Eurographics Workshop on Visual Computing for Biology and Medicine (2019)
B. Kozlíková, L. Linsen, and P.-P. Vázquez (Editors)

# `preha`: Establishing Precision Rehabilitation with Visual Analytics

Georg Bernold[1], Kresimir Matkovic[2], M.Eduard Gröller[1,2], Renata G. Raidou[1]

[1]TU Wien, Austria, [2]VRVis Research Center, Austria

**Abstract**

*This design study paper describes `preha`, a novel visual analytics application in the field of in-patient rehabilitation. We conducted extensive interviews with the intended users, i.e., engineers and clinical rehabilitation experts, to determine specific requirements of their analytical process. We identified nine tasks, for which suitable solutions have been designed and developed in the flexible environment of `kibana`. Our application is used to analyze existing rehabilitation data from a large cohort of* 46,000 *patients, and it is the first integrated solution of its kind. It incorporates functionalities for data preprocessing (profiling, wrangling and cleansing), storage, visualization, and predictive analysis on the basis of retrospective outcomes. A positive feedback from the first evaluation with domain experts indicates the usefulness of the newly proposed approach and represents a solid foundation for the introduction of visual analytics to the rehabilitation domain.*

**CCS Concepts**

• ***Human-centered computing*** → ***Visual analytics;*** • ***Applied computing*** → ***Life and medical sciences;***

## 1. Introduction

Personalized medicine aims to improve the treatment of a patient, by allowing doctors to select a strategy that takes the characteristics of the individual patient into account [Nat11, CV15]. In precision medicine, treatment is based not only on individual patient data—it additionally takes into account data within a cohort of similar patients. This approach comes in contrast to "one-size-fits-all" approaches that focus rather on finding the best possible treatment for the "average" patient. In rehabilitation, personalization and precision do not occur, due to a lack of available tools. Deploying a tool for analytics and predictions on available electronic health record (EHR) patient data offers a possibility to determine individual patient rehabilitation plans, supporting *precision rehabilitation*.

Predictive analysis within precision rehabilitation relies on having access to large EHR datasets and being able to process and analyze them. This entails three major challenges: First, large datasets require significant *computational resources*. Second, EHR data is *high-dimensional* and *heterogeneous* (dichotomous, numeric, scales), adding another layer of complexity in terms of interpretation. Third, rehabilitation data analysts have *different kinds of expertise*, e.g., can be clinical domain experts and engineers, and each one of them is interested in *specific tasks*. The visual exploration and analysis of the available precision rehabilitation data is currently a challenge, as there is no dedicated framework for this. Although general frameworks for the analysis of multi-dimensional heterogeneous data could be used for some parts of the workflow, there is no unified solution so far. This design study deals with our newly proposed approach, which takes the whole workflow of the data analysts into account.

In this paper, we are not concerned with introducing novel visualization techniques. Our focus is on leveraging the capabilities of dashboard visualizations and multiple coordinated views to propose solutions for the analysis of rehabilitation data. This is motivated by the tasks that data analysts need to perform, as determined through interviews. Therefore, our *contributions* can be summarized as follows: (1) A detailed data–users–tasks analysis to determine requirements for the exploration and analysis of rehabilitation data (Section 4). (2) `preha`, an integrated Visual Analytics approach that fulfills these predetermined requirements and supports the entire workflow of rehabilitation data analysts (Section 5).

## 2. The Precision Rehabilitation Pathway

The main goal of medical rehabilitation is to enable patients to actively reparticipate in their life, regardless of the origin of their disease [Eng77, HH13]. Our work focuses on *in-patient rehabilitation*, where the patient is treated in a clinic that is specialised in rehabilitation of a specific discipline, e.g., orthopedics or neurology. The patient is admitted to the clinic and remains there for the entire duration of the treatment. To ensure and increase the effectiveness of the pursued rehabilitation treatment, the interventions applied to a patient need to be actively revised. Furthermore, to quantify the effect of this interventions, outcome or status measures have been defined as a tool of evidence-based medicine [Sto11, EJP13, LK14]. These measures are often referred to as *scores*, and they are used to determine the effect of therapies and to make predictions on the rehabilitation progress [NTC*16].

The usual in-patient rehabilitation *pathway* involves a number of steps. First, the patient is referred to a rehabilitation facility,





DOI: 10.2312/vcbm.20191234

where his or her initial state is recorded. This covers the collection of *demographic*, e.g., age, and *medical*, e.g., state of the disability, information. Then, the iterative rehabilitation process begins, consisting of several sequences. At the initial stage of each sequence, an *assessment* aims to quantify the present problems in an objective way [EJP13]. This measurement is a standardized method for assessing those problems. The better the constructs used to perform the measurements, the more valid the measurement becomes, producing more reliable results. Measures with a high degree of confidence are referred to as *standardized outcome measures*. While some measures are easy to describe by objective measurement, such as physical items, other abstract concepts, such as pain, are harder to observe. To overcome this issue, indicators that enable quantification are created, for instance, the Visual Analogue Scale [Sto11, EJP13] for pain measurement.

Based on the assessment, a rehabilitation *goal* is determined. This goal has to be a measurable state, which is checked at every reassessment. An interdisciplinary team is involved in this, including doctors, nurses, and therapists. The doctors adapt the patients treatment plan according to the insights gained up to this point and the other carry out the adaptation. It is worth mentioning that these adaptations are only to be made in a limited range, as the intervention strategy is mainly determined by the payer of the rehabilitation, i.e., often the insurance company. At this point, the planned intervention is carried out and the patient works through his or her therapy plan as scheduled. At the final stage of each iteration, a *reassessment* is performed, using the same measures as in all other assessments for comparability. However, even though the process above is described iteratively, the frequency of the performed measurements differs in reality [KMH01, SO08]. Some measurements are only taken at admission, in order to set up the intervention strategy, and at the time of discharge to objectively measure the rehabilitation progress. In between, these measurements are often missing.

## 3. Related Work

A variety of visualization techniques are used in rehabilitation, which mainly relate to Virtual and Augmented Reality applications that provide visual feedback on the status of the patient [RYS16, DZK*12, DdOL*18]. Visualization in rehabilitation has also been for the presentation of clinical data [LSR*16]. In such approaches, clinical researchers can quickly visualize relationships among rehabilitation variables and efficiently share data, to support hypothesis generation. However, these approaches are mainly targeting the improvement of clinical trial design [LPW*18]. The most closely-related work in precision medicine has presented within the context of cancer treatment by Marai et al. [MMB*18]. Other solutions, specific to the domain of rehabilitation, are not available. We hereby review previous work that tackles similar problems as ours—namely, visualizations for *electronic health record (EHR) data* and for *population or cohort data*.

An extensive survey of information visualization for *EHR data* has been presented by Rind et al. [RWA*13]. Although applications for the analysis of individual patients exist [FN11, PWR*11, PMS*03, PFH07, BSM04], we focus on population and cohort approaches [WPQ*08, WPS*09, WG11, WGGP*11, WS09]. A typical technique applied to visualize data from EHRs is *filtering* or *querying* through a user interface [GWP14]. In environments where visual querying is not available, clinicians rely on database experts or other technologists to create SQL queries for them [ZGP15]. With visual queries, users can build queries in several ways: adding filter elements to a query via drag and drop [GWP14, KPS16], choosing subspaces in visualizations [ZGP15, AHN*17a], or selecting a range in histograms [RSN*19]. Hierarchical data is often used in electronic health records, therefore techniques for *hierarchical data visualization* are applied. To visualize such structures from EHRs, sankey diagrams may be used [ZGP15]. Another approach done by Krause et al. [KPS16] is based on tree maps. *Temporal event analysis* with mining-based and visualization-based methodshas also been proposed [ARH12]. With temporal event analysis, patterns in the timing of this event can be determined to gain insights from clinical event sequence data. Visualization of temporal events in EHR data is used in many applications [GWP14, ZGP15, RPOC18, KPS16].

Visualizing *population and cohort data* has recently become a common task in medical visualization. Especially in medicine, cohort analysis is used to identify risk factors among sub-populations. Cohort analysis can be performed in prospective or retrospective. In retrospective, cohort analysis is performed on a previously collected dataset, while in prospective, the data is collected prior to the analysis. Retrospective analysis is used, e.g., for determining the behavior of a cohort concerning a specific treatment [RCMA*18]. Prospective studies can possibly be used to predict, e.g., the course of disease for the health status of the population. Previous work on the interactive visual analysis of cohorts has been conducted by Steenwijk et al. [SMB*10] and Klemm et al. [KOJL*14], looking more into medical imaging data. Preim et al. [PKH*16], Bernard et al. [BSM*15] and Alemzadeh et al. [AHN*17b] propose applications closer to ours, but none of them tackles all aspects of precision rehabilitation analysis and prediction.

In precision rehabilitation, a comprehensive approach to support the entire worfklow—from the preprocessing (profiling, cleansing, wrangling), to the visualization of the available data for exploration, analysis and presentation of the results, and to the use of predictive analysis for the approximation of rehabilitation outcomes is not available. All aforementioned approaches tackle only *specific parts of the workflow* and there is no unified application to address all stages. The commonalities of our approach with previous related work can be found in the nature of the employed *data*, which are large, multi-dimensional and heterogeneous. However, we differ with regard to the *users* and *tasks* of our application. For users, although the exploration and analysis of the data is conducted by data analysis, who can be clinical domain experts or engineers, the outcome of the analysis is often presented also to patients. Not all involved stakeholders are familiar with visualization or prediction analysis. For the tasks, the data analysts target specific tasks, which cannot be all addressed with one of the existing approaches. The details of these tasks will be discussed in the next section.

## 4. Design Study Analysis: Data—Users—Tasks

In this section, we introduce the available data, the involved users and the identified tasks that guided the design of preha.

### 4.1. Available Data

Our application incorporates functionality for the analysis of an EHR dataset of 46,000 patient cases from multiple rehabilitation centres in Austria. Mainly orthopedic and neurological patients are featured in the dataset. Each case comprises *typical demographic features*, such as age, sex, or residence of the patient, *medical indicators*, such as the primary diagnosis, i.e., the cause for the rehabilitation, and several *health assessment scores*. Our data comprises of nominal (e.g., health insurance provider), dichotomous (e.g., smoker/non-smoker), and ordinal variables (e.g., pain assessment), as well as interval scales and ratios (e.g., height/weight). The data has been collected over a time span of up to seven years, from 2012 to 2019. Other important aspects of the data are missing values, when measurements are not taken or noted, and inconsistencies, for example when an instance of a score is entered as "10m" and another instance as "10 meters".

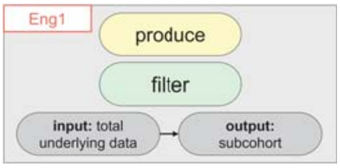
### 4.2. Involved Users

The staff in rehabilitation centres is a multi-disciplinary team, involving both medical and administrative staff. For this work, we focus on two main groups of users: clinicians and IT staff. First, we chose the clinicians, as they are involved in the total rehabilitation process, with ultimate responsibility for treatment decision making and coordination. They have a thorough background knowledge in the field of rehabilitation, they are aware of the EHR data and—to some point—are familiar with data analysis for research purpose. As their asset is their knowledge about the rehabilitation process, we refer to these users as *Domain Experts*. Additionally, we include staff that is responsible for the IT systems that are being deployed within a rehabilitation facility. They have deep knowledge of the EHR data and are familiar with thorough data analysis within the rehabilitation context. Given their deep technological expertise, we refer to these users as *Engineers*. As already mentioned, rehabilitation is an inter-disciplinary process. Thus, we anticipate that the two user groups very often collaborate to improve the rehabilitation process of the patients.

### 4.3. Identified Tasks

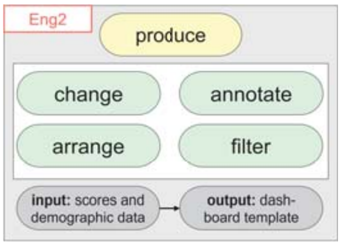
At the beginning of this work, we approached two Domain Experts and four Engineers. We observed their usual workflow with regard to the rehabilitation data and we conducted semi-structured informal interviews with each one of them, to identify their most significant task. Unstructured interviews were avoided as being too vague, while structured interviews leave too little space to freely share thoughts and ideas [LBI*12]. Furthermore, the two user groups have different backgrounds, making it not suitable for sharing the same prepared set of interview questions. The interviews were designed to get an idea of current technologies and work practise in rehabilitation, and lasted between half and one hour. All interview sessions were recorded and the audio recording was transcribed. In our task analysis, we have used the multi-level typology of abstract visualization tasks described by Brehmer and Munzner [BM13]. We hereby summarize nine tasks, which are the results of the interview sessions with the two involved user groups. These are prefixed as `Eng` and `Exp`, for Engineers and Domain Experts respectively. For each of the described tasks, we include a visual notation following the typology of Brehmer and Munzner [BM13], denoting in the schemes *why* with yellow, *how* with green, and *what* with grey.

#### `Eng1`: Provide meaningful data partitions



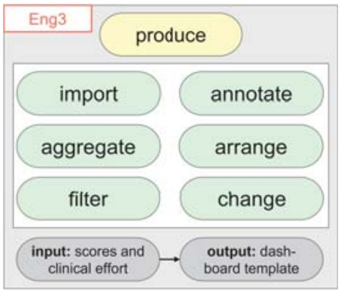

Research is carried out by rehabilitation facilities in order to improve the quality of care in the long term, using population data. The Engineers design meaningful queries, i.e., filters, that produce a subcohort, i.e., a subset of the data, based on a given set of characteristics from the total underlying data structure.

#### `Eng2`: Prepare templates for patient assessment




To discuss the assessment of the patients, it is the task of Engineers to provide the data in the desired format to the Domain Experts. The Engineers must import visualizations to the template and annotate additional information on the displayed score results. As the dashboard is often shown to patients, it is necessary to arrange the visualizations in a clean way, to filter only necessary data and change visualizations to be as simple as possible.

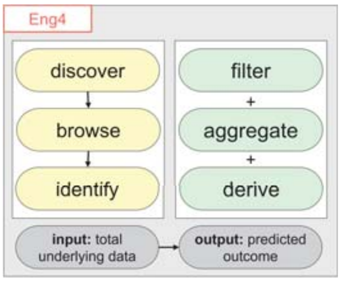
#### `Eng3`: Prepare templates for clinical benchmarking




Clinical benchmarking is a tool used by healthcare facilities to monitor and improve their quality and efficiency [CCCD12]. Payers, government regulators, and affiliated healthcare delivery organizations [RMM01] often demand this instrument to monitor healthcare facilities. The task of Engineers is to produce a template that is used by the Domain Experts (or other entities) for clinical benchmarking. For this, a data structure containing outcome measures and clinical effort, e.g., the total minutes used for therapy, is employed. New entities must be introduced to the dashboard like the import of the required visualizations and annotations for additional descriptions. This can be done by aggregating data, arranging visualizations, filtering the underlying data or changing some visualizations.

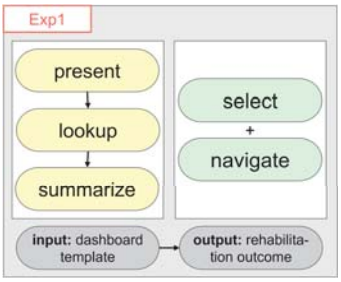
#### `Eng4`: Predict rehabilitation outcome

All kinds of data that have been collected over years can be utilized to predict specific rehabilitation outcome scores. This enables the Domain Experts to predict the outcome of the planned intervention strategy at the start of the rehabilitation. For the Engineers, this prediction enables insight on correlations of certain data features—



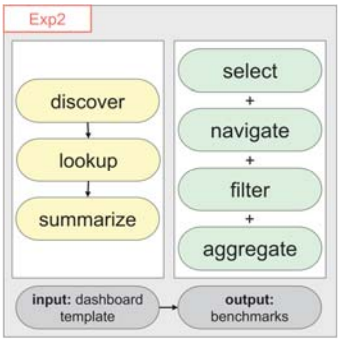

aiding them to gain further knowledge of the underlying data set and the effect of specific features on the outcome success. The Engineers must be able to browse the dataset to identify the features responsible for a successful rehabilitation. Deriving the prediction is possible by filtering and aggregating.

### `Exp1`: Show rehabilitation outcome to patients



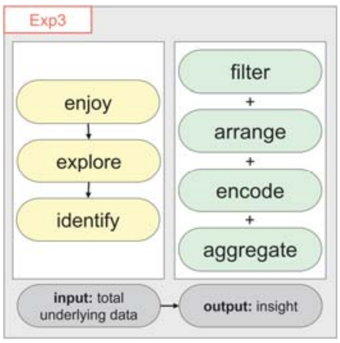

The Domain Experts present the rehabilitation outcome to the patients. This task is based on the dashboard templates created by the Engineers in Task `Eng2`. This task is designed to support a standardized procedure, as opposed to more exploration-focused tasks. The interaction with the data is minimal. The Domain Experts must be able to select items of interest and perform navigation tasks such as zooming, but no modifications of the templates are needed.

### `Exp2`: Perform clinical benchmarking



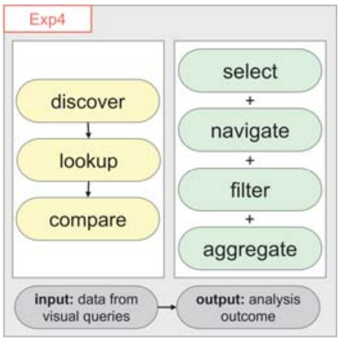

The Domain Experts perform the clinical benchmarking based on the dashboard templates created by the Engineers in Task `Eng3`. The benchmarks must be summarized, so they can be investigated in a clear way through the course of time. It is not desirable to change the clinical benchmarks frequently, as they need to be monitored over time. The users must be able to navigate through the time axis of the visualization, and selection of specific points of interest may be helpful through filtering and aggregation.

### `Exp3`: Explore clinical datasets



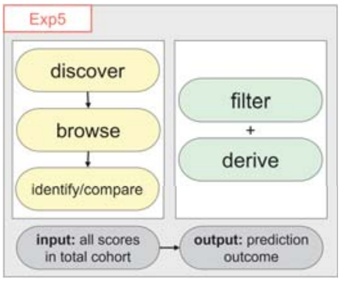

The Domain Experts state that this way of looking into the data may lead to new ideas for scientific research. The motivation for this task is to enjoy the visualization, while exploring the data in order to identify features that are of particular interest for retrospective studies. No restrictions are made on the input data, in order to preserve the free idea of the approach. All tools are available for the users of this task from filtering, arranging, aggregating to even encoding new visualizations.

### `Exp4`: Analyze data for clinical studies




Scientific research is part of the clinical work of Domain Experts. The Domain Experts use the dashboard to discover, lookup and compare data corresponding to a specific rehabilitation treatment strategy or a specific assessment tool. In this task the Domain Experts use the data structure resulting from the visual queries created by the Engineers in Task `Eng1`. Selection, navigation, filtering or aggregation support the analysis.

### `Exp5`: Intervention planning




Domain Experts have to modify the clinical intervention setup, in order to maximize the rehabilitation outcome. This task relates to Task `Eng4`. Correlations can be discovered by browsing through the data to identify or compare outcome measures of interest. The input for this task is the data structure, containing all health assessments in the total cohort. Outcome predictions for varying subcohorts can be derived by applying filters to the population.

## 5. The Basic Modules of `preha`

Our application consists of a set of independent modules that are interconnected, as depicted in Figure 1. First, the *preprocessing* module is responsible for collecting the data from various sources (e.g., database tables), reformatting all data to a single data-structure and standardizing the quality of the data. Then, the *storage* module is

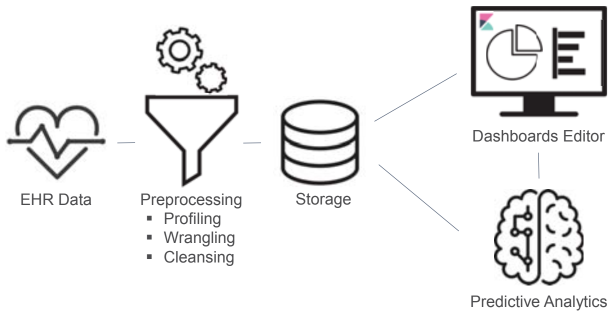



**Figure 1:** *The main components (modules) of* `preha`.

the primary persistence unit for preha. Once the preprocessing module stores data in it, the data are not modified further. The *dashboards editor* module is the user interface of the application, where the tasks discussed in Section 4 are answered. It features dashboards with rich sets of visualizations that are used for data analysis by the data analysts. The conceptual choices for the design of these visualizations will be discussed in detail at the end of this section. The *predictive analytics* module is responsible for advanced data analytic tasks, such as machine learning and predictions. The last three modules—dashboards editor, predictive analytics and data storage—may interact in an iterative process. We hereby discuss the four basic components of *preha*.

### 5.1. Preprocessing

The preprocessing module is the point where the unprocessed, raw data enters the application. The raw data is often obtained from many different sources. In our dataset, the data is either entered by a human (e.g., through forms) or automatically generated (e.g., by the rehabilitation information system). Before any kind of data analytics, data quality control [GAM*14, GE18] is mandatory for resolving the issue of "dirty data". It is the task of the preprocessing module to solve data quality problems by applying a variety of mechanisms that aim to address these issues. The preprocessing module of preha covers data profiling, wrangling, and cleansing. Although visualization could be employed for data preprocessing, we do not follow this approach, as the preprocessing needs to be done only once (or only if the dataset changes) and it does not need to be shown to the data analysts. Only a robust outcome is required, which will be used as input to the next step of the workflow.

*Data profiling* deals with the identification and communication of data quality problems [GGAM12]. Errors can be recognized by applying a regular expression on every entry of a feature. Entries that do not match these regular expressions can be seen as dirty. While simple rules can be automatically applied, they give no insight into the reason for the data quality issues. In order to overcome this issue, we involve data Domain Experts (especially the Engineers) in the data profiling process [GAM*14, KCH*03, GGAM12] to establish a set of rules for each feature and to keep the potential automation level high.

*Data wrangling* is about modifying the structure of the data to make it suitable for further processing (e.g., removing unnecessary rows or columns, splitting variables, merging data from different sources) [KCH*03]. In our case, it is performed to standardize the structure of the diagnoses tables and the scores tables, and to obtain datasets that include an entire patient case per row.

*Data cleansing* is the process of correcting dirty data by repairing or removing it [GAM*14]. There are no generic approaches for *data cleansing*. Dirty data can have multiple causes, so there is no standardized treatment for it. What we know for our dataset is how the correct data should look like for each column. Similarly to profiling, domain expert rules are applied one-by-one to the data, e.g., *all characters must be upper case*. Each cleansing program, i.e., rule, includes boundaries that determine where the values of specific scores can be. A feedback loop for refinement is provided via a result after the cleansing process that displays the percentage of successfully validated values in the dataset. If the user deems that the rate of success is low, new rules may be introduced.

### 5.2. Data Storage

This module is where the data is persisted. All data for the visualization dashboard or the predictive analytics engine is stored in a single table structure, as received from the preprocessing module. The dataset used in preha has a relative small size (46,000 patients), and will not increase significantly over time. Also, the memory will be sufficient even for cohorts that are larger for an order of magnitude. Therefore, we consider the solution as efficient and scalable. With regard to the access to the data, one or few users access the entirety of the data at a time. As the data is used by the visualization dashboard and the predictive analytics modules, speed is a critical facet, in order not to interrupt the thinking process of the user. Hence, we store the data in a simple .csv file. This format enables a high speed for read operations without overhead.

### 5.3. Dashboards Editor

The dashboards editor module is the interface between the users and the underlying dataset. It works as a means to create, maintain, and use dashboards of different levels of complexity that support the tasks of different data analyst groups. The dashboard interface is a crucial component, therefore thorough examination of possible implementations needs to be carried out. Our data analysts are the key users to derive knowledge from the data, but they are not necessarily visualization-literate. A rather minimalistic approach is needed. Additionally, a possible issue is that the tasks may change over time. This could be caused by new visualization technologies, changes in the workflow or issues with existing visualizations. In a static environment, we would have to update the visualization and introduce these changes to all users. Or, if this change is only demanded by parts of the users, a new software branch would be created. Due to these reasons, we decided to resort to dashboards [RF14, SCB*19]. The dashboard components are selected appropriately to answer one-by-one the tasks presented in Section 4, as determined by the intended users of our application, while this approach allows us to exchange or extent easily the employed components. A dynamic dashboard is highly adaptable and capable of supporting the needs of multiple users and multiple tasks, and there is no need to maintain multiple versions.

To support the tasks of Section 4, we decided to base our dashboards on kibana—a simple, yet powerful interface for analytics and visualization from elasticsearch [Gup15]. Elasticsearch is a free and open source technology used for a near realtime analysis of large data sets. Our choice offers capabilities for *categorical visualizations*, e.g., with (stacked) bar charts [Mun14], for *hierarchical visualizations*, e.g., with treemaps [Shn92, JS91] or sunburst diagrams [SCGM00], for *distribution visualizations*, e.g., with dot charts and line charts [Mun14], for *tabular representations* and for *geographical representations*, e.g., choropleth maps. Other visualizations that are not included in kibana can be either added through D3.js [BOH11] or by using the builtin vega [SWH14]. Multiple linked views are also supported. Finally, it provides functionality for traditional interactions [Mun14], such as zooming/panning, filtering, selecting, Focus+Context, and

Brushing and Linking. We will discuss the specific visualization components employed for each one of the tasks in Section 5.5.

### 5.4. Predictive Analytics

We need to implement means for the prediction (approximation) of rehabilitation outcome values for individual patients based on the characteristics of the clinical case, and calculation models based on statistical analysis of the dataset. In the present case of rehabilitation, we know that the data is not completely random, and that there are certain patterns occurring between the features. In this module, we can construct a good and useful approximation of those patterns. Even though we cannot identify all patterns with 100% accuracy, we can make use of them to create predictions for specific features in the future—under the assumption that the future data does not differ much from the data used for machine learning. The predictive analytics module is also based on `kibana`, which contains a built-in machine learning tool. This enables us to conduct *regression analysis tasks* [Fre09], e.g., predict a specific score at the discharge phase of the rehabilitation process based on other features of the patient, and *classification tasks* [Alp10].

A multitude of supervised machine learning algorithms can be used. We decided to employ *random forests* [Bre01] that use a randomized set of decision trees to solve both classification and regression problems. Random forests have high accuracy and robustness, while they do not suffer from overfitting and they provide feature importance information. Within this module, whenever a new filter is added to `kibana`, e.g. by selecting or filtering on a specific visualization, or whenever it is removed, a new request is issued. The predictive analytics engine performs a regression or classification analysis, depending on required task, and when the predictive analytics engine responds to the request, the outcome is presented to the user. We show to the user the predicted value, the accuracy of the prediction, the mean absolute error of the prediction, and—for the regression task—the influence of other features.

To evaluate the performance of the predictive analytics module, we use the interface `kibana` provides. Even though numerous parameters are incorporated in the random forest algorithm, we focus on the two most important for the complexity and time performance of our module: the number of trees and the maximal depth of each tree. A forest with eleven trees and a max depth of three has been determined as the sweet spot yielding good accuracy at a moderate calculation time of about 22 seconds. From our population of $46,000$ patients, we see an increase of our performance metrics from a sample size of $10,000$ patients, where the accuracy of the prediction is about 90%.

### 5.5. Proposed Dashboards for Each Task

We hereby highlight how we addressed the requirements of each task in the design of our application, focusing on the dashboards editor and the predictive analytics modules of `preha`. Each one of the tasks discussed in Section 4 results into an individual dashboard, which comprises specific visualization components. To facilitate the description, we exemplify the solutions for each task with usage scenarios. The visualization components employed for each tasks are always the same. What changes is the insight that they provide (e.g., different patients, different scores).

**`Eng1`: Provide meaningful data partitions**

The aim of this task is to generate meaningful partitions of the entire dataset. The dashboard is presented in Figure 2. The entire cohort and its hierarchical structure shown in a *treemap representation*, where selections are possible. Details on the cohort data can be seen on demand in a *tabular view* (A), and the location of the patients is shown in a *choropleth map* (B). As stated by the Engineers in the interviews, the data is filtered according to common characteristics like age or geographical location of the patient. In our solution, filters can be applied to the data in the form of *simple Brushing and Linking*, or as *textual queries* to create a desired subcohort on visualizations that have been imported into the dashboard. All views are linked in the dashboard. In Figure 2, we show a filtering based on the gender applied on the tabular representation (C) and reflected on the treemap (A) and the map (B).

**`Eng2`: Prepare templates for patient assessment**

For this task, the Engineers provide the assessment data to the Domain Experts, who discuss with their patients how their therapy progressed in comparison to previous patients. The dashboard is presented in Figure 3. Given that the patients are most probably not familiar with visualization, the results of their assessments are displayed to them as *simple metrics* (top). To give the patients a sense of how they compare to others with similar characteristics, a *distribution chart* is used for the assessment data of the entire cohort (bottom). All visualizations need to be arranged so that admission and discharge are clearly separable (left and right). The way the visualizations are prepared is critical for this task, as patients are not used to interpret complex charts. All used visualizations need to be annotated, so it becomes clear to the patient what data they are shown. The views in the dashboard are static. In Figure 3, we show a comparison of a score distribution at admission (left) and at discharge (right) compared to the respective distributions of the entire cohort.

**`Eng3`: Prepare templates for clinical benchmarking**

Clinical benchmarking dashboard templates are also prepared by the Engineers for the Domain Experts to provide data on clinical efficiency, based on specific information that need to be evaluated. Four visualizations are included in the dashboard of Figure 4: a *metric visualization* (A) that displays the total number of patients in the current selection, a *bar chart visualization* (B) of the top five payers of the rehabilitation ordered by number of patients, a *dis-*

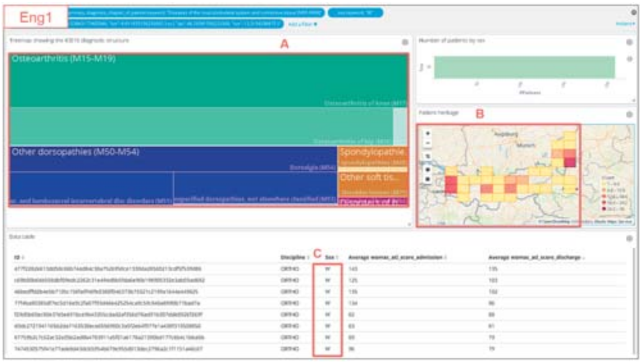



**Figure 2:** *The dashboard of* `Eng1`.

*tribution chart* (C) that displays the development of characteristic patient assessments in a given time frame (C) and a *line chart* (D) that displays the development of patient admissions over the same time frame, grouped by the two facilities. All views are linked in the dashboard.

### Eng4: Predict rehabilitation outcome

The purpose of this task is to predict the discharge value of an assessment for a specific subcohort of patients. The dashboard is presented in Figure 5. The Engineers need to write textual *queries* that are used to define this subcohort (A), from the entire cohort (shown in the treemap). If the query is changed, a new machine learning algorithm is executed and its results are displayed in a dedicated panel (C) in the dashboard. This panel includes metrics for the predicted value, the accuracy of the prediction, and the prediction error. Also, two *bar chart* representations are employed to show the importance of the variables for the prediction (C), and the distribution of this variable in the population (E). All views are linked in the dashboard. In Figure 5, a prediction for the WOMAC score (an index specific to osteoarthritis rehabilitation) is conducted, with a predicted value of 59.33 and a prediction accuracy of 80.92%. Additional visualizations (B),(D) encode characteristics of the queried subcohort, as in task Eng1.

### Exp1: Show rehabilitation outcome to patients

This task is based on the templates created by the Engineers in Task Eng2, and deploys the same representations. The dashboard is presented in Figure 6. The interaction of the Domain Experts with the presented visualizations is limited to setting the *filters* to a subcohort that corresponds to the respective patient, at admission (left) and at discharge (right). The views in the dashboard are static. For example, a neurological male patient at the age of 75 is shown the typical results of his corresponding subcohort at admission and at discharge. His distribution (blue line) is shown in comparison to the population (green line).

### Exp2: Perform clinical benchmarking

Similar to Exp1, also Exp2 is based on a dashboard prepared by the Engineers in Eng3. The dashboard is presented in Figure 7. The Domain Experts apply certain *filters* to the data and monitor the corresponding results. For example, how the developments of specific assessments differ among the rehabilitation facilities can be evaluated by selecting the corresponding segment from the *distribution chart*. Navigating through the time frame can also reveal additional details on the development of assessment or admission figures (right). All views are linked in the dashboard. In Figure 7, we see some periodical "dips" in the development of the therapy over time, which coincides with the holiday period around Christmas, when the patients are temporarily discharged and go home.

### Exp3: Explore clinical datasets

This task is not defined very strictly, and its aim is to provide the


**Figure 3:** *The dashboard of* Eng2.


**Figure 4:** *The dashboard of* Eng3.

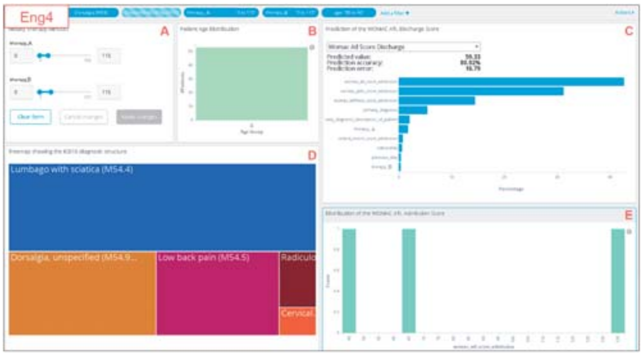



**Figure 5:** *The dashboard of* Eng4.


**Figure 6:** *The dashboard of* Exp1.

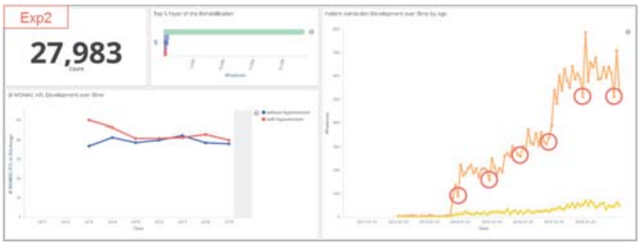



**Figure 7:** *The dashboard of* Exp2.

clinicians with the tools to explore the dataset. This allows the clinicians to utilize all possibilities of the visualization dashboard without any constraints, i.e., all visual representations freely. Possible actions include encoding data in visualizations, creating personal dashboards and defining custom queries. We show in Figure 8 a dashboard that includes *bar carts* (A), *metrics* (B), a *treemap* (C) and a *choropleth map* (D) for the comparison of two subcohorts in the data (visible in A). All views are linked in the dashboard.

### `Exp4`: Analyze data for clinical studies

This task aims at discovering the dataset of the subcohort created by the visual queries defined by the Engineers in `Eng1`. The dashboard is depicted in Figure 9. The Domain Experts interpret the data extracted in the *tabular structure* (D) from the entire cohort shown in the *treemap* (A). This mainly means locating measures of interest in the dataset and comparing them across the subcohort in *bar charts* (B) or *choropleth maps* (C). *Filtering* actions can be performed to view results more individually or to refine the subcohort in the tabular view and the treemap.

### `Exp5`: Intervention planning

In this task, the Domain Experts interact with the machine-learning module. For this task, the dashboard template created by the Engineers in `Eng1` is reused. This is shown in Figure 10. It consists of a *bar chart* (A) that displays the number of patients per facility and the total number of patients (B), a *treemap* view on the entire cohort (C) and a *choropleth map* showing their geolocation (D). The *predictive analytics panel* (E) is added to the visualization, providing all the necessary assessment information as discussed also in `Eng4` in a *histogram* view for the most significant features and simple *metrics* for the accuracy scores. All views are linked.

## 6. Evaluation Results

For the evaluation of each task, we conducted brief evaluation sessions with the users who were also included in the interviews for the design study analysis. For this evaluation, we decided to include only the Engineers to obtain a first feedback, and to include at the next stage the Domain Experts, after incorporating the proposed changes. We documented the overall impression of the users about the designed application, but no metrics were recorded, as the sample of participants is too small for statistically relevant results ($n = 4$). We would like to conduct a more thorough evaluation in the future, so we consider this first evaluation as a pilot study where we gather comments and concerns with regard to `preha`. We anticipate that the results will provide the necessary feedback to revise our approach in the future.

Each evaluation session held with the users started with an explanation of the system. In this, we included an overall introduction of the system, in particular `kibana` and how to use it in order to generate the dashboards. We explained the relationship between visualizations and dashboards, as well as which different types of visualizations are supported in the environment of `preha`. Another important point was which filters are available in `preha`, as well as the available intereaction capabilities, and also how a dashboard is affected when these are applied. We also provided a short reminder of the tasks. We only discussed the tasks described in Section 4 and analyzed in Section 5. For this, we determined a number of well-defined assignments that we provided as real-world scenarios to be accomplished by the evaluation participants. For example, we asked them to perform a meaningful partitioning of the cohort (according to `Eng1`) that includes female patients above 50 years old with a specific primary diagnosis and to answer how many patients are present in the cohort and to visualize the distribution of one of their rehabilitation scores with a bar chart. A second case was to prepare a dashboard where they can use the predictive analytics module (according to `Eng4`) to predict the rehabilitation outcome of a specific cohort partition for two different therapies.

After the users completed the assignments, we asked them to provide statements on how the application helped them to accomplish each task. All users agreed that `preha` is capable of realis-

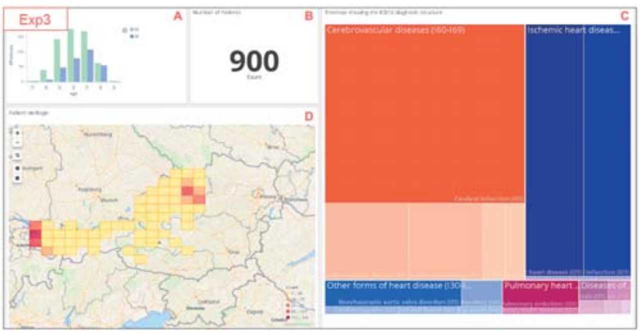



**Figure 8:** *The dashboard of* `Exp3`.

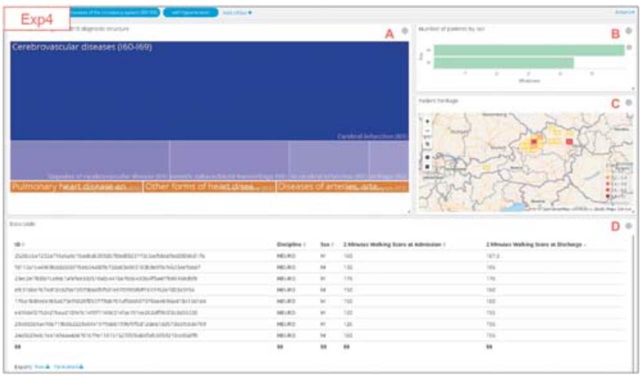



**Figure 9:** *The dashboard of* `Exp4`.

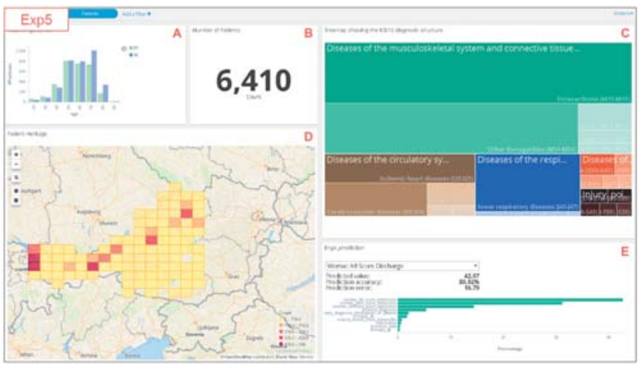



**Figure 10:** *The dashboard of* `Exp5`.

ing all the tasks, as they had been discussed in the first session. Furthermore, the users highlighted that the multiple coordinated views [WBWK00] in preha are a central feature and main advantage. The flexibility of the dashboards, including rearranging and resizing visualizations, is a functionality especially appreciated by the Engineers. In the course of work, we designed preha entirely in English. However, the users suggested to provide preha in their native language. All users stated that they would have required more knowledge to start with the first assignment. On the other hand, the users reported that exploring the system on their own helped them getting to know preha in their own working style. The Engineers further suggested to prepare extensive training material including a lot of examples, before approaching the Domain Experts. One Engineer stated that *"the Domain Experts are not used to work with tools such as preha, they lack required technical knowledge"*. Therefore, they recommended to adapt certain aspects to make the application more digestible for the Domain Experts. Our preha features some highly specialised and complex features, as the first evaluation stage demonstrated. In the course of this evaluation, some particular aspects were pointed out. For example, a lack of extensive documentation, a steep learning curve at the beginning, the placement of the time filter, which made it hard to distinguish, and the oversupply of visualization types, which is not always an easy choice to make. Incorporating the feedback from the Engineers will help us to improve the functionality of preha and design appropriately an evaluation that will include also the clinical domain experts.

## 7. Conclusions and Future Work

This design paper presents a novel application called preha. In tight collaboration with the users, nine tasks have been identified and suitable solutions have been designed for them. Our application incorporates all functionalities that have been set by the task analysis, including data preprocessing, storage, visualization and prediction analysis. Our application is used by its intended users to analyze rehabilitation data and utilizes the outcome to enable precision rehabilitation. preha has been developed on kibana providing flexibility in the creation of the dashboards and in the adaptation to other tasks, and possibly to other domains. An initial evaluation with the intended users of our work has been conducted, with promising results for the future.

The main directions for future work include the incorporation of the feedback of the evaluation participants to rework certain aspects of the tool and the conduction of a more thorough evaluation with all intended users. The visualizations are prototypical approaches and are to be tested further. Other processes from other facilities might need to be incorporated in the future. As visualizations are often used in the communication with patients, this user group should be considered as well. In addition to this, we foresee that the predictive analytics module might need extensions in the future to be able to accommodate more scalable solutions that will be still interactive. Moreover, going towards the direction of Guided Visual Analytics [CGM*16] would be an interesting extension of our work, both for the choice of visualizations to employ and for the conduction of the analysis. This would allow preha to accommodate also users less familiar with visualization.

**Acknowledgments:** This paper was partly written with the VRVis Competence Center, which is funded by BMVIT, BMDW, Styria, SFG and Vienna Business Agency in the scope of Competence Centers for Excellent Technologies (854174), managed by FFG.

## References

[AHN*17a] ALEMZADEH S., HIELSCHER T., NIEMANN U., CIBULSKI L., ITTERMANN T., VÖLZKE H., SPILIOPOULOU M., PREIM B.: Subpopulation discovery and validation in epidemiological data. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2017), The Eurographics Association. 2

[AHN*17b] ALEMZADEH S., HIELSCHER T., NIEMANN U., CIBULSKI L., ITTERMANN T., VÖLZKE H., SPILIOPOULOU M., PREIM B.: Subpopulation Discovery and Validation in Epidemiological Data. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2017), The Eurographics Association. 2

[Alp10] ALPAYDIN E.: *Introduction to machine learning*, 2nd ed. The MIT Press, 2010. 6

[ARH12] AIGNER W., RIND A., HOFFMANN S.: Comparative evaluation of an interactive time-series visualization that combines quantitative data with qualitative abstractions. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 995–1004. 2

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (dec 2013), 2376–2385. 3

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (dec 2011), 2301–2309. 5

[Bre01] BREIMAN L.: Random forests. *Machine Learning 45*, 1 (2001), 5–32. 6

[BSM04] BADE R., SCHLECHTWEG S., MIKSCH S.: Connecting time-oriented data and information to a coherent interactive visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2004), ACM, pp. 105–112. 2

[BSM*15] BERNARD J., SESSLER D., MAY T., SCHLOMM T., PEHRKE D., KOHLHAMMER J.: A visual-interactive system for prostate cancer cohort analysis. *Computer Graphics and Applications (CG&A), IEEE 35*, 3 (2015), 44–55. 2

[CCCD12] COMPLETO J., CRUZ R. S., COHEUR L., DELGADO M.: Design and implementation of a data warehouse for benchmarking in clinical rehabilitation. *Procedia Technology 5* (2012), 885–894. 3

[CGM*16] CENEDA D., GSCHWANDTNER T., MAY T., MIKSCH S., SCHULZ H.-J., STREIT M., TOMINSKI C.: Characterizing guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2016), 111–120. 9

[CV15] COLLINS F. S., VARMUS H.: A new initiative on precision medicine. *New England Journal of Medicine 372*, 9 (2015), 793–795. 1

[DdOL*18] DEBARBA H. G., DE OLIVEIRA M. E., LADERMANN A., CHAGUE S., CHARBONNIER C.: Augmented reality visualization of joint movements for physical examination and rehabilitation. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (mar 2018), IEEE. 2

[DZK*12] DYCK E., ZELL E., KOHSIK A., GREWE P., WINTER Y., PIEFKE M., BOTSCH M.: Octavis: An easy-to-use VR-system for clinical studies. *Virtual Reality Interaction and Physical Simulation* (2012). 2

[EJP13] ENDERBY P., JOHN A., PETHERAM B.: *Therapy outcome measures for rehabilitation professionals: Speech and language therapy, physiotherapy, occupational therapy*. John Wiley & Sons, 2013. 1, 2

[Eng77] ENGEL G. L.: The need for a new medical model: A challenge for biomedicine. *Science 196*, 4286 (1977), 129–136. 1

[FN11] FAIOLA A., NEWLON C.: Advancing critical care in the icu: a human-centered biomedical data visualization systems. In *International Conference on Ergonomics and Health Aspects of Work with Computers* (2011), Springer, pp. 119–128. 2

[Fre09] FREEDMAN D. A.: *Statistical models: theory and practice*. Cambridge University Press, 2009. 6

[GAM*14] GSCHWANDTNER T., AIGNER W., MIKSCH S., GÄRTNER J., KRIGLSTEIN S., POHL M., SUCHY N.: Timecleanser: A visual analytics approach for data cleansing of time-oriented data. In *Proceedings of the 14th International Conference on Knowledge Technologies and Data-Driven Business* (2014), ACM, p. 18. 5

[GE18] GSCHWANDTNER T., ERHART O.: Know your enemy: Identifying quality problems of time series data. In *2018 IEEE Pacific Visualization Symposium (PacificVis)* (apr 2018), IEEE. 5

[GGAM12] GSCHWANDTNER T., GÄRTNER J., AIGNER W., MIKSCH S.: A taxonomy of dirty time-oriented data. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2012, pp. 58–72. 5

[Gup15] GUPTA Y.: *Kibana essentials*, 1 ed. Packt Publishing, 2015. 5

[GWP14] GOTZ D., WANG F., PERER A.: A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *Journal of Biomedical Informatics 48* (apr 2014), 148–159. 2

[HH13] HOFMARCHER-HOLZHACKER M. M.: *Das österreichische Gesundheitssystem: Akteure, Daten, Analysen*. MWV Medizinisch Wissenschaftliche Verlagsgesellschaft mbH & Co. KG, 2013. 1

[JS91] JOHNSON B., SHNEIDERMAN B.: Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *IEEE Conference on Visualization 1991* (1991), IEEE, pp. 284–291. 5

[KCH*03] KIM W., CHOI B.-J., HONG E.-K., KIM S.-K., LEE D.: A taxonomy of dirty data. *Data Mining and Knowledge Discovery 7*, 1 (2003), 81–99. 5

[KMH01] KAY T. M., MYERS A. M., HUIJBREGTS M. P.: How far have we come since 1992? A comparative survey of physiotherapists' use of outcome measures. *Physiotherapy Canada 53*, 4 (2001), 268–275. 2

[KOJL*14] KLEMM P., OELTZE-JAFRA S., LAWONN K., HEGENSCHEID K., VOLZKE H., PREIM B.: Interactive visual analysis of image-centric cohort study data. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1673–1682. 2

[KPS16] KRAUSE J., PERER A., STAVROPOULOS H.: Supporting iterative cohort construction with visual temporal queries. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 91–100. 2

[LBI*12] LAM H., BERTINI E., ISENBERG P., PLAISANT C., CARPENDALE S.: Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics 18*, 9 (2012), 1520–1536. 3

[LK14] LEWIS G., KILLASPY H.: Getting the measure of outcomes in clinical practice. *Advances in Psychiatric Treatment 20*, 3 (may 2014), 165–171. 1

[LPW*18] LOHSE K. R., PATHANIA A., WEGMAN R., BOYD L. A., LANG C. E.: On the reporting of experimental and control therapies in stroke rehabilitation trials: A systematic review. *Archives of Physical Medicine and Rehabilitation 99*, 7 (jul 2018), 1424–1432. 2

[LSR*16] LOHSE K. R., SCHAEFER S. Y., RAIKES A. C., BOYD L. A., LANG C. E.: Asking new questions with old data: The centralized open-access rehabilitation database for stroke. *Frontiers in Neurology 7* (sep 2016). 2

[MMB*18] MARAI G. E., MA C., BURKS A. T., PELLOLIO F., CANAHUATE G., VOCK D. M., MOHAMED A. S., FULLER C. D.: Precision risk analysis of cancer therapy with interactive nomograms and survival plots. *IEEE Transactions on Visualization and Computer Graphics 25*, 4 (2018), 1732–1745. 2

[Mun14] MUNZNER T.: *Visualization analysis and design*. AK Peters/CRC Press, 2014. 5

[Nat11] NATIONAL RESEARCH COUNCIL (US) COMMITTEE ON A FRAMEWORK FOR DEVELOPING A NEW TAXONOMY OF DISEASE: Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease. *National Academies Press (US)* (2011). 1

[NTC*16] NG Y. S., TAN K. H., CHEN C., SENOLOS G. C., CHEW E., KOH G. C.: Predictors of acute, rehabilitation and total length of stay in acute stroke: A prospective cohort study. *Annals of the Academy of Medicine, Singapore 45*, 9 (2016), 394–403. 1

[PFH07] PIECZKIEWICZ D. S., FINKELSTEIN S. M., HERTZ M. I.: Design and evaluation of a web-based interactive visualization system for lung transplant home monitoring data. In *AMIA annual symposium proceedings* (2007), vol. 2007, American Medical Informatics Association, p. 598. 2

[PKH*16] PREIM B., KLEMM P., HAUSER H., HEGENSCHEID K., OELTZE S., TOENNIES K., VÖLZKE H.: Visual analytics of image-centric cohort studies in epidemiology. In *Visualization in Medicine and Life Sciences III*. Springer, 2016, pp. 221–248. 2

[PMS*03] PLAISANT C., MUSHLIN R., SNYDER A., LI J., HELLER D., SHNEIDERMAN B.: Lifelines: using visualization to enhance navigation and analysis of patient records. In *The Craft of Information Visualization*. Elsevier, 2003, pp. 308–312. 2

[PWR*11] POHL M., WILTNER S., RIND A., AIGNER W., MIKSCH S., TURIC T., DREXLER F.: Patient development at a glance: An evaluation of a medical data visualization. In *IFIP Conference on Human-Computer Interaction* (2011), Springer, pp. 292–299. 2

[RCMA*18] RAIDOU R., CASARES-MAGAZ O., AMIRKHANOV A., MOISEENKO V., MUREN L. P., EINCK J. P., VILANOVA A., GRÖLLER M. E.: Bladder runner : Visual analytics for the exploration of RT-induced bladder toxicity in a cohort study. *Computer Graphics Forum 37*, 3 (jun 2018), 205–216. 2

[RF14] RATWANI R. M., FONG A.: Connecting the dots: Leveraging visual analytics to make sense of patient safety event reports. *Journal of the American Medical Informatics Association* (oct 2014). 5

[RMM01] RICCIARDI T. N., MASARIE F. E., MIDDLETON B.: Clinical benchmarking enabled by the digital health record. *Studies in Health Technology and Informatics 84*, Pt 1 (2001), 675. 3

[RPOC18] RAJABIYAZDI F., PERIN C., OEHLBERG L., CARPENDALE S.: Personal patient-generated data visualizations for diabetes patients. In *IEEE VIS 2018 Posters* (2018). 2

[RSN*19] ROGERS J., SPINA N., NEESE A., HESS R., BRODKE D., LEX A.: Composer—visual cohort analysis of patient outcomes. *Applied Clinical Informatics 10*, 02 (mar 2019), 278–285. 2

[RWA*13] RIND A., WANG T. D., AIGNER W., MIKSCH S., WONGSUPHASAWAT K., PLAISANT C., SHNEIDERMAN B., ET AL.: Interactive information visualization to explore and query electronic health records. *Foundations and Trends® in Human–Computer Interaction 5*, 3 (2013), 207–298. 2

[RYS16] RINCON A. L., YAMASAKI H., SHIMODA S.: Design of a video game for rehabilitation using motion capture, EMG analysis and virtual reality. In *International Conference on Electronics, Communications and Computers (CONIELECOMP)* (2016), IEEE, pp. 198–204. 2

[SCB*19] SARIKAYA A., CORRELL M., BARTRAM L., TORY M., FISHER D.: What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (Jan. 2019), 682–692. 5

[SCGM00] STASKO J., CATRAMBONE R., GUZDIAL M., MCDONALD K.: An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies 53*, 5 (nov 2000), 663–694. 5

[Shn92] SHNEIDERMAN B.: Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG) 11*, 1 (1992), 92–99. 5

[SMB*10] STEENWIJK M. D., MILLES J., BUCHEM M., REIBER J., BOTHA C. P.: Integrated visual analysis for heterogeneous datasets in cohort studies. In *IEEE VisWeek Workshop on Visual Analytics in Health Care* (2010), vol. 3. 2

[SO08] STOKES E. K., O'NEILL D.: Use of outcome measures in physiotherapy practice in Ireland from 1998 to 2003 and comparison to Canadian trends. *Physiotherapy Canada 60*, 2 (2008), 109–116. 2

[Sto11] STOKES E. K.: *Rehabilitation outcome measures*. Churchill Livingstone, jan 2011. 1, 2

[SWH14] SATYANARAYAN A., WONGSUPHASAWAT K., HEER J.: Declarative interaction design for data visualization. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (2014), ACM, pp. 669–678. 5

[WBWK00] WANG BALDONADO M. Q., WOODRUFF A., KUCHINSKY A.: Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (2000), ACM, pp. 110–119. 9

[WG11] WONGSUPHASAWAT K., GOTZ D.: Outflow: Visualizing patient flow by symptoms and outcome. In *IEEE VisWeek Workshop on Visual Analytics in Healthcare, Providence, Rhode Island, USA* (2011), American Medical Informatics Association, pp. 25–28. 2

[WGGP*11] WONGSUPHASAWAT K., GUERRA GÓMEZ J. A., PLAISANT C., WANG T. D., TAIEB-MAIMON M., SHNEIDERMAN B.: Lifeflow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 1747–1756. 2

[WPQ*08] WANG T. D., PLAISANT C., QUINN A. J., STANCHAK R., MURPHY S., SHNEIDERMAN B.: Aligning temporal data by sentinel events: discovering patterns in electronic health records. In *Proceedings of the SIGCHI Conference on Human Factors in computing systems* (2008), ACM, pp. 457–466. 2

[WPS*09] WANG T. D., PLAISANT C., SHNEIDERMAN B., SPRING N., ROSEMAN D., MARCHAND G., MUKHERJEE V., SMITH M.: Temporal summaries: Supporting temporal categorical searching, aggregation and comparison. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1049–1056. 2

[WS09] WONGSUPHASAWAT K., SHNEIDERMAN B.: Finding comparable temporal categorical records: A similarity measure with an interactive visualization. In *2009 IEEE Symposium on Visual Analytics Science and Technology* (2009), IEEE, pp. 27–34. 2

[ZGP15] ZHANG Z., GOTZ D., PERER A.: Iterative cohort analysis and exploration. *Information Visualization 14*, 4 (2015), 289–307. 2