Towards A Task Taxonomy of Visual Analysis of Electronic Health or Medical Record Data

Xiaohui Bian[1,2,*], Hadi Kharrazi[3,*], Jesus J. Caban[4,*], Gaoqi He[5,*], Zhiquan Feng[6,7,*], Jian Chen[8,*]
1. Department of Computer Science and Technology, East China University of Science and Technology, Shanghai, China
2. Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Maryland, USA
3. Department of Health Policy and Management, The Johns Hopkins School of Public Health, Baltimore, USA
4. The National Intrepid Center of Excellence, Walter Reed National Military Medical Center, Bethesda, USA
5. Department of Computer Science and Software Engineering, East China Normal University, Shanghai, China
6. School of Information Science and Engineering, University of Jinan, Jinan, China
7. Shandong Provincial Key Laboratory of Network-based Intelligent Computing, Jinan, China
8. Department of Computer Science and Engineering, The Ohio State University, Columbus, USA
*All authors are corresponding authors.
Email address: y30150667@mail.ecust.edu.cn

***Abstract*—We integrate literature- and data-driven task analysis methods to derive an initial task taxonomy for electronic health record (EHR) and electronic medical record (EMR) data analysis. An EHR (EMR) is a digital and longitudinal version of a patients health(medical) information and may include all key clinical events relevant to that persons health (medical) history, such as provider, demographics, progress notes, medicine, diagnosis, etc. Our goal is to arrive a task taxonomy for analyzing EHR (EMR) datasets because tasks play an important role in the design and evaluation of visualization techniques. Our method has three stages: data collection, task modelling, and task taxonomy summary. In data collection, we first survey related literature from the past two decades and extract typical tasks and corresponding data by extracting goals and scenarios of the particular work. We introduce *multiple continuous relations* to describe specific binary or multiple continuous relation-seeking tasks. Finally, we arrive an initial set of task types for EHR/EMR analysis that guide the design and evaluation of visualization techniques.**

***Keywords*—Task taxonomy, visualization task analysis, electronic health record, literature-driven method, data-driven method.**

I. Introduction

With the development of data collection and storage techniques, the scale and complexity of electronic health records (EHR) and electronic medical records (EMR) are increasing rapidly and the relationships among entities in big EHR/EMR data are becoming challenging to understand. To help researchers and users easily discover the underlying patterns and outliers of complex data, visualization techniques and systems have become indispensible tools, as evidenced by recent publications [1] [2] [3].

Task taxonomy plays an important role in the design and evaluation of visualization systems [4]. A task is a collection of activities to accomplish a specific goal, and a task taxonomy aims to help data analysts understand users demands accurately and clarify their research target. For practitioners, task taxonomy helps characterize users tasks and data types in specific application domains, ensuring that the visualization system best supports users needs [5] [6]. For visualization evaluations, task taxonomy helps benchmark performance for comparative analysis and design recommendations in specific application domains (e.g., [7].)

A number of general task taxonomy frameworks have been applied successfully in several fields, such as hierarchical task analysis (HTA) framework [8] and the Andrienko-Andrienko data-driven framework [9]. In using them to EHR/EMR analysis, it is still necessary to consider domain-specificity when choosing tasks to support visualization design and evaluation. Here we combine literature-based and data-driven methods for EHR/EMR task analysis. The main contributions of our work are as follows:

- Select, collect, and analyze EHR/EMR papers over the past two decades. Typical tasks and related data are analyzed by the motivations and goals of the particular work.
- Extend the data-driven task analysis method to multivariate continuous relationship findings to describe longitudinal trajectory tasks.
- Construct an initial EHR/EMR task taxonomy, using literature- and data-driven methods, that can be used for subsequent design and evaluation.

II. Related work

Traditional task analysis involves asking the users. Recent novel task taxonomies also include data-driven solutions and meta-level task-analysis methods.

A. Task-Analysis Methods

User-driven methods use questionnaires, surveys, and field studies to derive needs from stakeholders [10]. To form a task typology, Schulz et al. characterize tasks in a design space of five dimensions: goal, means, characteristics, target, and cardinality [11]. Brehmer and Munzer extend Roths taxonomy [12] to distinguish the goals, objectives, operators, and operands to form a typology of *why*, *what*, and *how* [5] that has been widely adopted in tool designs, user studies, and nascent technologies [13].

Most similar to ours are the data-driven methods. Andrienko and Andrienko [9] formalize a data-driven approach to derive tasks in geospatial and temporal data explorations. This framework has two parts: a data model and a task framework. The data model divides data components into two categories, *referential* and *characteristic*. Referential components define the context in which the data are obtained, including the moment when the measurements are made, the location where the measurements are made, and the entities that are measured (e.g. time, space, population). Characteristic components are results of measurements, observations, calculations, etc. obtained in that context. In addition, a data function is proposed to define the correspondence between referential and characteristic components. Hence, the dataset structure in the data model is represented as a combination of three key components: the set of all references, the set of all possible characteristics, and the data function.

The task framework distinguishes tasks according to the level of data analysis, dividing them into elementary and synoptic tasks. Elementary tasks involve individual elements of the reference set; synoptic tasks involve the entire reference set or its subsets. Furthermore, tasks are further divided into three subcategories according to their constraints and targets: *lookup*, *comparison*, and *relation seeking* tasks.

This framework, though originally designed for geospatial data analysis, has found many uses in other domains. In particular, Kerracher et al. [4] extended this data-driven solution to graph data task analysis. Kerracher et al. pointed out that the original categories of references (time, space and population) are difficult to apply to graph data, and the categories of relations between references (continuity, order and distance) are difficult to use for edges between nodes in graph data. Thus, they extended the references with "graph" and extended the relations between references with "link".

B. EHR Data Visualization

Many tools are designed for solving specific application problems related to single items in datasets. For example, Plaisant et al. [14] designed lifeLines to explore cohort-graphs from single patient records in timelines. Wang et al. [15] design lifeLines2 to support query and other operations such as align, rank, and filtering to visualize estimates of the intervals in order to find temporal patterns across health records.

Others have examined aggregated data visualizations. For example, Wongsuphasawat et al. [16]'s LifeFlow provides an interactive visualization of cohort event sequences. Wongsuphasawat and Gotz [1] design Outflow to explore flow as well as factors and outcomes of temporal event sequences to help clinicians understand how certain disease-progression paths may lead to better or worse outcomes. Outflow visualizes aggregated event progression pathways together with associated statistical analysis. Perer and Sun [18] proposed MatrixFlow to track symptom evolution during disease progression through temporal event analysis. Gotz and Stavropoulos [19] designed a milestone-demotion algorithm in DecisionFlow for showing patients medical history ordered by time and types. Perer et al. [3]s Care Pathway Explorer supports interactive exploration for researchers to examine the level of detail relevant to user tasks. Debek et al. [20] designed a visualization tool for symptom transformations in EHR data. In addition, PatternFinder [21], COQUITO [22], TimeSpan [23] and Eventpad [24] support interactive query interfaces to specify the temporal queries and often aggregate temporal event patterns. Gotz and Stavropoulos [17] design a mileston demotion algorithm in DecisionFlow for showing patients' medical history ordered by time and types. Perer et al. [3]'s Care Pathway Explorer supports interactive exploration for researchers to examine the level-of-detail relevant to user tasks. Debek et al. [18] designed a visualization tool for symptom transformations in EHR data. Besides, PatternFinder [19], COQUITO [20], TimeSpan [21] and Eventpad [22] support interactive query interfaces to specify the temporal queries and often aggregate temporal event patterns.

III. Our Method

A. Literature Data Collection

Our first stage in deriving the tasks is to examine those studied in the literature. We extract typical tasks with researchers interest by surveying the literature related to medical data visualization literature and then analyze them by data types.

We collect related literature in two ways. One is to follow the work of visualization experts known for their EHR/EMR-related research, such as Ben Shneiderman, Catherine Plaisant, Adam Perer, David Gotz, Fei Wang, all of whom have made great contributions to EHR/EMR data visualizations over the past two decades. We collected 18 papers. The second way is to retrieve related literature over the past five years with keywords "electronic health data (record)" and "EHR data visualization." We used the top 24 most-cited papers.

B. Task Extraction from Literature

Our first step is to manually curate tasks from our literature collection by tabling *who* (user), *what* (function), *when* (condition), *why* (goal), and *how* (method). Sentences related to tasks are extracted and recorded. We especially focus on the motivation and goals (why). Table I shows the analysis results of part of the literature collection.

We take the Perer et al. paper "Mining and exploring care pathways from electronic medical records with visual analytics" [3] as an example to explain the task extraction steps. The main goal of this work came from the abstract:

"*The goal is to utilize historical EMR data to extract common sequences of medical events such as diagnoses and treatments, and investigate how these seuences correlate with patient outcome*."

Perer et al. further described the scenarios related to diabetes patients. We read the scenarios and subsequently summarize the task as follows:

TABLE I
4W+H ANALYSIS FOR MEDICAL DATA VISUALIZATION LITERATURE

Who (user)	When (condition)	Why (goal)	What (functionality)	How (method)
Physicians [3]	Temporal event sequence data	Utilize historical EMR data to extract common sequences of medical events such as diagnoses and treatments, and investigate how these sequences correlate with patient outcome	1. Give an overview of the frequent patterns; 2. Examine the frequent patterns and select specific patterns of interest; 3. Compute the patient subsets that match physicians specified subtraces; 4. The Frequent Pattern Analytics mines frequent patterns and displays them in the visualization.	1. Frequent-sequence-mining algorithm; 2. Bubble chart for overview visualization and flow visualization
Clinicians [1]	Temporal event sequence data	Provide important insights into how diseases evolve over time and help clinicians understand how certain progression paths may lead to better or worse outcomes.	1. Aggregate multiple event sequences; 2. Display aggregate pathways; 3. Summarize the pathways corresponding outcomes; 4. Let users explore external factors.	Flow visualization
Analysts and epidemiologists [17]	High-dimensional temporal event sequence data	Help analysts and epidemiologists study data from patient cohorts to understand what factors may influence particular outcomes.	1. Issue a query to retrieve subsequences of interest; 2. Construct a DecisionFlow Graph aggregated to the matching data; 3. DecisionFlow Graph is analyzed to extract statistics and visualized; 4. Interaction allows exploratory analysis	1. Milestone demotion algorithm; 2. Horizontal layout algorithm of milestone nodes
Investigators [23]	Temporal event sequence data	Help to understand the patterns of events observed within a population that most correlate with differences in outcome.	1. A visual query module to specify episode definitions interactively; 2. A pattern-mining module to help discover important intermediate events within an episode; 3. An interactive visualization module that helps uncover event patterns most impacting outcome and how those associations change over time.	1. Visual query capabilities; 2. Pattern mining techniques; 3. Interactive visualization techniques
Physicians [14]	Personal medical history records	Design appropriate visualization and navigation techniques for presenting and exploring personal medical history records.	1. Present a personal history overview on a single screen; 2. Provide direct access to all detailed information from the overview with one or two clicks of the mouse; 3. Make critical information or alerts visible at the overview level.	Medical record is summarized as a set of lines and events on a zoom-able timeline.
Physicians [15]	Multiple records of categorical temporal data	Find hidden patterns contained in EHR/EMR and other temporal datasets.	1. Select subsets of the records from multiple patients. 2. Use control panel to align, rank, and filter the display.	Timelines with the same absolute time scale
Physicians [19]	Multivariate and categorical data	Search and discovery of temporal patterns within multivariate and categorical datasets.	1. Visual temporal query languages 2. Query result visualization	1. Define a temporal pattern as a sequence of events separated by time spans so that it can be queried by events and time spans components. 2. Multiple timelines for query results.

What are the common medical conditions after one year for hyperlipidemia patients with hypertension and diabetes preconditions?

Further decomposing this task, we obtain the subtasks as follows:

- *Who in the patient cohort were diagnosed with both hypertension and diabetes? and are these conditions preconditions to hyperlipidemia?*
- *When are patients diagnosed with hyperlipidemia?*
- *What are the common medical conditions after a year in the cohort?*

By knowing these tasks, a visualization designer can create exploratory interfaces to investigate what techniques best correlate patients with outcomes. Then users can further answer questions such as *Which (sub)cohorts lead to negative outcomes?* Using this approach, we derived from these papers 107 typical tasks in EHR and EMR data visualization tasks.

C. Data Type Characterization

We characterize data types needed to accomplish the tasks derived above and then fit the data into the Andrienko-Andrienko data-driven framework by characterizing them into referential and characteristics data types, as shown in Table II.

TABLE II
DATA COMPONENTS IN MEDICAL DATA VISUALIZATION RESEARCH

Data type	Data	Detail
Reference	Time	Admission time Discharge time Medical event time
	Patient	Date of birth Date of death Sex Race State
Characteristics	Medical event	Diagnosis Lab test Medication order Treatment Transfer among hospital departments
	Outcome	Positive outcome / Negative outcome

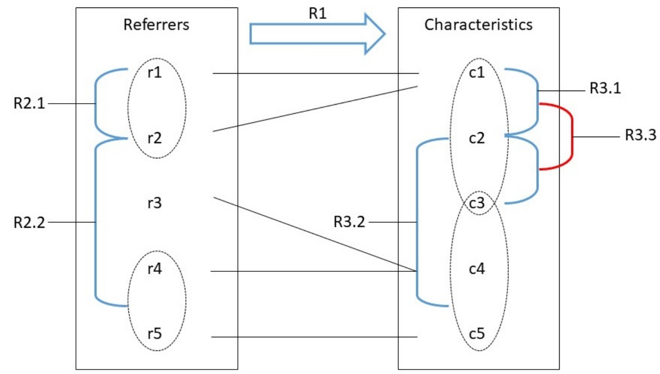

Fig. 1. Our extended relation descriptions between and among data components.

IV. TASK MODELLING

The original data-driven framework defines five relation types: (1) R1 is the relations between references and characteristics; (2) R2.1 is the relations between individual references and contains continuity, order and distance; (3) R2.2 is the relations between references sets (i.e., continuity, order, distance and set relations;) (4) R3.1 is the relations between individual characteristics (i.e., equality, order, distance and set relations); (5) R3.2 is the relation between characteristics sets (i.e., similarity, difference, opposition, correlation, dependency and structural connection). These five types of relations are shown by the blue lines in Fig. 1.

When analyzing the relations between data, we found that it was difficult to describe some EHR/EMR tasks. For instance, in the task *"Find the patients who have been diagnosed with D1 followed by new diagnosis D2 and finally D3."*, the relation among diagnoses does not conform to any relations between characteristics such as equality, order, distance and set relations. Although the task case we described seems similar to the *order* relation, they are not the same because events can co-occur. For example, if a patient is diagnosed with cold at first, then *fever*, and finally *cold* again, the *cold* occurred or co-occurred before and after *fever*. Thus the relation between *cold* and *fever* cannot be simply described as an ordered relation.

Kerracher et al. previously extended the categories of *references* with "graph" and categories of relations between references with "link." We have classified the medical/health events into characteristics rather than references. The "graph" and "link" concepts can still be used to describe the medical events and relations between medical events. In the Andrienko-Andrienko framework, the relations between characteristics data components are of two sorts: (1) relations between individual data components (R3.1); (2) relations between data groups or relations between individual data components and data groups (R3.2). For a given example task, Event D3 is neither directly related to event D1 or event D2, nor simply related to the group D1 and D2; it is related to the relation between D1 and D2. Thus, in order to describe the relation among D1, D2 and D3 in the task *"Find the patients who have been diagnosed with D1, and then diagnosed with D2, and then diagnosed with D3,"* we derive a new type of relation R3.3 among characteristics, shown by the red line in Fig. 1. R3.3 refers to the relations among multiple continuous separate characteristics. And we call this kind of relations as *multiple continuous relation*.

The Andrienko-Andrienko data-driven task-model approach subdivides tasks into a hierarchy according to their constraints and targets: elementary tasks and synoptic tasks at the first level and then lookup, comparison, and relation-seeking tasks, which are children of the synoptic tasks, in the second level.

(1) Lookup tasks give a reference or characteristic as constraints, and ask for another kind of data component. Tasks with references as constraints and characteristics as targets are direct lookup tasks; tasks with characteristics as constraints and references as targets are inverse lookup tasks.

(2) Comparison tasks ask for the relations between references or characteristics, and usually include lookup tasks as their subtasks. Tasks with references as constraints and relations between characteristics as targets are direct comparison tasks; tasks with characteristics as constraints and relations between references as targets are inverse comparison tasks.

(3) Relation seeking tasks give the relations as constraints and ask for references or characteristics.

After introducing the concept of multiple continuous relation, we extend the task model by dividing the relation seeking tasks into two subcategories: binary relation-seeking and multiple continuous relation-seeking tasks. The extended task space is shown in Table III.

V. TASK TAXONOMY

Table II summarizes tasks related to the data (Table III) and tasks (Table IV). The introduction of *multiple continuous relation* and the division into binary and multiple continuous relation-seeking tasks, helped us classify tasks. For instance, the task "*Find patients with medical event E4 occurring before/after the medical event sequence* $E1 \rightarrow E2 \rightarrow E3$", which cannot be generated by the original data- driven framework, can be described using multiple continuous relation-seeking tasks.

TABLE III

OUR PROTOTICAL TASK AND DATA MODELS FOR EHR/EMR TASK ANALYSIS

Task types			Constraints	Targets
Elementary tasks	Lookup	Direct lookup	References	Characteristics
		Inverse lookup	Characteristics	References
	Comparison	Direct comparison	References	Relations
		Inverse comparison	Characteristics	Relations
	Relation-seeking	Binary relation-seeking	Binary relations	References / Characteristics
		Multiple continuous relation-seeking	Multiple continuous relations	References / Characteristics
Synoptic tasks	Lookup	Direct lookup	References	Characteristics
		Inverse lookup	Characteristics	References
	Comparison	Direct comparison	References	Relations
		Inverse comparison	Characteristics	Relations
	Relation-seeking	Binary relation-seeking	Binary relations	References / Characteristics
		Multiple continuous relation-seeking	Multiple continuous relations	References / Characteristics

TABLE IV

OUR PROTYPICAL TASK TAXONOMY FOR ANALYZING EHR/EMR DATASET

	Elementary tasks	Synoptic tasks
Direct lookup tasks	Which medical events occurred at time T for patient cohort P? What is the outcome for patient cohort P?	What kinds of medical event patterns (e.g. frequency distribution of medical events, etc.) occurred in patient cohort P during time period T? What is the average outcome for patient cohort P?
Inverse lookup tasks	When did medical event E occurred for Patient P? For whom did the medical event E occurred at time T? Find those patient cohorts with positive/negative outcome.	For which patient cohort to which the medical event patterns (e.g. frequency distribution of medical events, etc.) occurred during time period T conform to the given pattern? Find patient cohort with positive/negative average outcome.
Direct comparison tasks	Compare the medical events (e.g. lab test results, medication doses, etc.). Compare the outcome of patient cohorts P1 and P2.	Compare the medical event patterns (e.g. the frequency distribution of medical events, the average dose, etc.) of patient cohorts P1 and P2. Compare the average outcomes of patient cohorts P1 and P2.
Inverse comparison tasks	For patient P, compare the time when medical event E1 occurred at the first time and the time when medical event E2 occurred at the first time (e.g. the time order, the time interval, etc.).	Compare the patient cohorts (e.g. the gender distribution, the average age, etc.) with the medical event pattern M1 and M2. Compare the specific cohorts with positive and negative outcomes.
Binary relation seeking tasks	Find patient cohort with medical event E2 occurring before or after medical event E1. Find sub-cohorts with a better/worse outcome in a cohort. Which medical events lead to positive/negative outcome?	Find those patient cohorts with similar medical event patterns (e.g. frequency distribution of medical events, etc.) to cohort P. Find patient cohort with medical event E2 occurring most/least often before/after medical event E1. Find the patient cohort P1 with a better/worse average outcome than patient cohort P2. Which medical events can lead to positive/negative outcome?
Multiple continuous relation seeking tasks	For patient cohort P, which medical events (sequence) occurred before/after the medical event sequence $E1 \rightarrow E2 \rightarrow E3$? Which medical event sequence lead to positive/negative outcome?	Find those patient cohort with medical event E4 occurring most/least often before/after medical event sequence $E1 \rightarrow E2 \rightarrow E3$. Which medical event sequences can easily lead to positive/negative outcome?

VI. CONCLUSION

We have expanded and adapted the Andrienko-Andrienko data-driven task analysis method data collection and coupled it with a literature survey to derive an initial set of EHR/EMR tasks. Future work includes searching for all these relationships that cannot be described and creating patterns for a complete EHR/EMR task taxonomy.

ACKNOWLEDGMENTS

This work was supported by DoD USAMRAA-13318046 through the Geneva Foundation, USA. X. Bian carried out this work when she visited University of Maryland, Baltimore County. G. He was supported in part by National Key Research and Development Program (No. 2016YFA0502304) and the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (No. VRLAB2018B12). Z. Feng was supported in part by the National Key R&D Program of China (No. 2018YFB1004901.)

REFERENCES

[1] K. Wongsuphasawat and D. Gotz, "Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2659–2668, 2012.

[2] D. Liu, P. Xu, and L. Ren, "TPFlow: Progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 1–11, 2018.

[3] A. Perer, F. Wang, and J. Hu, "Mining and exploring care pathways from electronic medical records with visual analytics," *Journal of Biomedical Informatics*, vol. 56, pp. 369–378, 2015.

[4] N. Kerracher, J. Kennedy, and K. Chalmers, "A task taxonomy for temporal graph visualisation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 10, pp. 1160–1172, 2015.

[5] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2376–2385, 2013.

[6] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Visual Languages, 1996. Proceedings., IEEE Symposium on*. IEEE, 1996, pp. 336–343.

[7] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry, "Task taxonomy for graph visualization," in *Proceedings of the AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*. ACM, 2006, pp. 1–5.

[8] J. Annett and K. D. Duncan, "Task analysis and training design," *Occupational Psychology*, pp. 1–22, 1967.

[9] N. Andrienko and G. Andrienko, *Exploratory Analysis of Spatial and Temporal Data*. Springer Berlin Heidelberg, 2006.

[10] S. Theis, P. Wilhelm, V. Rasche, C. Bröhl, M. Wille, and P. Alexander, "Task-data taxonomy for health data visualizations: Web-based survey with experts and older adults," *JMIR Medical Informatics*, vol. 6, no. 3, 2018.

[11] H. Schulz, T. Nocke, M. Heitzler, and H. Schumann, "A design space of visualization tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2366–2375, 2013.

[12] R. E. Roth, "An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2356–2365, 2013.

[13] K. Marriott, J. Chen, M. Hlawatsch, T. Itoh, M. A. Nacenta, G. Reina, and W. Stuerzlinger, "Just 5 questions: Toward a design framework for immersive analytics," in *Immersive Analytics*. Springer, 2018, pp. 259–288.

[14] C. Plaisant, R. Mushlin, A. Snyder, J. Li, H. Dan, and B. Shneiderman, "LifeLines: Using visualization to enhance navigation and analysis of patient records," *AMIA Annual Fall Symposium*, vol. 5, no. 1, pp. 76–80, 1998.

[15] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman, "Aligning temporal data by sentinel events: Discovering patterns in electronic health records," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 2008, pp. 457–466.

[16] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman, "Lifeflow: Visualizing an overview of event sequences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2011, pp. 1747–1756.

[17] D. Gotz and H. Stavropoulos, "Decisionflow: Visual analytics for high-dimensional temporal event sequence data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1783–1792, 2014.

[18] F. Dabek, J. Chen, A. Garbarino, and J. J. Caban, "Visualization of longitudinal clinical trajectories using a graph-based approach," in *Proceedings of the Workshop on Visual Analytics in Healthcare*, 2015, pp. 1–7.

[19] C. Plaisant, S. Lam, S. J. Lam, B. Shneiderman, M. S. Smith, D. Roseman, D. H. Roseman, G. Marchand, M. Gillam, and C. Feied, "Searching electronic health records for temporal patterns in patient histories: a case study with microsoft amalga," *AMIA Annual Symposium Proceedings*, vol. 2008, pp. 601–605, 2008.

[20] J. Krause, A. Perer, and H. Stavropoulos, "Supporting iterative cohort construction with visual temporal queries," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 91–100, 2016.

[21] M. H. Loorak, C. Perin, N. Kamal, M. Hill, and S. Carpendale, "Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 409–418, 2016.

[22] B. C. M. Cappers and J. J. van Wijk, "Exploring multivariate event sequences using rules, aggregations, and selections," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 532–541, 2018.

[23] D. Gotz, F. Wang, and A. Perer, "A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data," *Journal of Biomedical Informatics*, vol. 48, pp. 148–159, 2014.