Identifying Diagnostic Paths for Undifferentiated Abdominal Pain from Electronic Health Record Data

Yiye Zhang, PhD, MS¹, Rema Padman, PhD², Paul Epner, MBA, MEd³, Victoria Bauer, BS⁴, Anthony Solomonides, MSc⁴, Goutham Rao, MD⁵
¹Weill Cornell Medical College, New York, NY; ²Carnegie Mellon University, Pittsburgh, PA; ³Society to Improve Diagnosis in Medicine; ⁴NorthShore University Health System, Evanston, IL; ⁵Case Western Reserve University School of Medicine, Cleveland, OH

Abstract
The diagnostic process is a complex, uncertain, and highly variable process which is under-studied and lacks evidence from randomized clinical trials. This study used a novel visual analytics method to identify and visualize diagnostic paths for undifferentiated abdominal pain, by leveraging electronic health record (EHR) data of 501 patients in the ambulatory setting of a single institution. A total of 63 patients reached diagnoses in the study sample. We illustrate steps in identifying diagnostic paths of the study patients, both individually and collectively, and visually present the diversity in their diagnostic processes. Patients in whom diagnoses were obtained generally had more clinical encounters and health services utilization, although their diagnostic paths were more variable than those of the undiagnosed group. The capability of identifying diagnostic paths demonstrated from this study allows us to study larger data sets to determine diagnostic paths associated with more timely, accurate, and cost-efficient diagnosis processes.

Introduction
Diagnosis is an inherently complex, uncertain, and highly variable process¹. Variation in diagnostic approaches can be attributed to different patient characteristics, different disease presentations, and system-related factors such as the availability of specific tests². A crucial part of the variation, however, is associated with differing physician heuristics, experiences, and knowledge². Similar to widespread variation in treatment patterns, which is frequently associated with high costs and suboptimal outcomes, variation in diagnostic processes is often associated with delayed or incorrect diagnosis, unnecessary expense, over and under treatment, and genuine harm³. Diagnostic error is the top reason for malpractice claims in the outpatient setting. Roughly 12 million American adults are affected by diagnostic errors annually in ambulatory settings alone⁴. A 2006 study of claims involving missed or delayed diagnosis revealed that 59% of diagnostic errors were associated with serious harm and 30% were associated with death⁵. The root causes of diagnostic error are complex and multi-factorial, and many errors are the result of both system-related and cognitive factors⁶. In an analysis of 100 cases of diagnostic errors, cognitive factors, such as faulty knowledge, and premature closure contributed to diagnostic error in 74% of cases⁶. A more recent study by Zwann et al. of 7,926 patient records revealed that human causes, especially faulty knowledge, contributed to 96.3% of cases where a diagnostic adverse event took place⁶. Among the greatest diagnostic challenges are “undiagnosed” complaints, which are non-specific symptoms that have not yet manifested into identifiable illnesses⁷. These include undiagnosed abdominal pain, dizziness, fatigue, and fever without an obvious source⁸. Undiagnosed complaints are associated with especially significant variation in physician diagnostic practices. A survey of primary care physicians found that undiagnosed complaints were the presenting complaint in 64% of cases in which physicians recalled making a diagnostic error⁹.

Guidance for systematic approaches to diagnosis, especially in ambulatory settings, is scarce and often of poor quality⁹. By contrast, the majority of treatment recommendations are based on randomized trials and high quality systematic reviews¹⁰. Relatively few clinical practice guidelines address diagnosis specifically¹¹. Diagnostic recommendations in guidelines that address diagnosis and evaluation alone or diagnosis and treatment, are often based on expert consensus or similar weak levels of evidence¹¹. For example, the majority of recommendations in a recently updated guideline for fever of uncertain source in infants 60 days of age or less, for example, were based on either weak level of evidence or consensus opinion only⁹. Therefore, there is a significant gap in knowledge and evidence for the optimal diagnostic strategies, which have major consequences on early identification of diseases and proper referrals to specialists.
In order to identify diagnostic steps that lead to more timely diagnosis, greater accuracy, and more efficient use of diagnostic resources, we must obtain a clear understanding of the current practices. We previously defined a diagnostic path to be “the steps taken for diagnostic evaluation of a complaint from initial presentation until either a diagnosis is achieved or the patient and/or physician choose to end the evaluation without obtaining a diagnosis.” A path begins with a clinical evaluation, which includes a detailed history and in most cases, a physical examination. This is followed by any number of possible steps, including laboratory or imaging studies, referrals, observation, follow-up visits, or a trial of therapy. A referral may result in additional steps being added to the path. In this study, we define “reaching a diagnosis” as having a diagnosis recorded in the medical record on a minimum of 2 encounters for the complaint in question, and which becomes the basis for further management. A diagnosis must also be different from the chief complaint. For instance, “low back pain” may be a chief complaint, and may also be used as an interim diagnosis by the physician. However, the diagnosis must be more specific than “low back pain” and could include, for example, “lumbar disc herniation.” Alternatively, the diagnostic path may end if the patient or physician chooses to stop the evaluation without achieving a stable diagnosis. This may occur because the patient’s symptom resolves prior to receiving a diagnosis. The patient may no longer seek evaluation for the complaint, even though the symptom may continue to persist, or the patient may be lost to follow-up. The physician may recommend not pursuing the evaluation for a number of reasons such as harmful further diagnostic testing and initiate empiric treatment instead. Thus, a path may be brief, or lengthy and complex, depending on individual cases.

Understanding diagnostic paths for an individual patient or for a large number of patients in a way that is meaningful to clinicians presents a significant challenge. Previously, the only method to identify diagnostic paths was either prospectively, by observing and recording steps taken as patients go through the diagnostic process or through analysis of paper-based charts, which is cumbersome and imprecise. With the availability of Electronic health records (EHRs), we now are presented with an extremely useful opportunity to identify and analyze diagnostic paths using data-driven approaches. Yet, EHR data are heterogeneous, complex, and usually not collected for large-scale analysis of decision making. To address this challenge, we applied process mining techniques combined with visual analytics, which is the combination of advanced statistical analytics and visualization techniques for evidence discovery from data. We believe analysis based on visualization of large number of paths can provide insights that are complementary to a purely quantitative approach.

Process mining, originated from business process management, has been used in recent years across clinical domains to identify time-series patterns from data. Visual analytics, when applied to individual diagnostic paths or paths from thousands of patients, also has promoted easier consumption of data and knowledge discovery such as EventFlow and Harvest. One common challenge from previous literature on process mining and visual analytics has been the complexity due to events that share the same timestamps. In this paper, we apply a previously developed methodology, which overcomes such complexity, to the construction of diagnostic paths from time stamped EHR data. This methodology has been applied to investigate patterns of care for patients with chronic kidney disease and left-ventricular assist device (LVAD) implantation. We aim to show that this methodology, with appropriate technical refinements, can provide a visual representation of the common diagnostic paths of a cohort, as well as an individual diagnostic path of a patient. Both can be highly informative for clinicians and patients in understanding the diagnostic process. Given the scarcity of the current research into best diagnostic practices, this preliminary study can potentially lead to future research that shift current research and clinical practice paradigms, and establish the foundation for a new type of diagnostic guidance for these types of common, challenging, undiagnosed complaints. To the best of our knowledge, this work is the first to aim to identify and learn diagnostic paths from EHR data.

This paper is organized as follows. Details of the visual analytics methodology and its application to diagnostic path are described in Methods. Data section describes the patient sample characteristics. Preliminary insights and illustration of diagnostic paths are presented in Results. Finally, we conclude with study limitations, extensions, and conclusions in the Discussion section.
Method
There are 3 essential steps in visualizing diagnostic path from EHR data through visual analytics: transformation of the data, construction of common paths, and visualization. With these steps, we will instantiate the abstract flow diagram of the diagnostic process by defining key activities during encounters to generate event sequences representing each patient’s actual encounters.

Data Transformation: The purpose of this step is to reduce the complexity of the data documenting different aspects of the diagnostic process in an ambulatory setting, such as having multiple clinical events documented with the same-day time-stamp. The complexity in data is addressed by developing a novel representation for the clinical data structure, shown in Figure 1, that summarizes the co-progression of all information into a one-dimensional path of chronologically sequenced events for each patient. The flexibility of the methodology allows these data categories to be altered and expanded depending on the availability of appropriate data and variables. Data transformation is applied to all study patients’ records, and each patient’s data is summarized into one and only one path. This step prepares the data for the application of a variety of existing sequential data analytic methods, which is a crucial part of visual analytics.

![Figure 1. Schematic view of methodology: transformation of data for one patient](image)

Construction of common paths: A significant challenge is the heterogeneity or “noisiness” of the underlying data. Identification of diagnostic patterns, combined with research team clinical experts’ input, is an iterative process that is repeated many times to determine which variables and factors to include in the paths. To begin with, we compare patients with similar paths to understand the characteristics of the paths, as well as to detect if there are any differences among the underlying patient population. Similarity among paths is measured by computing a metric called the longest common subsequence (LCS) distance. LCS is the longest subsequence that 2 sequences have in common, while preserving the order of occurrence, but possibly separated. We believe that LCS is a reasonable metric for measuring similarity of diagnostic paths because it gauges similarity by the overall direction or flow of a path, rather than only by the initial or ending points. In this paper, we use sequence and path interchangeably.

For instance, if patient 1 has a sequence:

\[
[\text{clinical evaluation}] \rightarrow [\text{clinical evaluation + PPI}] \rightarrow [\text{clinical evaluation + referral to GI}] \rightarrow [\text{clinical evaluation + EGD}]
\]

and, patient 2 has a sequence:

\[
[\text{clinical evaluation}] \rightarrow [\text{clinical evaluation + referral to GI}] \rightarrow [\text{clinical evaluation}]
\]

then, their LCS is of length 2, being:

\[
[\text{clinical evaluation}] \rightarrow [\text{clinical evaluation + referral to GI}]
\]

. Then their LCS distance (dLCS) is:

\[
dLCS = |\text{sequence 1}| + |\text{sequence 2}| - 2 \times \text{LCS} = 4 + 3 - 2 \times 2 = 3
\]

Subsequently, in order to identify common patterns in the diagnostic process, we first elicit all transitions seen in the diagnostic paths of all patients. This is performed for each subgroup identified earlier. In a diagnostic path, we refer to an encounter occurring first as source and the successor as target. For example, given a path with encounter
identifier \([\text{clinical \ evaluation}\rightarrow \text{clinical \ evaluation} + \text{PPI}]\rightarrow \text{clinical \ evaluation} + \text{referral \ to \ GI}]\rightarrow \text{clinical \ evaluation}\) \(\rightarrow \text{clinical \ evaluation} + \text{EGD}\), there are 4 transitions: \((\text{clinical \ evaluation}\rightarrow \text{clinical \ evaluation} + \text{PPI})\), \((\text{clinical \ evaluation} + \text{PPI}\rightarrow \text{clinical \ evaluation} + \text{referral \ to \ GI})\), \((\text{clinical \ evaluation} + \text{referral \ to \ GI}\rightarrow \text{clinical \ evaluation})\), and \((\text{clinical \ evaluation}\rightarrow \text{clinical \ evaluation} + \text{EGD})\). In the first transition, the source is \([\text{clinical \ evaluation}]\) and target is \([\text{clinical \ evaluation} + \text{PPI}]\). In order to ensure that common patterns occur in the data beyond a certain probability, we compute the conditional probability of transitioning into a target given each fixed source, defined as

\[
\text{weight} = \frac{\text{[target = T | source = S]}}{\text{[source = S]}}
\]

where \([\text{source = S}]\) is the number of times \(S\) appears as a source, and \([\text{target = T | source = S}]\) is the number of times \(T\) appears as a target given \(S\) is the source. In the sequence used as example above, \((\text{[clinical \ evaluation}\rightarrow \text{clinical \ evaluation} + \text{PPI})\) and \((\text{[clinical \ evaluation}\rightarrow \text{clinical \ evaluation} + \text{EGD})\) have a weight of 0.5, respectively, because they both have \([\text{clinical \ evaluation}]\) as the source, but 2 different targets. The other 2 transitions in the sequence above both have a weight of 1. In addition to weight, edge frequency is another threshold for filtering, to ensure that transitions that occur in single patient, and therefore have high weight, are appropriately adjusted. While the example was using 1 patient’s path, with a large number of patients, we expect that same transitions will appear more than once across patients, hence providing a measure for common and meaningful transitions. Thresholds for dominant patterns will be determined depending on the volume of patient data and its variations. Thus, these procedures allow efficient extraction of the key patterns in diagnostic paths by filtering out events that are due to data error or out of scope of the research questions. We discuss the possibility of filtering out rare events in the Discussion section.

In addition to the weights mentioned earlier, to define the duration between encounters in visualized diagnostic paths, we capture the difference in days between source and target as \(\text{time \ delta}\). For example, in Figure 1, the time delta between the first two encounters (2/1/2016 and 2/3/2016) is 2 days. If the same transition of visits has more than one time gap observed in multiple paths, such as the case of \((\text{[clinical \ evaluation}\rightarrow \text{clinical \ evaluation} + \text{PPI})\) occurring in 10 patients’ diagnostic paths but with different value of time delta, we take the average of all in \(\text{days}\).

Visualization: Visualizations are prepared by taking the common transitions extracted using methods from above. Diagnostic paths are represented in graphical form using \(\text{nodes}\) and \(\text{edges}\). Nodes represent single encounters, and edges represent continuation of one encounter to the next. We adjust for path characteristics such as encounter type and frequency using colors and sizes of nodes and edges, described in detail as we present the results in the Results section. We used Gephi 0.9.1\textsuperscript{24}, an open-source software for graph and network analysis that uses a 3D render engine to display large networks. The diagnostic paths were plotted first using Gephi’s built-in ForceAtlas2 algorithm, which is scaled for small to medium-size networks and suitable for qualitative interpretation\textsuperscript{25}. Following the layout algorithm, we manually adjust the path visualization to clarify the start of all diagnostic paths to various routes before diagnoses. Due to the high variability and thus large number of nodes expected from our data, we chose to use network graphs over other visualization format such as Sankey diagram\textsuperscript{26} and algorithms such as LifeFlow\textsuperscript{27} for its compactness and interpretability.

**Data**

**Table 1. Characteristics of patients in the diagnosed and undiagnosed groups**

| Characteristic | Diagnosis (N=63) | No Diagnosis (N=438) | P-value for significance of difference across 2 groups |
|---------------|-----------------|----------------------|--------------------------------------------------------|
| Treatment duration in days (mean, SD) | 585.8±239.45 | 507.3±290.13 | 0.4769 |

Demographics

| Characteristic | Diagnosis | No Diagnosis | P-value |
|---------------|-----------|--------------|---------|
| Age (mean, SD) | 53.5±16.95 | 51.4±15.88 | 0.332 |
| BMI (mean, SD) | 27.7±5.51 | 27.6±5.57 | 0.897 |
| Sex – Female (%) | 49.2% | 50.0% | 1.00 |

Race

| Ethnicity | Diagnosis % | No Diagnosis % | P-value |
|-----------|-------------|----------------|---------|
| African American (%) | 1.6% | 3.2% | 0.900 |
| Asian (%) | 3.2% | 5.5% | |
| White Caucasian (%) | 66.7% | 63.7% | |
| Other (%) | 28.6% | 27.6% | |

Ethnicity
We obtained EHR (Epic Systems, Verona, WI) data from ambulatory settings from January 2010 to December 2012. We used a small database of records from 501 adult patients with abdominal pain from a single institution in Illinois. All patients presented a chief complaint of new onset of abdominal pain, stomach pain, or epigastric pain, and their data were extracted from the available structured chief complaint fields in the EHR. Based on consensus of the research team, only structured field data relevant to abdominal pain was extracted, including the most common patterns of care: referrals, orders including diagnostic tests and procedures, medication prescriptions, follow-up intervals, and International Classification of Diseases (ICD)-9 codes for diagnoses associated with abdominal pain. Table 1 describes the characteristics of patients and treatments in the study data by patients who received diagnoses and those who did not. We applied t-tests to compare the significance of differences across the 2 groups and listed the p-values for each characteristic. No significant differences were observed.

Table 2. Data element included in the diagnostic paths

| Category (N) | Classes (N) |
|--------------|-------------|
| Referral (19) | Addiction (1), Allergy (14), Cardiology (31), Endocrinology (18), Gastroenterology (113), Gynecologic oncology (92), Hematology (11), Nephrology (14), Neurology (42), Neurosurgery (9), Nutrition (10), Oncology (17), Otolaryngology (49), Pain (8), Psychiatry/Psychology (21), Pulmonary (17), Surgery (53), Radiation Oncology (17) |
| Order (23) | Albumin (3), Allergen-wheat (2), Bilirubin (1), C. Difficile (43), CBC (1134), Celiac (3), Chemistry (1174), Colonoscopy (95), C-reactive protein (16), E. Coli (2), EGD (53), Fecal Blood (7), Hepatitis (140), H. Pylori (69), Imaging (317), Lipase (124), Pathology (10), Nutrition consult (3), Stool (43), Urinalysis (535), Urology consult (3) |
| Medication (10) | Analgesics (892), Anti-infective agents (892), Cardiovascular agents, Central Nervous System Drugs (1111), Endocrine & Metabolic Drugs (863), Gastrointestinal Agents (634), Genitourinary Products (610), Nutritional Products (165), Respiratory Agents (522), Others (1449) |
| Diagnosis (noted at least twice) | Anal fissure and fistula (2), Calculus of kidney and ureter (5), Cholelithiasis (4), Diverticulitis of colon (4), Dyspepsia and other specified disorders of function of stomach (3), Esophageal reflux (5), Intestinal infection due to clostridium difficile (2), Unspecified gastritis and gastroduodenitis (2), Other disorders of urethra and urinary tract (14) |

Due to the limited size of the study data, data elements were summarized into related classes by the research team. For example, both ‘US Transvaginal Screen and ‘Ultrasound Abdomen or Pelvis’ are summarized as imaging orders. We also performed chart reviews to validate the relevance to abdominal pain of specific diagnostic steps. For example, abdominal ultrasound was always relevant to abdominal pain. Table 2 describes data elements in diagnostic paths that were analyzed.

Results

Characteristics of the diagnostic paths in the diagnosed and undiagnosed groups, including encounters before and after reaching diagnoses are shown in Table 3. Using elements listed in Table 2, we found 490 distinct diagnostic paths out of a total of 501. There was a total of 1107 distinct encounters, consisting of different combinations of referrals, orders, medication prescriptions, and diagnoses (or no diagnoses). As the table shows, despite the small size of the data (501 patients), there is tremendous diversity in the diagnostic paths. For example, the average LCS distance is 60.6 and 37.4 in the 2 patient groups, respectively. Larger LCS distance suggests more variable
diagnostic paths, both in terms of encounter content and path length. Therefore, the diagnosed group is more variable compared to the undiagnosed group, and it is also evident from the ranges of path length. The diagnosed group’s path ranges from 4 to 215 encounters, whereas the undiagnosed group’s path ranges from 2 to 146.

Table 3 also lists most frequent orders, referrals, and medications taken by patients in the 2 groups, and the average number of observation per patient. Patients who received diagnoses have had consistently more services across orders, referrals, and medications. The most common 5 orders are the same in both diagnosed and undiagnosed groups, while in different order, and the diagnosed group received larger number of orders compared to the undiagnosed group. For referrals, both Gynecologic oncology and Gastroenterology are among 2 of the most common 3. Due to the smaller size of the referrals, we only listed the 3 most common referrals. The diagnosed group has surgical referral, and undiagnosed group has Otolaryngology referral as the other. Similarly, for medications, diagnosed group has Genitourinary Products, whereas the undiagnosed group has Endocrine & Metabolic Drugs, as one of the most common 5 medications, respectively.

Table 3. Characteristics of the diagnostic paths in the diagnosed and undiagnosed groups, including encounters before and after reaching diagnoses

|                         | Diagnosed (N=63) | Undiagnosed (N=438) |
|-------------------------|------------------|---------------------|
| Number of distinct paths| 63               | 427                 |
| Min, mean, max number of encounters in paths* | 4, 43.2, 215 | 2, 26, 146 |
| Number of unique encounter types | 505            | 936                 |
| Number of unique transitions from one encounter to the next | 1647           | 5215                |
| Average LCS distance (SD) | 57.5 ± 45.03 | 37.4 ± 28.56        |
| Most frequent orders (average number of observations per patient) | CBC (3.9), Chemistry (3.7), Urinalysis (2.3), Imaging (1.4), Hepatitis (0.5) | Chemistry (2.2), CBC (2.0), Urinalysis (0.9), Imaging (0.5), Hepatitis (0.3) |
| Most frequent referrals (average number of observations per patient) | Gynecologic oncology (0.5), Surgery (0.3), Gastroenterology (0.2) | Gastroenterology (0.2), Gynecologic oncology (0.1), Otolaryngology (0.1) |
| Most frequent medications (average number of observations per patient) | Other (4.8), Central Nervous System Drugs (2.9), Genitourinary Products (2.7), Anti-Infective Agents (2.6), Analgesics & Anesthetics (2.6) | Other (2.6), Central Nervous System Drugs (2.1), Endocrine & Metabolic Drugs (1.7), Anti-Infective Agents (1.7), Analgesics & Anesthetics (1.7) |

Figure 2 displays a visualization of the diagnostic paths of all 63 patients containing all transitions of encounters until reaching stable diagnoses, but not after. In the figure, each node, except for the green node, represents an encounter, and labels on the node show actions taken by providers during the encounter, such as placing a medical order or making a referral. We color-coded the node by index visit, encounters after the index visit, and diagnoses. Since there can be multiple events taking place during one encounter, as reflected in the over 500 unique encounter types, we did not color-code nodes beyond the 3 types. The green node signals the start of each diagnostic path and is not an encounter; purple nodes are an encounter without diagnoses; and the orange nodes are encounters where at least one diagnosis was reached. Size of the nodes and labels represent commonality of the encounter represented by the nodes and labels. Edges thickness represents the number of patients who have experienced the transitions; the thicker the edge, the more patients who have experienced the transition. Also, purple edges represent transitions of encounters before reaching diagnoses, and orange edges represent transitions leading to at least one diagnosis. Edge lengths are determined for layout purposes only and are not representative of any characteristic of the paths. While both Table 3 and Figure 2 show the diversity that exists in the data, the diagnostic paths visualized in Figure 2 more clearly show the dramatically diverse processes that patients go through via the differing number of evaluations, from 1 up to 127 encounters, before reaching stable diagnoses.
Figure 2. Diagnostic paths of all 63 patients with abdominal pain who reached diagnoses.

Figure 3. Eliciting only common (experienced by at least 2 patients) diagnostic paths of 63 patients with abdominal pain who reached diagnoses.

On the other hand, Figure 3 shows the common paths encountered by the 63 patients who received at least one diagnosis. Compared to Figure 2, Figure 3 filters out edges that have edge frequency that is fewer than 2, thereby reducing the complexity seen in Figure 2 and allows easier interpretation of the diagnostic paths. Due to the small sample size, we only used frequency as the threshold and not weight. It is much easier to see in Figure 3 that some patients received their diagnosis after completing imaging orders, whereas others received their diagnosis after 11 encounters with only evaluations and no other actions. However, we did not find significant associations among
specific diagnoses and time of encounter, or previous clinical events. Ten patients reached diagnoses at their initial encounters, including Other symptoms involving abdomen and pelvis (3 patients), Esophageal reflux (2 patients), Diverticulitis of colon (2 patients), Other and unspecified noninfectious gastroenteritis and colitis (1 patient), Disorders of menstruation and other abnormal bleeding from female genital tract (1 patient), Infectious diarrhea (1 patient), Dyspepsia and other specified disorders of function of stomach (1 patient), Calculus of gallbladder without mention of cholecystitis (1 patient), Inguinal hernia, without mention of obstruction or gangrene, unilateral or unspecified (1 patient). Diagnoses obtained after 11th encounter with evaluation only include Urinary tract infection (1 patient), Regional enteritis of unspecified site (1 patient). Diagnoses after performing imaging include Calculus of gallbladder without mention of cholecystitis (1 patient), Diverticulitis of colon (1 patient), and Other calculus in bladder (1 patient).

Furthermore, to illustrate temporality, Figure 4 displays simplified diagnostic paths among a subset (56 patients out of 63 patients) who reached diagnoses. The paths are constructed by summarizing events in Table 2 by their categories. We excluded 7 patients from this figure whose paths were distinctly different from others, and whose number of encounters ranged from 9 to 87, which were too long and complex to fit appropriately in the figure. Unlike Figures 2 and 3, the edges in Figure 4 represent relative duration in time between 2 encounters. Time durations are calculated as average time in days when there is more than one patient. Hence, we see that while 10 patients received diagnoses at their first encounter, there is one patient who received a diagnosis 137 days after having medication prescriptions, and another 5 patients who received diagnoses 18 days, on average, after receiving orders and medication prescriptions. The dashed edges refer to omitted encounters to minimize complexity. For example, the diagnostic path at the very top are experienced by 24 patients who received diagnoses after different number of encounters with evaluations only.

Discussion
In this preliminary study, we aimed to demonstrate that diagnostic paths can be learned and visualized from EHR data. We found tremendous amount of diversity in the diagnostic paths for patients who all presented undifferentiated abdominal pain in a single institution’s ambulatory setting. In this paper, due to the small sample size, our goal was not to make any conclusions about practices that lead to, or not lead to diagnoses. Yet, we did find
that patients who received diagnoses received more clinical services including laboratory orders, referrals, and medication prescriptions. We expect that future studies with a much larger sample size will lead to discovery of diagnostic paths which allow us to conclude association of evaluation processes with greater timeliness and accuracy in reaching diagnoses, as well as efficient use of clinical resources.

One limitation of the study is the completeness of the data. For example, 3 patients received diagnoses after receiving an imaging order. However, patients in the undiagnosed group received imaging orders for 231 times without reaching diagnoses, as indicated in Table 3. Our data for this study does not include sufficient information about the results of the imaging orders for us to understand whether some patients failed to receive diagnoses because of the inconclusiveness of the imaging results, or whether diagnoses occurred after our study period. Similarly, we do not have results for the laboratory orders listed in Table 2. Therefore, while it is very likely that diagnoses are made based on laboratory results, we are not able to identify such relationship due to the lack of data on laboratory results. If patients were to visit providers from outside institutions, we also would be missing those encounters. These data limitations may explain why we observed a very wide range of time it took before reaching diagnoses. For example, we found that a patient had 127 encounters with evaluation only and no other actions, before reaching diagnoses. In the undiagnosed group, a patient has had 77 encounters with evaluations only, 18 orders of CBC and Chemistry, and 3 imaging orders, without being able to reach any diagnosis. It is valuable to be able to find patients such as this, whose records likely need detailed review to better understand the reason behind it. When sufficient data are available, rigorously matching patients who received diagnoses versus those who did not, may allow us to discover best practices, as well as exceptions from them, in the diagnostic processes. Moreover, information on providers and patients, such as years of experience, patient demographics, and comorbidities may allow us to classify diagnostic paths by provider and patient types. Furthermore, with longer time period we will identify diagnostic paths which go beyond the first diagnosis to detect changes in diagnoses as more results become available. Advanced analytical techniques for diagnostic paths such as sequential pattern mining can also be applied to generate interesting patterns across subgroups.

Nevertheless, we observed tremendous variability in their diagnostic paths. As shown in Figure 3, filtering out minor events and being able to see the common paths in the population are useful in grasping dominant patterns, but we realize that for specific research questions, low-frequency events need to be evaluated equally to avoid missing important information. Therefore, future work includes developing a visual analytics platform that allows interested users to input EHR from their own practice, and generate diagnostic paths for evaluation, under user-defined thresholds for filtering. More importantly, it is our future aims to understand and reduce the source of variation to only patient-specific factors.

Conclusion
This paper describes a preliminary study using electronic health records (EHRs) data where we, applying a novel visual analytics methodology, examined precisely how a limited sample of patients presenting a chief complaint of new onset of abdominal pain, stomach pain, or epigastric pain, were evaluated before a clear diagnosis was reached/not reached. We aim to demonstrate the potential of innovations in visual analytics for improving the challenging task of diagnostic evaluation for undifferentiated complaints through informatics and advanced visual analytics methods. Leveraging EHR data of 501 patients in a single institution’s ambulatory setting, we examined the diagnostic paths of a total of 63 patients reached diagnoses after a range of number of evaluations, as well as 438 patients who did not reach any diagnosis. We illustrate steps in learning and visualizing diagnostic paths, through which we present the diversity in the diagnostic processes in a more visual way compared to traditional analytical methods. We envision that future work with larger data may lead to identification of common diagnostic evaluation strategies and their associations with timely and accurate diagnoses, as well as efficient use of clinical resources.

References
1. Ball JR, Balogh E. Improving Diagnosis in Health Care: Highlights of a Report From the National Academies of Sciences, Engineering, and Medicine. Ann Intern Med. 2016;164(1):59-61.
2. Saber Tehrani AS, Lee H, Mathews SC, et al. 25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National Practitioner Data Bank. BMJ Qual Saf. 2013;22(8):672-680.
3. Sloane PD, Coeytaux RR, Beck RS, Dallara J. Dizziness: state of the science. Ann Intern Med. 2001;134(9 Pt 2):823-832.
4. Bishop TF, Ryan AM, Casalino LP. Paid malpractice claims for adverse events in inpatient and outpatient settings. *JAMA*. 2011;305(23):2427-2431.
5. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165(13):1493-1499.
6. Zwaan L, de Bruijne M, Wagner C, et al. Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med.* 2010;170(12):1015-1021.
7. Rao G, Epner P, Bauer V, Solomonides A, Newman-Toker DE. Identifying and analyzing diagnostic paths: a new approach for studying diagnostic practices. *Diagnosis*. 2017;4(2):67-72.
8. Song YJ. Regional Variations in Diagnostic Practices (vol 363, pg 45, 2010). *New Engl J Med.* 2010;363(2):198-198.
9. Goldman RD, Scolnik D, Chauvin-Kimoff L, et al. Practice Variations in the Treatment of Febrile Infants Among Pediatric Emergency Physicians. *Pediatrics*. 2009;124(2):439-445.
10. Institute of Medicine (U.S.). Committee on Standards for Developing Trustworthy Clinical Practice Guidelines., Graham R. Clinical practice guidelines we can trust. Washington, DC: National Academies Press; 2011.
11. Quality AfHRa. Guideline Index. 2015; http://www.guideline.gov/browse/index.aspx?alpha=A.
12. Cook KA, Thomas JJ. Illuminating the path: The research and development agenda for visual analytics. 2005.
13. Caban JJ, Gotz D. Visual analytics in healthcare - opportunities and research challenges. *J Am Med Inform Assn.* 2015;22(2):260-262.
14. Zhang Y, Padman R, Patel N. Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data. *Journal of biomedical informatics*. 2015.
15. Aalst WMPvd, Weijters T, Maruster L. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*. 2004;16(9):1128-1142.
16. Yang W, Su Q. Process mining for clinical pathway: Literature review and future directions. Paper presented at: 11th International Conference on Service Systems and Service Management 2014.
17. Huang Z, Lu X, Duan H. On mining clinical pathway patterns from medical behaviors. *Artificial intelligence in medicine*. 2012;56(1):35-50.
18. Furniss SK, Burton MM, Grando A, Larson DW, Kaufman DR. Integrating Process Mining and Cognitive Analysis to Study EHR Workflow. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium*. 2016;2016:580-589.
19. Monroe M, Lan RJ, Lee H, Plaisant C, Shneiderman B. Temporal Event Sequence Simplification. *Ieee Transactions on Visualization and Computer Graphics*. 2013;19(12):2227-2236.
20. Hirsch JS, Tanenbaum JS, Lipsky Gorman S, et al. HARVEST, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association : JAMIA*. 2015;22(2):263-274.
21. Zhang YY, Padman R, Patel N. Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data. *J Biomed Inform*. 2015;58:186-197.
22. Movahedi F, Carey L, Zhang Y, Padman R, Antaki J. Care pathway after Left Ventricular Assist Devices (LVAD) implementation. International Society for Heart and Lung Transplantation; 2017; San Diego, CA.
23. Bergroth L, Hakonen H, Raita T. A survey of longest common subsequence algorithms. *Spire 2000: Seventh International Symposium on String Processing and Information Retrieval - Proceedings*. 2000:39-48.
24. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *ICWSM*. 2009;8:361-362.
25. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PloS one*. 2014;9(6).
26. Riehmann P, Hanfler M, Froehlich B. Interactive Sankey diagrams. *Infovis 05: Ieee Symposium on Information Visualization, Proceedings*. 2005:233-240.
27. West VL, Borland D, Hammond WE. Innovative information visualization of electronic health record data: a systematic review. *Journal of the American Medical Informatics Association : JAMIA*. 2015;22(2):330-339.