

Composer: Visual Cohort Analysis of Patient Outcomes

Jennifer Rogers¹, Nicholas Spina¹, Ashley Neese¹,
Rachel Hess¹, Darrel Brodke¹, Alexander Lex¹
¹University of Utah, Salt Lake City, UT

Abstract

Visual cohort analysis utilizing electronic health record (EHR) data has become an important tool in clinical assessment of patient outcomes. In this paper, we introduce Composer, a visual analysis tool for orthopaedic surgeons to compare changes in physical functions of a patient cohort following various spinal procedures. The goal of our project is to help researchers analyze outcomes of procedures and facilitate informed decision-making about treatment options between patient and clinician. With Composer, analysts can dynamically define a patient cohort using demographic information, clinical parameters, and events in patient medical histories and then analyze patient-reported outcome scores for the cohort over time, as well as compare it to other cohorts. We describe a usage scenario for use of the tool in a clinical setting.

1 Introduction

Determining the best treatment option for patients with back pain involves an assessment of their medical histories and a comparison to similar patients. Such comparisons have relied on a physician’s memory of related prior cases, which can be influenced by cognitive biases. With an increasing amount of data available for patient populations in electronic health records (EHR), visual cohort analysis has gained attention as an informative analytic tool in healthcare. Recent work has shown the efficacy of using subsets of similar patients, referred to as cohorts, for outcome analysis and prediction in a “patient-like-me” approach^{1,2}. This approach can help clinicians assess optimal treatment options for patients with pre-existing conditions (comorbidities) that can influence recovery and response to treatment.

In this paper, we introduce Composer, a visual analysis tool for comparison of patient outcomes in cohorts under alternative treatment options. Composer was developed in collaboration with domain experts at the University of Utah’s Orthopaedic Research Center. We incorporate outcome scores that are frequently measured over the course of treatment in the decision-making process, supplementing physicians’ memory of prior cases. We used the Patient-Reported Outcomes Measurement Information System (PROMIS)³ scores as the metric for patient physical function and well-being over time.

The technical contributions of Composer include methods to flexibly define multiple patient cohorts based on EHR data and demographic attributes as well as medical codes associated with a given medical visit. We provide functionality for PROMIS score normalization to allow for alignment of score trajectories based on events in patient medical histories, such as surgery or injection. We also provide the ability to normalize scores from absolute measurements to relative change to identify improvement of patient physical function. Finally, we introduce aggregation methods to deal with larger patient cohorts.

2 Background

Cohort Analysis. Most clinical guidelines are based on evidence from clinical trials and controlled studies. However, data collected from clinical trials, often sourced from a general population, may not provide an accurate reflection of potential outcomes for subsets of patients with pre-existing conditions and comorbidities². Clinicians are, therefore, interested in using EHR data and observational studies to better identify factors that can influence the recovery of such patients⁴. A cohort is defined as a subset of the general population that shares one or more defining characteristics. The analysis of cohorts has proven effective in the medical community for identifying factors that patient recovery and treatment.

In clinical applications, cohorts can be defined by utilizing patient data collected through the EHR. The medical community has relied on cohort subsets sourced from a large body of EHR data that can be used for retrospective analysis^{4,5}. Cohorts of patients formed from EHR data have the potential to be used for “patients-like-me” comparisons²,

in which clinicians can define a cohort with attributes mirroring a given target patient. These comparisons can help identify factors that influence patient recovery and has been used to develop predictive tools that help domain experts determine the best treatment options for a given patient⁶⁻⁸.

Patient-Reported Outcomes Measurement Information System PROMIS is a validated measurement system that evaluates a range of patient physical functions⁹. In this paper, we use only PROMIS physical function (PF) scores. The PROMIS system defines the abilities of a patient with a specific score, which is determined by patient response to a series of questions³. A patient who can run 10 miles without difficulty would have a PROMIS PF score of approximately 72, whereas a patient with a score of 32 can stand for a short period of time without difficulty¹⁰. If a patient has answered that they have trouble walking a mile, later questions will focus on a smaller range of physical abilities. The score system is converted to a t-score metric that ranges from 0 to 100, with an average ability score of 50 and a standard deviation of 10. All scores are scaled to values relative to the average score, for example, a score of 40 implies physical function that is one standard deviation lower than the score of the reference mean³. The literature cites changes between 2 and 6 points as a meaningful change for patients on the physical function scale.

The University of Utah Orthopaedic Research center has been a proponent in the use of PROMIS scores to assess patient outcomes¹¹. Recent research into PROMIS physical function scores to evaluate a given procedure relative to cost has identified PROMIS PF as a more accurate assessment of physical well-being for patients with spinal ailments than the Oswestry Disability Index, which is derived from patient reported questionnaire and is used to measure lower back pain. Due to its accuracy, PROMIS PF can be a valuable metric to evaluate patient well-being following treatment and assist in evidence-based decision-making for treatment options for patients with spinal conditions¹².

3 Domain Goals and Tasks

This project emerged from a collaboration with four medical researchers from the Orthopaedic Research Center and the Department of Population Health Sciences at the University of Utah, who are currently investigating the use of PROMIS scores as a measure of patient well-being and progression of physical function following various procedures for spinal ailments. In meetings with collaborators on a bi-weekly basis over 18 months, we collected notes on current EHR and PROMIS score use within the Orthopaedic Research Center to identify domain goals and inform the design of our tool.

Two of the collaborators are spinal surgeons who have not used visualization of EHR data when considering a patient's options for treatment. Instead, their assessments have been based on past patient experiences. When determining patient treatment options, they take into account demographics, medical comorbidities such as diabetes, prior treatments, and current symptoms and severity. They then choose the treatment that is likely to result in the best outcome while also considering other factors such as recovery time and cost. Because the medical histories and collected EHR data for the patient population are extensive and involve a variety of records and data types, we sought to develop a visual analysis solution that combines our collaborators' data into a comprehensive dynamic interface that helps them identify trends in patient outcomes. We identified three functionality requirements that inform the design of Composer, defined below:

1. **Define meaningful cohorts of patients** and analyze how this subset of patients reacts to various treatments and procedures. The clinicians need to be able to form cohorts from the EHR data based on patient demographic information, treatment history, medical records, and initial physical function scores.
2. **Compare the outcomes of different cohorts**, for example, physical function outcomes following different treatment options in otherwise identical cohorts, or to identify an effect of a comorbidity.
3. **Normalize Physical Function Scores** in several ways to successfully analyze and compare cohort outcomes, following an event, such as surgery.

4 Related Work

Visualization of patterns in patient medical histories helps identify risk factors that influence patient recovery following treatment¹³. Recently developed clinical tools provide visual support for users, often in the form of aggregated representations of patient data derived from EHR as well as visual comparisons for patient outcomes and trajectories^{5,7,14}. Composer is related to various tools and techniques for cohort definition and EHR analysis, which we discuss below.

Cohort Definition. Cohort definition is a vital first step for analysis. Emergent patterns identified in cohort behavior and outcome remain dependent on the accuracy of the cohort creation¹⁵ and therefore, cohort definition tools often provide visual feedback to track stages in cohort definition¹⁵. We included a visual representation of each filter layer for a cohort in Composer and have extended this idea to allow dynamic changes to filters.

Cohort Comparison. Current visual tools often provide users the ability to compare clinical pathways and outcomes of patients. These comparisons help users identify differences in patient outcomes between two defined cohorts and diverging event sequences within a given cohort's records⁵. Normalization to a standard time metric and alignment at events in the patient histories facilitate comparison and highlight patterns within the data¹³. This time metric, often in the form of days or visits, allows patient histories to be viewed along a common axis. A tool by Bernard et al.¹⁴ allows realignment of events, e.g., when metastases develop in cancer patients. By sorting and realigning, users can better see trends between events and their corresponding phases. Comparisons can be used for identifying both significant differences as well as similarities and recurring patterns. In contrast to Bernard et al., Composer represents patient trajectories as single lines layered over one another which allows visualization of a larger number of patient trajectories at once. In Composer, we normalize patient data to a standard day metric and allow users to realign scores to a common procedure event. This facilitates comparison of score fluctuation for cohorts containing several hundred patients after given events by viewing patient score change aligned on a common axis.

Aggregation. Much patient data includes event sequences and temporal information. With a large amount of patient data over a span of years, visualization of patient care pathways and events can prove difficult. Clinicians must be able identify patterns of events within a single patient's medical history and recurring trends between multiple patients' records¹⁶. Data, therefore, are often aggregated and summarized to identify emergent patterns within the cohort's medical time-lines and track progression¹⁷. Aggregation can help with pattern identification within complex temporal data by reducing the visual complexity, although it can also hide subtle trends in the data.^{16,18} Composer uses aggregation of individual scores to show emergent trends in PROMIS score fluctuation without the occlusion that happens in a cohort of hundreds of patient scores, but it retains the ability to view the individual score trajectories of patients in the cohort and allows the user to view an aggregation of patients scores as well as the separated scores at their discretion.

Making Relationships in the Data Explicit. Many recent tools facilitate cohort definition and analysis by making relationships between events and static attributes more explicit. Bernard et al.'s visual analysis tool for patients with prostate cancer visualizes distributions of static attributes in the data and indicates when an attribute's frequency is higher or lower in the cohort relative to the population. This visual information is valuable to the domain expert as it provides insight into filter constraints on attributes that might have influenced a subset of patient outcomes¹⁴. Du et al.'s EventAction is a prescriptive visual tool for event sequences. It provides plots showing positive and negative correlations between categories and outcomes¹⁹. Another method of highlighting significant relationships within the cohort data is through visual hierarchy and color. Many visual tools provide color coded highlighting to emphasize significant events^{14,20}. By making these relationships explicit, users can make informed decisions to determine the next steps. We have incorporated these methods in Composer by providing distribution plots to show the number of patients in the entire population who meet the requirements for each filter category. For example, users can see the distribution spread of patient BMI measurements. We also provide visual representation of each filter constraint on a given cohort along with the number of patients at each filter stage.

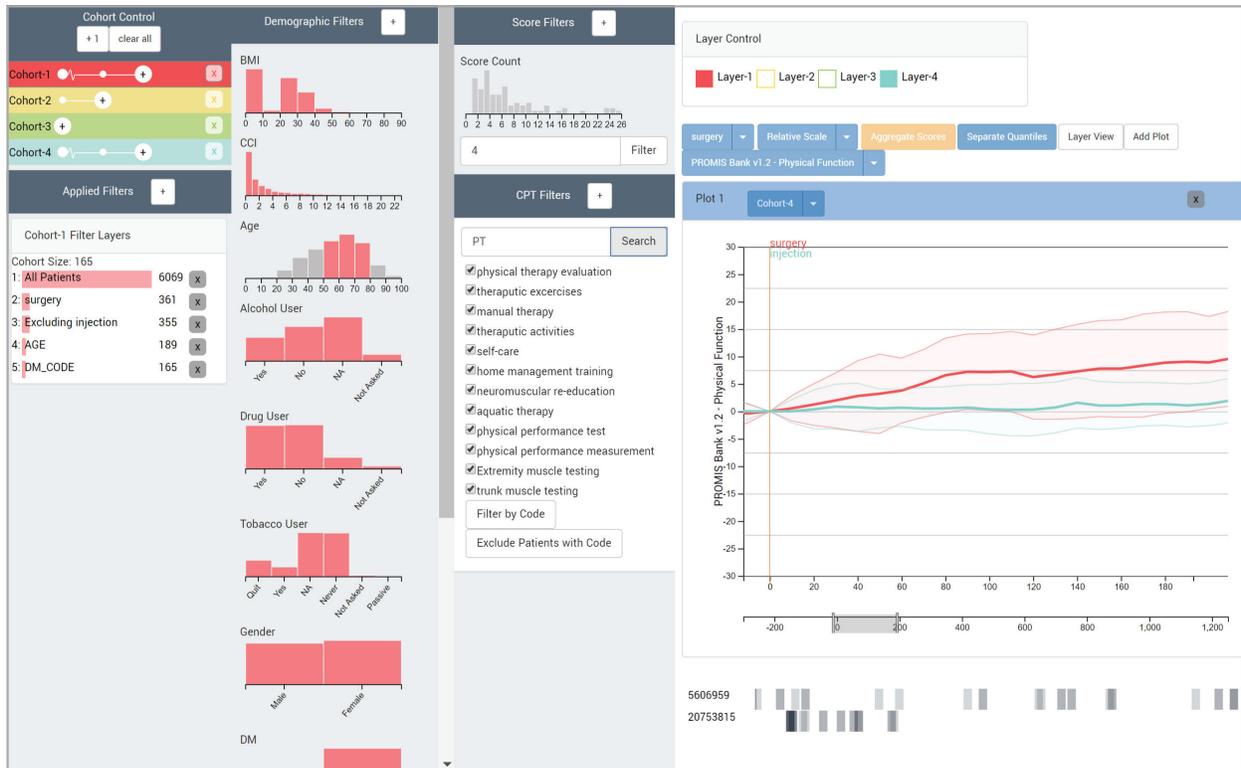


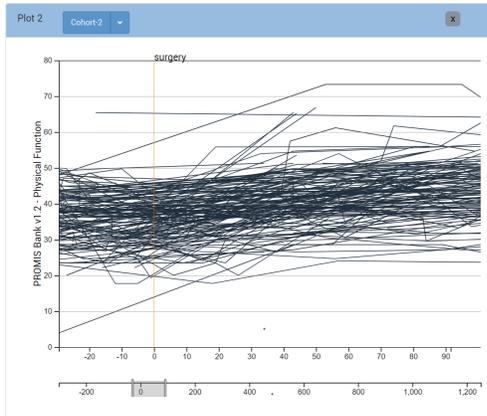
Figure 1: Composer consists of interfaces for displaying the physical function scores of patients treated for back problems over time, and for flexibly defining cohorts. Patient cohorts, shown in the leftmost panel, can be defined based on demographic variables (second panel from the right) and events in their medical history (third panel). The main interface is a chart showing either individual lines, or aggregated areas. A zero-point for the PROMIS scores can be flexibly defined to align all patients by a specific event, such as a medical intervention.

5 Composer Design

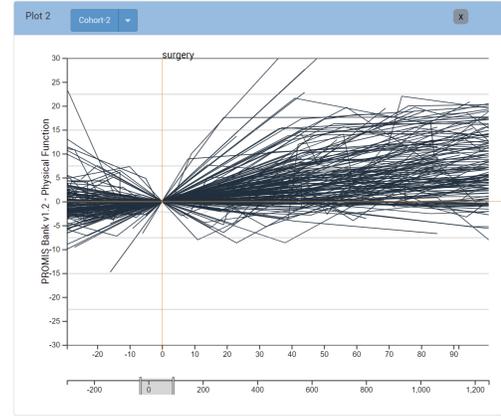
Composer, shown in Figure 1, consists of two components: the cohort definition interface, and the visualization of PROMIS physical function scores. Cohort definition is contained within the collapsible sidebars on the left side of the interface and score visualization and manipulation is placed on the right.

5.1 Cohort Creation

Our collaborators need the ability to define a cohort from a set of specific attributes and medical histories. In Composer's filter sidebar (see Figure 1) cohorts can be defined by demographic information such as age or gender, in addition to other factors deemed relevant, like smoking habits. The filter sidebar is divided into Demographic, Score, and CPT (Current Procedural Terminology; codes used to identify procedures) sections. Within the demographic filters, we use histograms to visualize the distributions of attributes in the patient population. The histograms also serve as means to interact with a filter through brushing for quantitative attributes and selections for categorical ones. In addition to demographic variables, cohorts can also be defined based on the presence or absence of procedure codes in patient histories. This allows analysts to, for example, separate patients that have received a specific surgery from those who have not. With each cohort refinement, a filter layer is added to the sidebar as a visual history of filters used and cohort size at the given filter. Individual filters and cohorts can be removed from the filter history or updated at any time. Composer enables analysts to define multiple cohorts simultaneously. Each cohort is assigned a unique label and color, which is kept consistent across the interface. Cohorts can either be set up independently or branched at any stage. This functionality addresses the need of the domain experts to compare patients of one cohort following various procedures (Figure 1). Once branched, the filter constraints of the parent cohort are duplicated in the branch.



(a) Absolute Score Scale



(b) Relative Score Scale

Figure 2: View of score plots using (a) absolute and (b) relative scales. Relative scales show change in PROMIS score, calculated from the score at the day zero event, in this case the surgery event.

5.2 Outcome Score Comparison

PROMIS physical function scores for the defined cohort are visualized as individual lines showing the course of physical function for each patient over time. The time-window can be resized as desired. By default, we align by the first PROMIS score, yet alignment by a specific clinical event, such as surgery or the start of physical therapy, are often more informative. When different cohorts are aligned by different events this way, the relative progression after the event can be evaluated. We use juxtaposition and superposition to compare between cohorts¹⁸. Juxtaposition visualizes two cohorts side-by-side, while superimposition layers the scores from defined cohorts over one another using color to distinguish them, as shown in Figure 4.

Dynamic Score Scales and Normalization. The physical function scores used by the domain experts are often subtle in absolute measured change (see Figure 2a), yet these subtle changes often have significant impact on the perceived well-being of patients. Change in patient scores are further obscured as patients in the same cohort have different baseline scores. To emphasize change and normalize the baseline, analysts can view scores on a normalized scale that visualizes relative score change for the patients, as shown in Figure 2b. With the option of both absolute and relative score scales, analysts can assess the cohort's overall trend in baseline score measurements as well as trends in score fluctuation. By showing relative score change and making the relationship between cohort scores more explicit, analysts can see differences in outcome trajectories during comparison more clearly.

Separation of Scores by Quantiles. Even in a well-defined cohort, patient outcomes can be markedly different. Due to this heterogeneity, our collaborators need the ability to separate the cohort into quantiles that communicate how, for example, the physical function changes for the top 25 percent of patients in the cohort (see Figure 3a). In Composer, a cohort can be divided by quartiles. We calculate these quartiles by the average change in score over a user-adjustable period of days following a given event.

Aggregation of Scores. Frequently, our collaborators do not need to view individual patients, but rather are interested in aggregate representation of scores. To address this requirement, we provide means to aggregate the scores of a cohort into an area chart centered on the median of the data and extending by plus/minus one standard deviation. This aggregation can be done with the cohort using absolute or relative scales. Cohort scores can also be separated by quantiles to more clearly identify any difference in score change within subsets of the cohort that have different baseline measurements, as shown in Figure 3b.

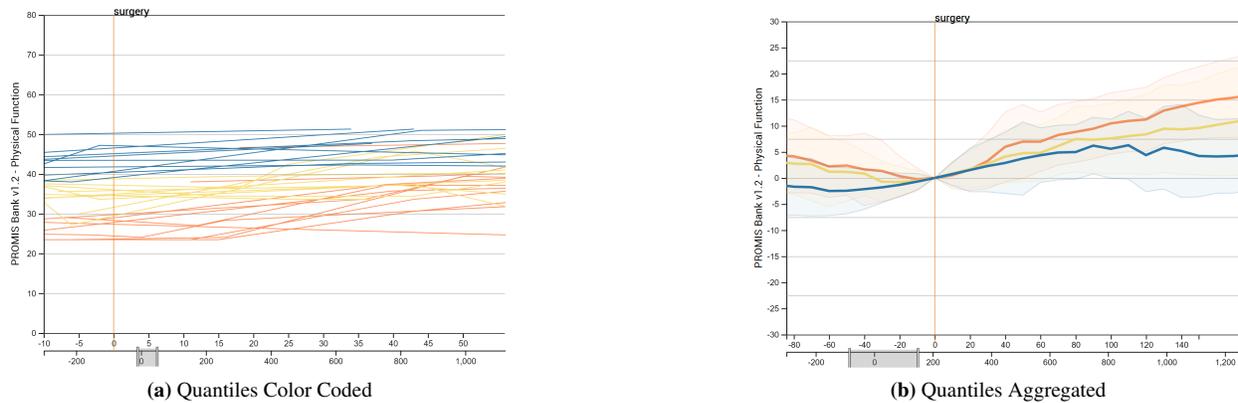


Figure 3: View of patient scores separated and color coded by quantiles. The PROMIS scores were separated into quartiles, shown as individual lines in (a) and aggregated area charts in (b). The orange marks represents the top quartile, the yellow marks the inter-quartile range, and the blue marks the bottom quartile.

Individual Patient CPT History View. For further analysis of procedure code distributions and procedure frequency, patients procedure code histories can be viewed by selecting individual patients in the score chart. These codes are shown as rectangles plotted along the same x-axis as the PROMIS score chart. These events can provide context but can also be used to further filter a cohort. Analysts can view patient histories by selecting the patient's PROMIS scores on a given plot. The events then appear below the plot, aligned on the same time.

5.3 Implementation

Composer is open source and was developed with Typescript using the D3.js library for visualization. The prototype is a Caleydo Phovea client/server application²¹. The Code for Composer can be found at <https://github.com/visdesignlab/Composer>. Data used for development and to inform the usage scenario was sourced from a sample of EHR provided by our collaborators from the Orthopaedic Research center.

6 Usage Scenario

A surgeon sees a patient suffering from a herniated disc. While evaluating potential treatment options for the patient, she defines a cohort in Composer by constraints similar to the given patient's medical history. She filters by the patient's age range, specifies the cohort to only include diabetic patients, and filters just those patients that have had physical therapy evaluation. The cohort defined by these patient specific filters contains 3317 patients. She branches the cohort and filters the initial branch by those that have had surgery, but have not had an injection. She then filters the secondary branch by those patients that have had an injection but not surgery. Aligning each cohort by the surgery or injection event they were filtered by, she can view the diverging cohorts superimposed over one another and visually compare differences in PROMIS score fluctuation between the two. She can then aggregate the individual scores to show only the median PROMIS score within the cohort. Next, she normalizes the PROMIS scores from the absolute score measurement to relative score change, so that she can visually compare the difference in score change between the two to determine what treatment appears to produce better outcomes (Figure 4). After comparing the change in score across a span of 150 days after treatment, she can see that surgery had a greater positive change in physical function, which is clearly visible after the first month (Figure 4a). She can take this into consideration when determining patient treatment options, and show this visualization to the patient when discussing treatment options.

7 Conclusion and Implications for Future Work

In this paper, we outlined the domain analysis for and the design of Composer, an application to visualize and compare patient cohorts and their physical function trajectories. This tool was developed in collaboration with domain experts from the Orthopaedic Research Center at the University of Utah, with their current research in the efficacy of PROMIS scores to evaluate physical function of patients with lower back conditions. Composer is still in development, with

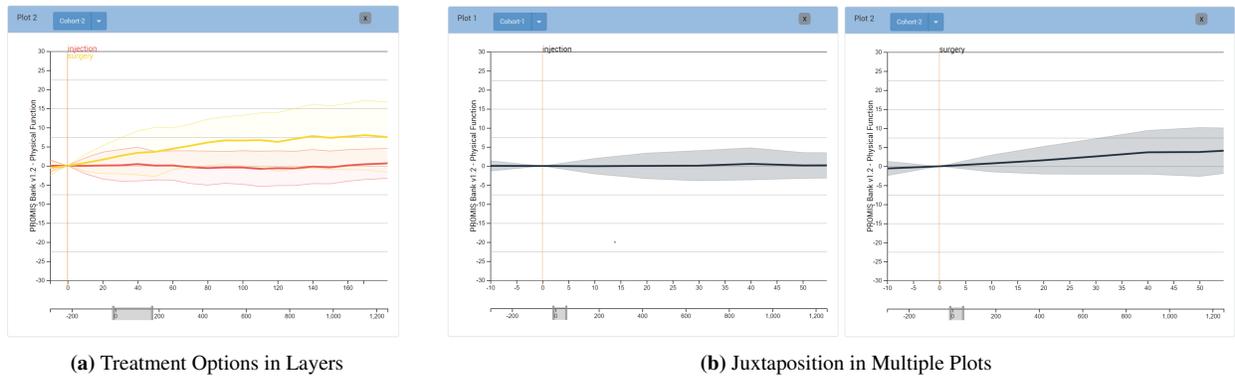


Figure 4: Differences in PROMIS scores after surgery and injection compared by (a) layering and (b) juxtaposition of multiple plots.

progressive iterations made in response to feedback received from meetings with collaborators. As we develop Composer beyond its proof-of-concept stage, we intend to integrate the tool with our collaborator’s EHR database. In the near future, we plan to provide a more extensive statistical breakdown of cohort medical history with the inclusion of ICD codes. As distributions of events and attributes become more explicit, users will be able to apply more accurate filtering constraints to define cohorts. Additionally, we plan to provide more control of the CPT filter codes as they appear within the patient record, and inclusion of sequence specific event filters. As recent literature has shown, medical event sequences can provide important clues on patient outcomes^{5,8,19}.

The long term goal for Composer is the addition of an interface for shared decision making in which insight from exploration in the current interface could be translated into visualizations that would facilitate the explanation of treatment choices and potential outcomes to the patient, and the integration of other measures, such as cost.

Acknowledgements

This project is funded by the Orthopaedic Research Center and NSF IIS 1751238.

References

- [1] Joon Lee, David M Maslove, and Joel A Dubin. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. *PloS one*, 10(5):e0127428, 2015.
- [2] Blanca Gallego, Scott R Walter, Richard O Day, Adam G Dunn, Vijay Sivaraman, Nigam Shah, Christopher A Longhurst, and Enrico Coiera. Bringing cohort studies to the bedside: framework for a green button to support clinical decision-making. *Journal of comparative effectiveness research*, 4(3):191–197, 2015.
- [3] David Cella, William Riley, Arthur Stone, Nan Rothrock, Bryce Reeve, Susan Yount, Dagmar Amtmann, Rita Bode, Daniel Buysse, Seung Choi, et al. Initial adult health item banks and first wave testing of the patient-reported outcomes measurement information system (promis) network: 2005–2008. *Journal of Clinical Epidemiology*, 63(11):1179, 2010.
- [4] Ravi Thadhani and Marcello Tonelli. Cohort studies: marching forward. *Clinical Journal of the American Society of Nephrology*, 1(5):1117–1123, 2006.
- [5] Adam Perer, Fei Wang, and Jianying Hu. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of biomedical informatics*, 56:369–378, 2015.
- [6] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. Finding similar people to guide life choices: Challenge, design, and evaluation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5498–5544. ACM, 2017.

- [7] David Gotz, Fei Wang, and Adam Perer. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *Journal of biomedical informatics*, 48:148–159, 2014.
- [8] Lyndsey Franklin, Catherine Plaisant, Kazi Minhazur Rahman, and Ben Shneiderman. Treatmentexplorer: An interactive decision aid for medical risk communication and treatment exploration. *Interacting with Computers*, 28(3):238–252, 2014.
- [9] Ann L Gruber-Baldini, Craig Velozo, Sergio Romero, and Lisa M Shulman. Validation of the promis[®] measures of self-efficacy for managing chronic conditions. *Quality of Life Research*, 26(7):1915–1924, 2017.
- [10] Jeff Houck, Zane Wise, Amanda Tamanaha, Judith Baumhauer, Luke Skerjanec, Alexandra Wegner, Chris Dasilva, and Michael Bass. What does a promis t-score mean for physical function? *Foot & Ankle Orthopaedics*, 2(3):2473011417S000200, 2017.
- [11] Man Hung, Shirley D Hon, Jeremy D Franklin, Richard W Kendall, Brandon D Lawrence, Ashley Neese, Christine Cheng, and Darrel S Brodke. Psychometric properties of the promis physical function item bank in patients with spinal disorders. *Spine*, 39(2):158–163, 2014.
- [12] Darrel S Brodke, Vadim Goz, Maren W Voss, Brandon D Lawrence, William Ryan Spiker, and Man Hung. Promis pf cat outperforms the odi and sf-36 physical function domain in spine patients. *Spine*, 42(12):921–929, 2017.
- [13] Taowei David Wang, Catherine Plaisant, Alexander J Quinn, Roman Stanchak, Shawn Murphy, and Ben Shneiderman. Aligning temporal data by sentinel events: discovering patterns in electronic health records. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 457–466. ACM, 2008.
- [14] Jürgen Bernard, David Sessler, Thorsten May, Thorsten Schlomm, Dirk Pehrke, and Jörn Kohlhammer. A visual-interactive system for prostate cancer cohort analysis. *IEEE computer graphics and applications*, 35(3):44–55, 2015.
- [15] Josua Krause, Adam Perer, and Harry Stavropoulos. Supporting iterative cohort construction with visual temporal queries. *IEEE transactions on visualization and computer graphics*, 22(1):91–100, 2016.
- [16] Megan Monroe, Rongjian Lan, Hanseung Lee, Catherine Plaisant, and Ben Shneiderman. Temporal event sequence simplification. *IEEE transactions on visualization and computer graphics*, 19(12):2227–2236, 2013.
- [17] Wathsala Widanagamaachchi, Yarden Livnat, Peer-Timo Bremer, Scott Duvall, and Valerio Pascucci. Interactive visualization and exploration of patient progression in a hospital setting. In *AMIA Annual Symposium Proceedings*, volume 2017, page 1773. American Medical Informatics Association, 2017.
- [18] Michael Gleicher. Considerations for visualizing comparison. *IEEE transactions on visualization and computer graphics*, 24(1):413–423, 2018.
- [19] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. Eventaction: Visual analytics for temporal event sequence recommendation. In *Visual Analytics Science and Technology (VAST), 2016 IEEE Conference on*, pages 61–70. IEEE, 2016.
- [20] Theresia Gschwandtner, Wolfgang Aigner, Katharina Kaiser, Silvia Miksch, and Andreas Seyfang. Design and evaluation of an interactive visualization of therapy plans and patient data. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, pages 421–428. British Computer Society, 2011.
- [21] Samuel Gratzl, Nils Gehlenborg, Alexander Lex, Hendrik Strobel, Christian Partl, and Marc Streit. Caleydo Web: An Integrated Visual Analysis Platform for Biomedical Data. In *Poster Compendium of the IEEE Conference on Information Visualization (InfoVis '15)*. IEEE, 2015.