine the medication type and date of the prescription being filled. We abbreviated the derived event types as follows: *emergency room visits (ER)*, *hospitalizations (HOS)*, *physician's office visits and clinic visits (PO)*, *asthma short-term medication (ASM)* and *asthma controlled medication (ACM)*. The first three event types were derived from the place of service and type of service codes of the claims code in the IP and OT files. The medications are grouped into ASM and ACM following the HEDIS nomenclature of asthma reliever medication and asthma controller medications. We also assume that if a prescription is filled, then the medication is utilized within the time of prescription.

The output of this translation step consists of individual-level sequences of utilization events with corresponding event time stamps. For example, consider a patient who visited the emergency room for an asthma attack on January 1st, 2005, who subsequently received a prescription for an inhaler which she filled one month later, along with a referral to a primary care physician. She then visited the same physician and refilled her asthma a three month interval. The resulting utilization sequence is (ER, ACM, PO, PO, ACM) with time stamps scaled to one-year intervals: (0.08, 0.25, 0.50, 0.75).

Additional details on the National Drug Codes used for asthma medication and other details on the translation process are provided in Web-Appendix B.

Model-based Clustering of Utilization Sequences

Modeling Utilization Sequences—We modeled patient-level utilization in the form of sequential event data over a period of time, where we considered both the order of the events (e.g. an ER visit precedes a PO visit) and the timing between events (e.g. the expected time between two PO visits).

A simple, but useful, model for summarizing time-ordered events with varying time intervals between events is the Markov renewal process (MRP) [44]. The MRP is the continuous-time analog of a discrete-time Markov chain (DTMC).

Let $\vec{X} = (X_1, X_2, \dots, X_L)$ the sequence of events and $\vec{T} = (T_1, T_2, \dots, T_L)$ the set of “arrival” times, where L is the length of the patient healthcare utilization sequence. Let $s_i, i \in \{1, \dots, S\}$ be all possible states in the sequences of events, in our case, they are ASM, ACM, ER, HOS and PO, where S is the number of states, in our case, $S=5$.

In an MRP, the sequence $\vec{X}$ is itself a DTMC, with corresponding transition matrix P where

P_{ij} is the transition probability between states s_i and s_j and $\sum_{j=1}^S P_{ij} = 1$. For example, the transition probability from ER to PO is the probability that a patient receives care in a physician’s office after an emergency room encounter. We estimate the transition probabilities P_{ij} using maximum likelihood estimation as presented in Web-Appendix C.

Defining the distribution for the sequence of interarrival times $\tau_i = T_{i+1} - T_i$, assume that for each pair $i, j \in \{1, \dots, S\}$, the distribution of the interarrival times between states s_i and s_j is an exponential distribution with rate parameter λ_{ij} . This assumption holds for the interarrival times of healthcare events as shown in Appendix C. To estimate the rate parameters we used maximum likelihood estimation provided in Web-Appendix C.

The output of the MRP consists of estimated transition probability and inter-event time matrices, specifying the transition probability and inter-event time expectation for each event type or state pair. Both output matrices are 5-by-5 matrices, where a cell in the transition probability matrix and in the inter-event time matrix respectively corresponds to the probability and the expected time of a patient with asthma to transition from one event type to the same event type or to a different event type. The matrix is not symmetric, since the transition probability from ER to PO may not be equal to the probability from PO to ER, for example.

In order to capture the probability distribution of the entire patient sequence we accounted for the choice of the first and last events by including left and right censor events. An illustration of an MRP output is in Web-Appendix C.

Clustering Analysis of Utilization Sequences—We complemented the MRP modeling of the utilization sequences with a clustering of the utilization sequences. We assumed that the cluster membership is a latent variable with a multinomial distribution, hence the resulting model is a mixture MRP. The MRP clustering algorithm simultaneously estimates the MRP model parameters and clusters patients into distinct utilization profiles.

We employed the estimation maximization (EM) with a hierarchical tree-based algorithm to derive the clustering output. In the first step, the algorithm searches for a division that maximizes the Bayesian information criterion score using the Kullback Leibler distance between the individual patient distributions and population distributions given a set of candidate divisions. We then used this candidate division as the initialization for the EM algorithm where patients move to clusters such that to maximize each individual posterior likelihood. After assigning patients to a new cluster, the parameters are re-estimated given current cluster membership via maximum likelihood estimation of both the transition probabilities and inter-event time distributions.

The computational complexity of our algorithm is $O(n \log n)$. The primary computational steps involved in fitting a patient sequence to an MRP rely on simple counting and averaging, while the computation of posterior likelihood relies on multiplication. The sorting step of the posterior likelihoods in determining the clustering membership is the most computationally expensive with order $O(n \log n)$. All other computations are of order $O(n)$. We present the computational time in hours with the number of patients or utilization sequences in Figure 1 in Web-Appendix C.

Once we identified a clustering tree, we further re-clustered nodes that have similar utilization patterns in terms of both the event types or states with higher probabilities as well as in terms of the expected interarrival times. This allows a sparser representation of the utilization profiles.

The details of the model structure and the model estimation are in Web-Appendix C.

Utilization Profile Visualization—By employing stochastic models for clustering utilization sequences we further derived utilization networks via the transition matrices, allowing for visualization of the utilization behaviors as networks across different healthcare types or states. The primary inputs for the utilization networks are the transition matrices. Specifically, the five event types, ASM, ACM, ER, HOS and PO, are the nodes in a directed graph. The directed edges represent transition probabilities between two event types, for example, the transition from the ER to a PO visit. For a simplified representation, the networks only include nodes such that a total of 90% of volume is represented. We used different types of arcs for different levels of transition probabilities to better identify nodes that are most visited within each profile.

We provide a graphical representation of the translation of one MRP as a network in Figure 1. For each state and cluster, we estimated the raw transition matrices: the probability transition matrix containing the probability of a follow-up visit from one provider type to another and the average interarrival time matrix containing the average length of time between visits in years. In this example, the only events with a high probability of visits are the physician's office, a short term prescription fill and hospitalization. Therefore, the transition matrices used as inputs contain only these three provider types. The nodes ASM, PO and HOS are circled in the resulting utilization networks. Consider the transition from ASM back to ASM in this example. The probability of a repeated ASM visit after a ASM visit is 0.58 with an average time between visits of 0.55 years. The directed edges represent

the transition probability between events and the average interarrival time, measured in years, between two consecutive events in parentheses. The left censor (LC) and right censor (RC) nodes represent the beginning and end of the study time period, January 1st, 2005, and December 31st, 2012, respectively. These two nodes are surrounded by dashed circles to differentiate them from actual healthcare events. The three styles and shading schemes of the lines corresponding to transitions between providers help the reader to visualize the high-probability patterns through the network.

The primary data source was acquired with support from Children's Healthcare of Atlanta, Institute of People and Technology at Georgia Tech and the School of Industrial and Systems Engineering at Georgia Tech. The development of the methodological research and its implementation was supported by a QuickWin grant by Children's Healthcare of Atlanta and the Institute of People and Technology at Georgia Tech and by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number R56HL126761.

RESULTS

Data Summaries

The target population of this study consisted of 1.5 million patients with persistent asthma who contributed to 24 million events over the eight year period (2005–2012) in the 10 states. Detailed statistics of event types per state are in Table I.

The number of children with asthma per state ranged from 40,147 in Mississippi, to 476,345 in Texas. Since different patients will have different numbers of Medicaid eligibility months per year, we normalized the event counts by *member-year*. We divided the counts by the number of eligibility months for each patient and multiply by 12.

On average, patients included in the analysis were enrolled in Medicaid for 74 months (6.1 years) from 2005 to 2012. The average number of events per member-year was 2.74; in high utilization states such as NC, patients had 3.22 events per member-year, and in low utilization states such as LA, patients had 2.36 events per member-year.

Out of the 24 million total asthma events, 13% were PO visits and 83% were RX events, including ACM and ASM. The numbers of PO and RX events per member-year across all states were $0.35(\pm 0.08)$ and $2.26(\pm 0.28)$, respectively. NC had significantly higher RX utilization at 2.65 ($p < 0.005$) events per member-year compared to FL, with only 1.86 RX events ($p < 0.005$) per member-year. Patients in FL visited the PO most frequently, with 0.53 per member-year, whereas those in MS visited the PO least frequently, with 0.25 per member-year.

ER and HOS events were less frequent than PO and RX among all states. The average number of ER and HOS events per member-year across the entire study population were $0.05(\pm 0.02)$ and $0.07(\pm 0.02)$, respectively. AR had primarily HOS events with an extremely small number of ER events. The lowest aggregated rate of ER and HOS events occurred in TX (0.07) and the highest occurred in MS (0.17).

Clustering of Utilization Sequences & Visualization of the Utilization Profiles

We first describe the underlying utilization patterns for medication arising across all utilization profiles for the 10 states:

ACM – high probability controller medication but low probability short-term medication;

ASM – high probability short-term medication but low probability controller medication;

ACM <-- ASM – high probability controller and short-term medication with high probability link of following up with controller medication after short-term medication; and

ACM/ASM – high probability controller and short-term medication with low probability link between the two medications.

Each utilization profile is characterized by one of the four medication patterns described above or no medication, in addition to other types of care, including PO, ER and HOS, depending on which of the event types are present in the utilization networks. For example, the profile ASM+PO describes utilization of short term medication and physician office for asthma care. The profile ASM+(PO,ER,HOS) describes utilization of short term medication along with physician office, emergency department and hospitalizations.

Table 2 provides the description of the profiles for all 10 states. Figure 2 is the utilization network for Alabama as an illustrative examples. Figures in Web-Appendix D are the utilization networks for the rest of the states.

Common features across all profiles in the 10 states were:

- All states had profiles with high probability ACM utilization (ACM or ACM <-- ASM). The percentages of the study population within such profiles across the 10 states were: 85%-AL, 60.6%-AR, 61.6%-FL, 72.2%-GA, 80.6%-LA, 70.1%-MS, 89.7%-NC, 65.2%-SC, 72%-TN, and 90.2%-TX.
- All states had one or more profiles with severe outcomes (ER or/and HOS).
- For those profiles with ACM <-- ASM pattern and severe outcomes, all of the profiles in MS, except for profile 1, had a high probability transition into ACM; the percentages of the study population among such profiles are: 4.3%-AL, 72.2%-GA, 60.3%-LA, 18.4%-TN.
- For those profiles with ACM, ACM/ASM or ASM medication patterns and severe outcomes, the link from ER or HOS to ACM had low probability; the percentages of the study population among such profiles were: 15%-AL, 18.8%-AR, 6.4%-FL, 11.9%-GA, 13.4%-LA, 11.8%-MS, 10.2%-NC, 18%-SC, 6.5%-TN and 4.7%-TX.

Table 3 provides ranges (minimum and maximum) of the transition probabilities across all profiles by state for links pertinent to recommended guidelines for asthma care. The ranges illustrate the lower and upper bound on frequency of follow ups inbetween two events for

different clusters in each state. For each link, a low upper bound indicates a consistently low level of follow ups for every patient, and a high lower bound indicates state wide high frequency of follow ups.

1. Follow-up visit after a severe outcome: The probabilities for the links ER/HOS → PO were generally lower than for the links ER/HOS → ER/HOS; the maximum value across all profiles of the 10 states for ER/HOS → PO varied between 0.11 (AR) and 0.38 (TN) and for ER/HOS → ER/HOS varied between 0.20 (TX) and 0.59 (MS).
2. Prescription (re)fill of controller versus short term medication after ER, HOS or PO event: The probabilities for controller medication after a PO visit were larger than for short-term medication. The probabilities for controller medication after an ER or HOS were lower than for short-term medication for some states, including AR, NC, SC and TX.
3. Consecutive prescriptions of controller medication had higher probabilities with shorter frequency between (re)fills than those for short-term medication across all states. The probability of consecutive prescriptions for asthma controller medication ranged between 0.75 and 0.95. The frequency of the consecutive prescriptions ranged from 0.04 (approximately 2 weeks) to 0.17 (approximately 2 months), with one exception where it was 0.53 or approximately 6 months, as shown in the network plots.

DISCUSSION

This study brings to bear substantive contributions to inferring healthcare utilization from a large database of claims data and to decision making on the healthcare delivery for pediatric asthma. This paper provides a rigorous approach to model multiyear longitudinal utilization, accounting for both the order of the events and the inter-event time using stochastic modeling, thus not only estimating the frequency of care events, as is common in the existing healthcare utilization research, but also the transition from one care event to another and the expected time between events, relevant in making inferences on the compliance with recommended care. The objective is to infer and visualize underlying profiles of individual-level utilization behaviors. Our framework begins with manipulation and processing of large flat files of administratively coded claims into meaningful data in the form of streamlined utilization sequences. The individual-level utilization sequences are then the input for a scalable model-based clustering analysis for discovering the underlying utilization profiles. Our methods are both rigorous and general, with applicability beyond the case study in this paper.

The paper is also the first study to draw inferences on patient-level healthcare pathways across approximately 1.5 million children with persistent asthma across 10 states.

According to the National Heart Blood and Lung Institute [45], the guidelines for asthma treatment specify the use of controller medication for children with persistent asthma. For all 10 states, controller medication was the most prevalent care event, with North Carolina being the highest utilizer and Florida and Georgia the lowest utilizers of controller

medication. Importantly, those children taking controller medication did so also on a frequent basis. Among the 10 states, Florida had the smallest proportion of children with frequent and high utilization of controller medications (~62%) compared to Texas with the highest proportion (~90%). While short-term medication utilization is much lower than the controller medication across all states, we found that there were groups of children who entirely relied on short-term medication. Among those children who did not use controller medication regularly, the probability of a severe outcome, including emergency department encounter or hospitalization, was higher than those who used controller medication as provided in Table 2. Utilization profiles (except one in Mississippi and one in Florida) with severe outcomes also had no transition from short-term medication to controller medication, indicating the importance of controller medication for persistent asthma. If asthma controller medication was present with high probability in the utilization profile, then it was also the follow-up event after a hospitalization or Emergency Department visit. The proportion of children in clusters with high prevalence of severe outcomes but no transition in controller medication varied across the states, from 4.7% in Texas to 18 in South Carolina.

While we only observe the prescriptions filled by a pharmacy and not those prescribed by a provider using claims data, we generally found that there was a higher probability of a controller medication than a short term medication following a physician office visit. This was not the same after a hospitalization or an Emergency Department visit; there were states where short-term medication had a follow-up rate than controller medication.

An important care practice after a hospitalization or Emergency Department encounter is the follow up with a physician office [46]. For all states, the probability of such follow-ups was low, with a maximum of 0.38 in Tennessee but as low as zero for some utilization profiles in Alabama, Arkansas and South Carolina. In fact, there was a higher probability of yet another hospitalization or Emergency Department encounter, with a probability as high as 0.58 in some utilization profiles in Arkansas and 0.59 in Mississippi. Most utilization profiles with high probability of regular controller asthma medication do not have severe outcomes with probability higher than 0.05, an indication that asthma may be controlled for children in these clusters. This is consistent with findings from other existing research, which found that the use of controller medications over time was greater in the patients who had controlled asthma [36, 47, 48]. Moreover, across all states we found that the regular physician office visits have low probability, consistent with prior research [49].

Limitation

One shortcoming of this study is reliance on claims data to infer utilization. First, the MAX files only include claims that have been submitted for reimbursement. Second, many Medicaid-eligible children have intermittent enrollment and thus some utilization events may be missing; during unenrollment, it is expected that the utilization of emergency department is higher thus our study may underestimate the utilization for this type of care. Estimates on the healthcare utilization are also likely to be biased for certain subgroups of the Medicaid-enrolled children who have difficulty in maintaining Medicaid coverage or are susceptible to particularly disparate utilization [50, 51]. We also consider only asthma events with a primary diagnosis; it is possible that we would observe more physician office visits if

we would consider those with a secondary diagnosis also. Moreover, the medication events lack specificity in terms of number of days supplied and the timing of prescription filled versus when the medication is taken; such specificity would improve the interarrival times between consecutive medication fills and between medication with other events but it will not change the transition probabilities to the extent that will change our findings.

Another limitation is that there will be Medicaid-enrolled children who are undiagnosed due primarily to lack of healthcare access. We also may not capture all children with persistent asthma since the criteria for persistent asthma may be conservative.

Last, Medicaid MAX files can have data quality issues for the encounter data, especially for states with large populations on managed care [52, 53]. The only states with enough patients in comprehensive managed care (CMC) and hence with encounter data to be assessed were Florida, Georgia, South Carolina and Texas; among these states South Carolina did not provide enough encounter data for assessment and Florida was identified to not meet the completeness and quality thresholds by this study for years 2008–2009. Thus the results derived from this study for South Carolina and Florida are not as reliable as for other states.

While our model and its estimation and selection methods are computationally attractive, they can be extended further by relaxing some of the underlying assumptions. First, we do not include the mean times until the first event and the mean times between the last event because they are biased estimates of complete lifetimes due to the censored nature of our data. Therefore, we are unable to completely determine the consistency with which patients visit providers or take medication. For instance, with unbiased estimates of the arrival to the first event it would be clear if a patient waits a long time between groups of consecutive visits or utilizes the system at a fairly homogeneous rate across the complete study time span.

More importantly, it is likely that covariates including age, condition severity, comorbidities, enrollment status and access play a role in the frequency of the visits. However, this method does not capture the potential effects of these covariates on utilization.

Conclusions

Even though this study has several limitations, it has some important implications for health care providers and policy makers. In most profiles, children who take medication do so regularly. While there are children with persistent asthma who primarily rely on short-term medication, we also find multiple fold higher utilization of asthma controller medication versus short-term medication. Recommended care guidelines specify follow-up physician office visits after an emergency room encounter or hospitalization; however we find low utilization of physician office after severe outcomes. Finally, all states have a proportion of children who have uncontrolled asthma, meaning they do not take controller medication while they do have severe outcomes; the proportion varies significantly from one state to another with Texas having a very small proportion and South Carolina having the highest.

Other Research Directions

The results from this study can be applied to identify factors impacting the utilization profile of an asthma-diagnosed child, for example, socio-economic environment (zip code unemployment) and child's characteristics (e.g. race, ethnicity, clinical risk). A second direction is to simulate utilization profiles and evaluate the impact of interventions for reducing severe outcomes on healthcare cost for the Medicaid system. Such analyses can provide insights on the impact of health policy and interventions on reducing severe health outcomes for children diagnosed with asthma for the 10 states in this study.

While this study focuses on pediatric asthma, the data science framework for extracting, analysing and integrating large, highly-sensitive claims information for deriving visual interpretations of healthcare utilization applies generally to other conditions and sub-populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: Financial support for this study was provided in part by the a QuickWin grant by Children's Healthcare of Atlanta and the Institute of People and Technology at Georgia Tech. Research reported in this publication was also supported in part by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number R56HL126761. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, Children's Healthcare of Atlanta or the Institute of People and Technology. The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

The authors are thankful to Matt Sanders and Richard Starr in assisting with data safeguards and the information technology infrastructure. The authors are thankful to Dr. Julie Swann for the leadership in the protocol submission of the use of the MAX Medicaid claims data to the Centers of Medicare and Medicaid and in the Internal Review Board (IRB) approval process. The authors are also thankful to Dr. Jean O'Connor and Francesca Lopez for sharing the list of asthma controller medication used by the Georgia Department of Public Health.

References

1. Braunstein, ML. Health informatics in the cloud. Springer; 2012.
2. Centers for Disease Prevention and Control. Asthma: State Data Profiles (2011). 2011.
3. Centers for Disease Control and Prevention: National Center for Health Statistics. National Health Interview Study Raw Data. 2011.
4. Bacharier LB, et al. Classifying Asthma Severity in Children: Mismatch Between Symptoms, Medication Use, and Lung Function. American Journal of Respiratory and Critical Care Medicine. 2004;170.
5. Gottlieb DJ, O'Connor GT. Poverty, race, and medication use are correlates of asthma hospitalization rates: a small area analysis in Boston. CHEST Journal. 1995; 108(1):28–35.
6. National Heart, L., and Blood Institute. How Is Asthma Treated and Controlled?. [cited 2015; <http://www.nhlbi.nih.gov/health/health-topics/topics/asthma/treatment>]
7. Barnes PJ, Jonsson B, Klim JB. The costs of asthma. European Respiratory Journal. 1996
8. Bateman ED. The economic burden of uncontrolled asthma across Europe and the Asia-Pacific region: can we afford to not control asthma? European Respiratory Review. 2006
9. Juniper EF, Wisniewski ME, Cox FM. Relationship between quality of life and clinical status in asthma: a factor analysis. The European Respiratory Journal. 2004

10. McGrady, Meghan E., Hommel, Kevin A. Medication Adherence and Health Care Utilization in Pediatric Chronic Illness: A Systematic Review. *Pediatrics*. 2013; 132(4):730–740. [PubMed: 23999953]
11. Centers for Disease Prevention and Control. The 6|18 Initiative: Accelerating Evidence into Action. 2015.
12. Cooper RA, et al. Poverty, Wealth, and Health Care Utilization: A Geographic Assessment. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*. 2012; 89(5):828–847. [PubMed: 22566148]
13. Chang J, et al. Comparisons of Health Care Utilization Outcomes in Children With Asthma Enrolled in Private Insurance Plans Versus Medicaid. *Journal of Pediatric Health Care*. 2014; 28(1):71–79. [PubMed: 23312366]
14. Chen L, et al. Health care resource utilisation in primary care prior to and after a diagnosis of Alzheimer's disease: a retrospective, matched case-control study in the United Kingdom. *Bmc Geriatrics*. 2014;14. [PubMed: 24484314]
15. Chomi EN, et al. Health care seeking behaviour and utilisation in a multiple health insurance system: does insurance affiliation matter? *International Journal for Equity in Health*. 2014;13. [PubMed: 24502318]
16. Christian B, et al. Oral health care services utilisation in the adult US population: Medical Expenditure Panel Survey 2006. *Community Dental Health*. 2013; 30(3):161–167. [PubMed: 24151790]
17. Garg R, et al. Gout-Related Health Care Utilization in US Emergency Departments, 2006 Through 2008. *Arthritis Care & Research*. 2013; 65(4):571–577.
18. Harrison PL, et al. Evaluation of the Relationship Between Individual Well-Being and Future Health Care Utilization and Cost. *Population Health Management*. 2012; 15(6):325–330. [PubMed: 22356589]
19. Manski RJ, et al. Patterns of Older Americans' Health Care Utilization Over Time. *American Journal of Public Health*. 2013; 103(7):1314–1324. [PubMed: 23678922]
20. Peres H, et al. Understanding Utilization of Outpatient Clinics for Children with Special Health Care Needs in Southern Israel. *Maternal and Child Health Journal*. 2014; 18(8):1831–1845. [PubMed: 24414986]
21. van Oostrom SH, et al. Multimorbidity of chronic diseases and health care utilization in general practice. *Bmc Family Practice*. 2014;15. [PubMed: 24456299]
22. Yunus, NM., et al. The Determinants of Health Care Utilization among Rural Communities. In: Soliman, KS., editor. *Vision 2020: Innovation, Development Sustainability, and Economic Growth*. Vol. 1–3. 2013. p. 640–645.
23. Lau JS, et al. Young Adults' Health Care Utilization and Expenditures Prior to the Affordable Care Act. *Journal of Adolescent Health*. 2014; 54(6):663–671. [PubMed: 24702839]
24. Pylypchuk Y, Sarpong EM. Comparison of Health Care Utilization: United States versus Canada. *Health Services Research*. 2013; 48(2):560–581. [PubMed: 23003340]
25. Shen C. Determinants Of Health Care Decisions: Insurance, Utilization, And Expenditures. *Review of Economics and Statistics*. 2013; 95(1):142–153.
26. Graves JM, et al. Health Care Utilization and Costs Associated with Adherence to Clinical Practice Guidelines for Early Magnetic Resonance Imaging among Workers with Acute Occupational Low Back Pain. *Health Services Research*. 2014; 49(2):645–665. [PubMed: 23910019]
27. Huber CA, et al. Improved prediction of medical expenditures and health care utilization using an updated chronic disease score and claims data. *Journal of Clinical Epidemiology*. 2013; 66(10): 1118–1127. [PubMed: 23845184]
28. Locklear JC, et al. Treatment Patterns, Health Care Resource Utilization and Costs in Patients with Bipolar Disorder, Newly Treated with Extended Release or Immediate Release Quetiapine Fumarate using US Healthcare Administrative Claims Data. *Clinical Therapeutics*. 2013; 35(12): 1923–1932. [PubMed: 24275622]
29. Stephenson JJ, et al. Patients As Reporters Of Health Care Utilization: An Analysis Of All-Cause And Disease-Specific Patient-Reported Health Care Utilization Compared To Administrative Claims Data. *Value in Health*. 2013; 16(3):A238–A238.

30. Teitelbaum A, et al. Health Care Costs and Resource Utilization, Including Patient Burden, Associated With Novel-Agent-Based Treatment Versus Other Therapies for Multiple Myeloma: Findings Using Real-World Claims Data. *Oncologist*. 2013; 18(1):37–45. [PubMed: 23299776]
31. Wiley-Exley E, et al. The Impact of Assertive Community Treatment on Utilization of Primary Care and Other Outpatient Health Services: The North Carolina Experience. *Journal of the American Psychiatric Nurses Association*. 2013; 19(4):195–204. [PubMed: 23824135]
32. Zhao Z, et al. Health Care Costs And Resource Utilization In Working Age Patients With High Risk Vascular Disease: Findings From A Multi-Employer Claims Database. *Value in Health*. 2014; 17(3):A111–A111.
33. Heintzman J, et al. Using electronic health record data to evaluate preventive service utilization among uninsured safety net patients. *Preventive Medicine*. 2014; 67:306–310. [PubMed: 25124279]
34. James T, Fine M. Monitoring Asthma Control Using Claims Data And Patient-Reported Outcomes Measures. *Pharmacy and Therapeutics*. 2008; 33(8):454. [PubMed: 19750176]
35. Leone FT, et al. Pharmaceutically-based severity stratification of an asthmatic population. *Respiratory medicine*. 1999; 93(2):788–93. [PubMed: 10603627]
36. Berger WE, et al. The Utility of the Health Plan Employer Data and Information Set (HEDIS) Asthma Measure to Predict Asthma-related Outcomes. *Annals of Allergy, Asthma & Immunology*. 2004; 93(6):538–545.
37. Schatz M, et al. Asthma Quality-of-care Markers Using Administrative Data. *CHEST Journal*. 2005; 128(4):1968–1973.
38. Jacob, Christian, et al. Assessing Asthma Severity Based on Claims Data: A Systematic Review. *European Journal of Health Economics*. 2016:1–15.
39. Ramirez de Arellano, A., Wolfe, S. Unsettling Scores: A Ranking of State Medicaid Programs. 2007. 2011.
40. Pearson, William S., et al. State-Based Medicaid Costs for Pediatric Asthma Emergency Department Visits. *Preventing Chronic Diseases*. 2014; 11:140139.
41. Kuo, Dennis Z., et al. Comparison of health care spending and utilization among children with Medicaid insurance. *Pediatrics*. 2015; 136(6)
42. Wakefield, Dorothy B., Cloutier, Michelle M. Modifications to HEDIS and CSTE algorithms improve case recognition of pediatric asthma. *Pediatric Pulmonology*. 2006; 41(10):962–71. [PubMed: 16871628]
43. National Committee for Quality Assurance. Improving Outcomes in Asthma: Advancing Quality Using NCQA HEDIS Measures. 2011.
44. Foufoula-Georgiou E, Lettenmaier DP. A Markov renewal model for rainfall occurrences. *Water Resources Research*. 1987; 23(5):875–884.
45. National Heart Blood and Lung Institute. National Institute of Health. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. 2007.
46. National Asthma Education and Prevention Program, Third Expert Panel on the Diagnosis and Management of Asthma. *Clinical Practice Guidelines*. 2007.
47. Stempel DA, et al. Patterns of asthma control: a 3-year analysis of patient claims. *Journal of Allergy Clinical Immunology*. 2005; 115(5):935–9. [PubMed: 15867848]
48. Piccoro LT, et al. Asthma prevalence, cost, and adherence with expert guidelines on the utilization of health care services and costs in a state Medicaid population. *Health Services Research*. 2001; 36(2):357–371. [PubMed: 11409817]
49. Mintz M, et al. Assessment of asthma control in primary care. *Curr Med Res Opin*. 2009; 20(10): 2523–31.
50. DeVoe JE, et al. Electronic health records vs Medicaid claims: completeness of diabetes preventive care data in community health centers. *The Annals of Family Medicine*. 2011; 9(4):351–358. [PubMed: 21747107]
51. Piccoro LT, et al. Asthma prevalence, cost, and adherence with expert guidelines on the utilization of health care services and costs in a state Medicaid population. *Health Services Research*. 2001; 36(2):357–371. [PubMed: 11409817]

52. Byrd VL, Dodd AH. Assessing the usability of MAX 2008 encounter data for comprehensive managed care. *Medicare Medicaid Res Rev.* 2013; 3(1)
53. Byrd, Vivian LH., Dodd, Allison Hedley. Assessing the Usability of Encounter Data for Enrollees in Comprehensive Managed Care 2010–2011. Centers for Medicare and Medicaid Services: Mathematica Policy Research; 2015.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

Overall utilization summary for each of the 10 states, sorted by the total number of patients considered for each state. The average number of events, as well as averages of each event types(ACM, ASM, ER, HO, PO) are per member-year.

Table 1

State	# Patients	avgEvents	avgACM	avgASM	avgER	avgHOS	avgPO
MS	40,147	2.58	1.90	0.26	0.08	0.09	0.25
SC	65,175	2.78	1.95	0.37	0.06	0.08	0.32
AR	71,369	2.99	2.07	0.51	0.00	0.10	0.31
AL	104,531	3.05	2.08	0.47	0.04	0.09	0.37
FL	122,667	2.5	1.54	0.32	0.06	0.05	0.53
TN	137,148	2.84	2.04	0.38	0.05	0.06	0.31
GA	137,519	2.57	1.54	0.50	0.06	0.09	0.38
LA	142,608	2.36	1.63	0.34	0.07	0.05	0.26
NC	157,011	3.22	2.17	0.48	0.06	0.08	0.43
TX	476,345	2.44	1.66	0.40	0.04	0.03	0.31
AVG		2.74	1.86	0.40	0.05	0.07	0.35
STD		0.28	0.24	0.08	0.02	0.02	0.08

Table 2

Characterization of utilization profiles for each state along with the percentage of patients for each cluster in paranthesis.

State	# Enroll Years	Profile 1 Pop %	Profile 2 Pop %	Profile 3 Pop %	Profile 4 Pop %	Profile 5 Pop %
<i>AL</i>	637,639	ACM<--ASM + (PO) 80.7%	ACM<--ASM + (PO, HOS) 4.3%	ASM + (PO, HOS) 8.2%	ASM + (PO, HOS, ER) 6.8%	
<i>AR</i>	449,625	ACM<--ASM + (PO) 35.7%	ACM 24.9%	ACM/ASM + (PO) 20.6%	ASM + (HOS) 5.8%	ACM/ASM + (PO, HOS) 13%
<i>FL</i>	718,829	ACM/ASM + (PO) 32%	ACM<--ASM + (PO) 61.6%	ACM + (PO, ER, HOS) 6.4%		
<i>GA</i>	805,861	ACM<--ASM + (PO, HOS) 72.2%	ASM + (PO) 13.5%	+ (PO) 2.4%	ASM + (PO, ER, HOS) 11.9%	
<i>LA</i>	944,065	ACM 5.5%	ACM<--ASM + (PO) 14.8%	ACM/ASM + (PO) 6.1%	ACM<--ASM + (PO, ER) 60.3%	ASM + (PO, ER, H OS) 13.4%
<i>MS</i>	240,882	ACM + (PO) 8.2%	ACM/ASM + (PO) 5.5%	ACM<--ASM + (PO, HOS) 63%	ASM + (PO, ER, HOS) 11.2%	ASM + (PO, ER, H OS) 0.6%
<i>NC</i>	954,627	ACM<--ASM + (PO) 89.7%	+ (PO, ER, HOS) 2.6%	ASM + (PO, ER, HOS) 7.6%		
<i>SC</i>	410,603	ACM<--ASM + (PO) 70.1%	ASM + (PO) 8.2%	ASM + (PO, ER, HOS) 4%	ASM + (PO, ER, HOS) 14%	
<i>TN</i>	877,747	ACM (8%)	ACM<--ASM + (PO) 64%	ASM + (PO) 3.1%	ASM + (PO, ER, HOS) 6.5%	ACM<-- ASM + (PO, HOS) 18.4%
<i>TX</i>	2,810,436	ACM (8.6%)	ACM + (PO) 10%	ACM<--ASM + (PO) 71.6%	ACM/ASM + (PO) 5.1%	ASM + (PO, ER, HOS) 4.7%

Table 3

The range (minimum and maximum) of probabilities for different links between events across the utilization profiles of all the states.

State	ER/HOS → PO	ER/HOS → ER/HOS	PO → ACM	PO → ASM	ER/HOS → ACM	ER/HOS → ASM
AL	(0.00, 0.19)	(0.07, 0.34)	(0.45, 0.47)	(0.00, 0.28)	(0.32, 0.32)	(0.05, 0.51)
AR	(0.00, 0.11)	(0.00, 0.58)	(0.21, 0.46)	(0.14, 0.26)	(0.00, 0.00)	(0.00, 0.33)
FL	(0.23, 0.24)	(0.06, 0.26)	(0.23, 0.44)	(0.13, 0.18)	(0.14, 0.19)	(0.00, 0.00)
GA	(0.14, 0.18)	(0.00, 0.34)	(0.11, 0.40)	(0.11, 0.20)	(0.28, 0.28)	(0.12, 0.21)
LA	(0.12, 0.13)	(0.08, 0.25)	(0.05, 0.47)	(0.10, 0.24)	(0.05, 0.41)	(0.13, 0.21)
MS	(0.03, 0.13)	(0.08, 0.59)	(0.08, 0.51)	(0.09, 0.23)	(0.07, 0.40)	(0.05, 0.21)
NC	(0.11, 0.24)	(0.09, 0.36)	(0.41, 0.41)	(0.14, 0.15)	(0.00, 0.00)	(0.13, 0.17)
SC	(0.00, 0.19)	(0.00, 0.50)	(0.03, 0.52)	(0.05, 0.20)	(0.09, 0.09)	(0.18, 0.23)
TN	(0.16, 0.38)	(0.00, 0.21)	(0.42, 0.42)	(0.01, 0.42)	(0.22, 0.22)	(0.00, 0.20)
TX	(0.11, 0.12)	(0.00, 0.20)	(0.02, 0.47)	(0.14, 0.29)	(0.00, 0.00)	(0.24, 0.31)