**EHDViz: Clinical Dashboard Development Using Open Source Technologies**

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>BMJ Open</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>bmjopen-2015-010579</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Research</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>25-Nov-2015</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Badgeley, Marcus; Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Shameer, Khader; Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Glicksberg, Benjamin; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Tomlinson, Max; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Levin, Matthew; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
McCormick, Patrick; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Kasarskis, Andrew; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Reich, David; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Dudley, Joel; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness |
| **Primary Subject Heading**: | Health informatics |
| Secondary Subject Heading: | Communication, Health services research, Medical management |
| Keywords: | early warning systems, biomedical informatics, data visualization, clinical decision systems |
EHDViz: Clinical Dashboard Development Using Open Source Technologies

Marcus A Badgeley¹,²,#, Khader Shameer¹,²,#, Benjamin S Glicksberg², Max S Tomlinson¹,², Matthew A Levin³, Patrick McCormick³, Andrew Kasarskis², David L Reich³, Joel T Dudley¹,²,4*

¹Harris Center for Precision Wellness ²Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology; ³Department of Anesthesiology; 4Department of Population Health Science and Policy; Icahn School of Medicine at Mount Sinai, Mount Sinai Health System, New York City, New York, USA.

*Corresponding author: joel.dudley@mssm.edu

#Denotes equal contribution

Keywords: clinical dashboard, medical informatics, clinical decision support, real-time monitoring, big data, visualization
Abstract

Objective To design and prototype a clinical dashboard for integrated health assessment with interactive real-time data visualization of heterogeneous biomedical, clinical, and patient generated health data streams.

Materials and Methods We designed developed and implemented interactive web based clinical dashboards using the R language and the R/Shiny web server architecture called EHDViz. In three prototypes, we utilized an outpatient cohort of 14,221 patients, an inpatient cohort of 445 patients, and consumer activity monitors to demonstrate the contextual versatility of EHDViz. A custom web service function was designed to pull and integrate user defined data features in real-time from simulations of the clinical cohorts and fitness monitor.

Results Prototype web applications were developed for several scenarios. Users are able to select a set of health features or an ICD9 class for the most common corresponding health features to populate the main visualization panel. The system retrieves relevant health data to populate the main panel of the application. The source code for EHDViz and various visualization prototypes developed using EHDViz are available at http://ehdviz.dudleylab.org.

Conclusion We envisage that as an open-source visualization framework, EHDViz will be an indispensable toolkit for rapid design, development, and implementation of clinical data visualization.
STRENGTHS AND LIMITATIONS

- Developing clinical dashboard for visualizing clinical, and healthcare data require implementation of costly health information technology (HealthIT) projects, we aim to offer a complementary, open source technology based solution for design, development and implementation of clinical dashboards.
- Developing affordable and sustainable HealthIT solutions are critical to support affordable care organizations and improve the quality of healthcare care delivery.
- EHDViz is designed as a solution to address the challenge to develop clinical dashboards using open source technologies.
- Implementation and systematic comparison of clinical dashboards developed using EHDViz and other open access or commercial solutions are needed to understand various usability preferences of end users and HealthIT application developers.
INTRODUCTION

The dynamics of subclinical and symptomatic health features as a condition progresses or response to lifestyle and medical interventions vary greatly even among individuals with similar demographics, clinical profiles and disease burden. The timing and variety of lifestyle, medical, and procedural interventions a physician incorporates into a treatment plan depends largely on this disease course kinetics. Electronic Health Record (EHR) software is used to capture and record vitals and other parameters of health status longitudinally. However, the EHR data is often represented as tabular views or static text formats, which can fail to demonstrate the underlying trends in a patient’s disease progression. Decades of research have shown that graphical summaries of patient information provides both faster and more accurate interpretations.1–3 Reproducible studies have established that practicing clinicians are unlikely to adapt any information retrieval task that takes longer than 30 seconds.4,5 Therefore, with physicians being challenged with increasingly complex patient histories, we need to develop better ways of visualizing and interpreting EHR data, upon which physicians can base critical treatment decisions.

Electronic Health Data (EHD) includes the acquisition of physiological values, radiology reports, laboratory values, radiologic reports, physician consults, and clinically actionable genetic information in medical settings. This is further supplemented by information provided directly by patients, such as blood pressure and food logs, and continuous physiologic data from wearable devices. Remote monitoring of implanted device data streams such as cardioverter defibrillators has been implemented at some institutions6, leading to non-inferior outcomes7,8, cost savings8,9, and earlier identification of device malfunction.10 Clinical trials are also beginning to incorporate remote physiologic data collection including cloud computing of electrophysiological data11 and the Apple ResearchKit mobile studies12.
This expanding number of electronic health data streams contributes to increasing amount of big data in health care, a foundation for many prospective paradigm shifts in modern medicine. Recent reviews on the application of big data in healthcare identify actionability and distribution of data over several databases as key challenges to fulfilling big data’s potential. A patient’s electronic health data can be generated, aggregated and normalized using data from personal logs, consumer wearable devices, medical devices (e.g., continuous positive airway pressure pumps for sleep apnea), and hospital administration and operations data cannot be easily aggregated. Information retrieval systems that integrate multiple data streams have been developed or suggested for particular applications including antibiotic CDS, recording events during surgical operations, diabetic patient data collection programs, text-mining and concept extraction from clinical documentation using natural language processing, and consumer health devices using third-party software application programming interfacing services (Human API, Aqua.io, Vivametrica). As of yet, but there is no system that permits flexible retrieval and interactive visualization of EHR, medical, and patient collected data streams.

**Clinical dashboards using open source technologies:**

Clinical dashboards offers the healthcare provides with a birds eye view of variety of quality metrics, patient status, progress in cohort aggregation and other various patient safety and care-delivery quality metrics routinely assessed in the clinical setting for performance improvement for physicians, nurses and other healthcare providers. Clinical dashboards are often developed using commercial or custom tools, thus little to no interoperability with tools for statistical analysis, machine learning or predictive modeling. In this paper, we propose an open source clinical dashboard development approach called EHD Visualization framework (EHDViz). The clinical dashboard system is interactive and extensible and could use an clinical decision aid to empowers patients, biomedical informatics scientists, solution architects, health care executives, healthcare delivery management team and medical professionals to retrieve, integrate, and explore diverse healthcare data.
streams to assess patient health trends. The program and web app are implemented using the open source R programming language and the base implementation can be customized for disease specific, division-specific, or institutional applications. EHDViz also offers features for integration with risk prediction algorithms for patient stratification, data mining algorithms to utilize underlying data repositories to refine the user experience and automatically retrieve the most relevant data for a selected context. Integrating various risk assessment algorithms with traditional clinical dashboard style interface offers a powerful tool for clinicians to use EHDViz for more complex dashboard development projects. EHDViz was evaluated using an individual patient’s data including non-conventional data streams not captured in a clinical setting, and cohorts of 14,221 outpatients and 445 inpatients at Icahn School of Medicine at Mount Sinai (ISMMS), a hospital of the Mount Sinai Health System in New York City.

The organization of the remainder of this paper is as follows. We first review related works on EHR data visualization and our approach to developing an integrative visualization tool and web platforms. We then discuss several examples of use-case scenarios to demonstrate the versatility and extensibility of EHDViz, and discuss the novelty of this system over existing platforms. Finally, we conclude by summarizing the main accomplishments of this work and discuss opportunities for further work.

**Related Work:**

Visual descriptions of the health status of patients in clinical settings have been a challenging problem since the introduction of computer programs for care management. At that time, the principal limitations were resource availability including appropriate graphical engines for rendering and specialized hardware to visualize patient status using computer programs. Decades later, visualization of clinical information and communicating health, disease or risk status of a patient using visual cues remains as an emerging challenge in the current era of data-driven medicine. Visualization tools can track the physiological status of patients,
biochemical variations, and quantification of biomarkers; such visualization aids were part of particular medical devices designed to monitor one or more specific physiological variables, such as heart rate and pulse rate, both within and external to EHRs. Efficient tools, algorithms, and risk prediction models are now required to for visual communication of clinical information to manage the high volume of biomedical and healthcare data in hospital setting. Integrating such visualization tools with predictive models and risk estimation tools could support accelerated patient stratification for improved care.

**Clinical data visualization:**

Designing visual tools to graphically explain risk scores and predictive models would help to accelerate patient risk stratification for improved care. There are a variety of technical challenges for integrated visualization of multiple clinical visualization tools. Interoperability of EHR applications and data feeds from medical devices remains a major challenge. Vendor standards also hinder the integration of diverse data elements to a common platform. Often data feeds need extensive quality control, normalization or other preprocessing procedures before the utilization in risk scoring engines. Visualization tools are currently available for effective integration of actionable information in workflow of clinical care pathways. Deng et al used a tag-cloud from radiology reports, pathology reports, and surgical reports to summarize unstructured patient data.\textsuperscript{20} Recent platforms, such as HARVEST\textsuperscript{21} offer web-based infrastructure for integrating, discovering, and reporting data, but are restricted to visualization of data captured in the warehouse. Visualization also plays a key role in shared-decision making\textsuperscript{21,22} where a care-provider and patient participate in a discussion regarding therapeutic strategies or clinical pathways that are supported with a variety of tools including visualization tools. Shared decision making is currently used in the treatment and management of cardiovascular diseases\textsuperscript{23,24}, diabetes\textsuperscript{25–28}, and osteoporosis\textsuperscript{29}. Irrespective of the medical specialty, visualization improves the ability to understand trends in a patient’s health and the effects of interventions over time.
Visualization of risk estimation in clinical setting:

Inpatient adverse events including cardiac arrests and unanticipated Intensive Care Unit (ICU) transfers and death are frequently preceded by slow and progressive physiological decompensation, with antecedents retrospectively identified in 79%, 55%, and 54% of cases, respectively. Failure to recognize and respond to signs of deterioration include infrequent or incomplete vital sign assessments, poor design of vital sign charts, and poor accuracy of ‘track-and-trigger’ systems. Several single-parameter and multi-parameter risk scoring methods have proposed to implement a “track-and-trigger” method of alerting for patients in clinical wards within 24 hours of an adverse event for accelerated clinical intervention aids.

The most established methods are based on vital signs and neurologic status, including Modified Early Warning System (MEWS), Standardized Early Warning Systems (SEWS) and National Early Warning Score (NEWS) that differ with respect to the inclusion of oxygen saturation and supplemental oxygen and the weight of different features. When assessed retrospectively, these vital based systems have AUROCs of 0.76-0.83 for cardiac arrests, 0.73-0.77 for ICU transfers, and 0.87-0.88 for mortality, effective for triggering follow-up evaluation. Implementations of these warning systems have required that staff perform rounds and fill out paper sheets or electronically enter the vital signs. Reports of real-time EHR information retrieval based implementations of early warning systems have had some success in reducing adverse events in RCTs and cross-over trials, although the risk models have limitations due to the limited physiological feature space. Algorithms developed using the entire set of discrete health features in the clinical data warehouse have incorporated significantly predictive laboratory values, physician orders, and medications. When assessed retrospectively they consistently outperform the vital-constrained approaches. EHDViz provides an EHR-agnostic visualization framework that can be implemented in real-time to assist in identifying patients with decompensating physiology with a rich, visual aid.
METHODS

The client-server architecture of EHDViz is provided in Figure 1 and source code is available from http://ehdviz.dudleylab.org/. We designed multiple clinical dashboard prototypes using EHDViz, coupled with statistical analytics and visualization using R language (R version 3.0.2 (2013-09-25)), and EMR-agnostic web server architecture was implemented using Shiny (http://shiny.rstudio.com/). The Shiny server architecture was used as it can be deployed over multiple desktop and server environments and can be distributed as convenient software modules.

Clinical dashboard prototypes developed using EHDViz:

To test and evaluate the technical challenges in developing and deploying a dynamic, knowledge-based, real-time visualization tool we designed a prototype web application using R language and the R/Shiny web server architecture as outlined above. Clinical data was obtained from two cohorts and simulated for online demonstrations, data from fitness monitoring devices was aggregated using internal API of the device manufacturer, and a custom web service function was designed to pull and integrate user defined data features in real-time.

RESULTS

Dashboard 1: Quantified Self time series data integration:

The quantified self movement involves individuals tracking many types of biometric data in order to gain insight and awareness of their health.41 Patients are also increasingly able to access and control their clinically collected health data.42,43 Our first demonstration addresses the quantified self challenge of integrating and visualizing time series health data from multiple data sources. The example in Figure 3a demonstrates the integration of an individual patient’s electronic health data sources. For this example, the patient has 3 primary sources of health data: 1) clinical data from outpatient visits, 2) continuous activity data from a wearable device (Fitbit; San Francisco, CA) and 3) a self-recorded blood pressure log. The clinical data from outpatient wellness visits was simulated by randomly sampling
aggregated physiologic and lab values from 14,221 patients in an ISMMS outpatient cohort. The continuous activity data was scraped from one of the author’s (MAB) wearable device using the API at an interval of 15 minutes. The blood pressure log is simulated as weekly measurements from normal distributions $N(130,15)$ and $N(85,10)$.

The user interface features a main panel with sparklines for each health feature and a sidebar with widgets for the user to select the health features of interest. In this example, there is a checkbox group for each data source: (1) EMR, (2) data from fitness monitoring device, and (3) personal log. The user can select any combination of health features to be displayed. The main panel displays a sparkline stack with selected health features sorted to correspond to the sidebar. Minimums and maximums are highlighted with red and blue dots, respectively. In this application, the data source that updates most frequently was from the wearable device collected in 15-minute intervals; the application was programmed to auto refresh every 15 minutes to retrieve new data.

**Dashboard 2: High-velocity data visualization in clinical setting:**

Next we demonstrate the retrieval of continuous data contained in a collection of patient’s EMRs during an inpatient stay, where data will be much more dynamic than in the previous outpatient example. This implementation was tested with both a cohort of 445 inpatients at ISMMS with clinical labs recorded throughout their encounter (Figure 2a, 2b), as well as with simulated patients (Figure 3b).

For the cohort patient data the user can use the sidebar is used select a patient and the date range of interest, and then relevant information is retrieved from the EHR or data warehouse throughout the encounter (Figure 2a). Within a single encounter, several hospital units including the emergency department, inpatient units, surgical suites, and ICUs see the patient. This information of admission, transfer and discharge was integrated into EHDViz by coloring the
background by location to easily associate health status changes with the patient transfers (Figure 2b).

For the simulated patients, we randomly retrieved data from the EMR for a cohort of 14,221 patients to populate each of the 375 discrete continuous health features contained in the EMR. For each of the 7,000 unique diagnoses, we pooled corresponding patient data and found the most frequently measured health features for each ICD-9 class (See various examples of input formats and sample file at http://ehdviz.dudleylab.org/help.html#introduction). The simulated patient dashboard (Figure 3b) allows the user to select a patient and an ICD9 class from the drop-down menus in the side panel, and then populates the main panel with the most common health features measured for that ICD9 class. The list of health features corresponding to the selected ICD9 class is additionally displayed as a checkbox group in the side panel so the user can further refine the displayed feature set. This enables the user to rapidly retrieve and assess trends in the most relevant biomarkers. We also provide a demonstration at http://ehdviz.dudleylab.org/providers/full that allows a key word based search and multi-selection of all 375 health features to make customized dashboards. Real-time displays were also designed from the simulated data, demonstrated at http://ehdviz.dudleylab.org/providers/real-time.

Dashboard 3: Visualizing clinical information on a population scale:

The examples in Figure 2c, 2d, and 3c demonstrate the use of EHDViz for patient safety visualization and cohort analysis. This dashboard provides risk estimation visualization for users to track all patients simultaneously in a unit, which facilitates identifying atypical and destabilizing features to trigger interventions. Patient vital signs were retrieved from the EMR warehouse from 445 inpatients and processed to calculate the MEWS risk score. Figure 2c and 2d show the dashboard for monitoring these patients MEWS health stability. The user can select the clinical unit of interest with the drop-down menu, and sparklines with MEWS scores are displayed for each patient in the unit with alert triggering.
thresholds displayed for reference. When there are multiple patients in the unit, MEWS scores are colored by patient (Figure 2d). Data for online demonstrations was simulated as discussed in scenario 2 and the covariates “location” and “patient” was switched from a data coloring covariate to a user filtering covariate and vice versa for use in a cohort application. As shown in Figure 3c, a user can select the clinical unit of interest and text search 375 health features, and the main panel will display the values of these features for all the patients in the selected unit, colored by patient. This design allows rapid evaluation of any health feature or calculated status score for a population of patients. We provide online demonstrations of ICD9 class based feature selection at ehdviz.dudleylab.org/cohort/ICD9.html and a real time monitoring dashboard at ehdviz.dudleylab.org/cohort/real-time.

Discussion:

A physician’s treatment plan for a patient depends on a number of quantifiable factors that can be collected from a number of sources. Physicians can collect data from the Electronic Health Record (EHR), electronic patient diaries, fitness trackers, and the patient’s recollections of medical history. However, in most presentations this data overwhelms physicians instead of guiding decision making. Real-time clinical monitoring and automated alerting provides better tools to that have the capacity to improve patient safety, improve clinical outcomes and quality of healthcare in clinical setting. Tools are available to monitor infection, adverse events and customized tools targeting specific needs of the clinical unit including operating rooms or ICUs. Developing a unified visualization tool that can provide an overview of patient by integrating different vital signs remains as an open challenge. EHDViz offers a solution to this; as an open source data visualization framework capable of real-time data visualization. EHDViz unifies heterogeneous biomedical and healthcare data integration and utilize R, a popular and preferred programming language for big data analytics. R is typically used for desktop or client-cluster based visualization models. Here, we have used a static R visualization package and
rendered it as a real time data visualization engine. Close integration with R also enables visualization of advanced analytics and predictive modeling using EHDViz.

Conclusions:
The introduction of the affordable care act and the necessity for health information technologies to be more efficient and sustainable calls for the adoption of cutting edge technologies with limited spending. Open source technologies offers an alternative option for healthcare providers to design, develop and deploy very efficient clinical dashboards with no cost for software, thus reduce the healthcare spending. We designed EHDViz as an open source clinical dashboard and visualization framework to integrate with existing commercial and open source EHRs for heterogeneous biomedical and healthcare data visualization. We envisage that design and development of real-time patient status tools coupled with risk estimation that could enhance the quality of health-care delivery and reduce adverse patient outcomes in clinical settings. As hospitals and healthcare systems are emerging as learning health systems, in which data capture, smart clinical dashboards, and adaptive data visualization will play an integral role in managing the patient population.
Contributorship statement:
MAB, KS, BG, and MST contributed to the data integration, software package implementation, and web server development. KS, MAL, PM, AK, DLR and JTD formulate visualization strategy and designed illustrative examples. JTD contributed to the overall planning of the project, the development of an extensible software package for clinical dashboard development and the manuscript. All authors have contributed to the writing and compilation of the final manuscript. All authors approve their contributions and the final draft of the manuscript.

Competing interests:
MAB has received consulting fees from MetaMed; KS: None declared; BG: None declared; JTD has received consulting fees or honoraria from Janssen Pharmaceuticals, GSK, AstraZeneca, and Hoffman-La Roche. JTD holds equity in NuMedii Inc, Ayasdi, Inc. and Ontomics, Inc. No writing assistance was utilized in the production of this manuscript.

Funding:
Intramural funding to Dudley laboratory from Mount Sinai Health System funded study

Data sharing statement
Source code and data is available from the companion website: 
http://ehdviz.dudleylab.org Software repository can be downloaded to local development workstations and servers and clinical dashboards can be developed using the extensible software framework.

Acknowledgements
Authors would like to thank Adel Bassily-Marcus, Bruce Darrow, Kash Patel, Alan Moskowitz, Annetine Gelijns and Patricia Kovatch. Authors would like to thank Icahn Institute for Genomics and Multiscale Biology for infrastructural support.
References:


8. Raatikainen MJP, Uusimaa P, van Ginneken MME, Janssen JPG, Linnaluoto M. Remote monitoring of implantable cardioverter defibrillator patients: a safe,


**Figures:**

**Figure 1:** Client-server architecture of EHDViz

**Figure 2:** Different scenarios of implementing a visual aid for Modified Early Warning Score (MEWS) using EHDViz framework. a) Visualization of a single patient; b) Visualization of a single patient layered on patient admission, discharge and transfer data; c) Visualizing trends of MEWS in different in-patient units; d) visualizing multiple patients in a same unit

**Figure 3:** Example use cases and features of for EHDViz in the setting of a) quantified self - 1: user management; 2: dynamic selection; 3: integration with data streams; 4: integration with manual data input. b) Clinical evaluation – 1: selection of individuals; 2: options to control visual layouts 3: integration with ICD-9 codes. c) Group health management – 1: visualization of data from floor using admission-discharge-transfer data; 2: dynamic control of visualization; 3: real-time user interaction
Data aggregation using web-services

Quality control, preprocessing and normalization

Vendor agnostic visualization on web browsers
a) Quantified self

Welcome back, User

EMR Data
- Blood Pressure
- Blood Glucose
- Heart Rate
- Temperature
- Weight
- BMI

Fitbit Data
- Steps
- Distance
- Calories
- Heart Rate

Personal Logs
- Diaries
- Notes

b) Clinical evaluation

Mental Health
- Mood
- Sleep

Physical activity
- Steps
- Distance
- Calories

c) Population health management

Diabetes
- Glucose levels
- Blood pressure
- Heart rate

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml
# EHDViz: Clinical Dashboard Development Using Open Source Technologies

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>BMJ Open</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>bmjopen-2015-010579.R1</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Research</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>27-Jan-2016</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Badgeley, Marcus; Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness Shameer, Khader; Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness Glicksberg, Benjamin; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness Tomlinson, Max; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness Levin, Matthew; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness McCormick, Patrick; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness Kasarskis, Andrew; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness Reich, David; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness Dudley, Joel; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness</td>
</tr>
<tr>
<td>Primary Subject Heading:</td>
<td>Health informatics</td>
</tr>
<tr>
<td>Secondary Subject Heading:</td>
<td>Communication, Health services research, Medical management</td>
</tr>
<tr>
<td>Keywords:</td>
<td>early warning systems, biomedical informatics, data visualization, clinical decision systems</td>
</tr>
</tbody>
</table>
EHDViz: Clinical Dashboard Development Using Open Source Technologies

Marcus A Badgeley¹,²,#, Khader Shameer¹,²,#, Benjamin S Glicksberg², Max S Tomlinson¹,², Matthew A Levin³, Patrick McCormick³, Andrew Kasarskis², David L Reich³, Joel T Dudley¹,²,4*

¹ Harris Center for Precision Wellness ² Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology; ³Department of Anesthesiology; 4Department of Population Health Science and Policy; Icahn School of Medicine at Mount Sinai, Mount Sinai Health System, New York City, New York, USA.

*Corresponding author: joel.dudley@mssm.edu

#Denotes equal contribution

Keywords: clinical dashboard, medical informatics, clinical decision support, real-time monitoring, big data, visualization
Abstract

**Objective:** To design, develop and prototype a clinical dashboard for integrated health assessment with interactive real-time data visualization of heterogeneous biomedical, clinical, and patient-generated health or wellness data streams.

**Materials and Methods:** We designed, developed, and implemented interactive web-based visualization dashboards for medical data using the R language and the R/Shiny web server architecture. The electronic healthcare data visualization toolkit (EHDViz) and several example clinical dashboard development implementations developed using EHDViz are provided in the public domain. EHDViz leverages widely used R packages for data cleaning, extraction, transformation, loading (ETL) tasks, and delivering high-quality, scalable visualizations over the web. We have provided several use cases to illustrate the utility of EHDViz in a clinical setting and as a visualization aid for healthcare delivery.

**Results:** Prototype web applications were developed for several scenarios. For implementing three different prototypes, we utilized an outpatient cohort (n=14,221), an inpatient cohort (n=445), and wellness data from fitness activity monitor worn by a single individual (n-of-1), to demonstrate the contextual versatility of EHDViz. A custom web service function was designed to pull and integrate user-defined data features in real-time from simulations of the clinical cohorts and fitness monitors. Users customize the visualizations using health features or an ICD-9 class to populate the visualization panel in the dashboard. The system retrieves relevant health data to populate the main panel of the application. The source code for EHDViz and various visualization prototypes developed using EHDViz are available at http://ehdviz.dudleylab.org.

**Conclusion:** Collaborative data visualizations, risk estimation algorithms, acuity status and wellness trend predictors, and complex disease indicators are important components of implementing big-data driven medicine. As an open-source
visualization framework, EHDViz aims to be a useful toolkit for rapid design, development, and implementation of clinical data visualization dashboards.
STRENGTHS AND LIMITATIONS

- Currently, design, development, and deployment of the clinical dashboard for visualizing heterogeneous clinical and healthcare data requires implementation of costly health information technology (HealthIT) projects.
- We aim to offer a complementary, cost-effective, and open source technology based solution for design, development, and implementation of clinical dashboards.
- Developing affordable and sustainable HealthIT solutions are critical to support affordable care organizations and improve the quality of healthcare delivery.
- Implementation and systematic comparison of clinical dashboards developed using EHDViz and other public access or commercial solutions are needed to understand various usability preferences of end users and provide easy options for HealthIT application developers.
- EHDViz is designed to address the need to develop clinical dashboards using open source technologies while also enabling improved patient engagement and simulation-based learning.
INTRODUCTION

The subclinical features and primary diagnosis symptoms vary as a condition progresses and can change due to lifestyle and medical interventions. These variations deviate greatly even among individuals with similar demographics, clinical profiles, and disease burdens. The specific implementation of lifestyle, medical, and procedural interventions a physician incorporates into a treatment plan depends largely on the disease course. Electronic Health Record (EHR) software is widely used to longitudinally capture and record vital signs, medications, laboratory values, diagnostic reports, fluid inputs/outputs, mental states, patient transfers, and other health status parameters. However, the EHR software often presents data with tabular views or static text formats, which illuminates neither the underlying trends in a patient’s disease progression nor the similarities among patient trends within a given department. Decades of research have shown that graphical summaries of patient information provide faster and more accurate medical diagnoses, thus improving the healthcare quality.\(^1\)\(^2\)\(^3\) Reproducible studies demonstrate that practicing clinicians are unlikely to adopt any information retrieval task that takes longer than 30 seconds.\(^4\)\(^5\) Therefore, with physicians being challenged with increasingly complex patient histories, there is an unmet need to develop better ways of visualizing and interpreting EHR data, upon which physicians can base critical treatment decisions. Electronic Health Data (EHD) includes the acquisition of physiological values, diagnostic reports (radiology reports), laboratory values, pathology reports (biopsy report), physician consults, and clinically actionable genetic information.\(^6\)\(^7\)\(^8\) In recent years, this is further supplemented by information that patients provide directly, such as blood pressure and food logs, and by continuous physiologic data from wearable devices from patient portals\(^9\) or mobile phones.\(^3\) Some institutions have implemented remote monitoring of patients using implanted devices, like the implantable cardioverter-defibrillator (ICD), as well as augmented clinical management using data streams from health monitoring devices,\(^10\) leading to non-inferior outcomes,\(^11\)\(^12\) cost savings,\(^12\)\(^13\), and earlier identification of device malfunction.\(^14\) Clinical trials are
incorporating remote monitoring devices, including devices capable of collecting physiological data and cloud computing electrophysiological data\textsuperscript{15}. As an example, several ongoing clinical trials make use of Apple ResearchKit for evaluating patients with asthma\textsuperscript{16}, cardiovascular disease, diabetes, and Parkinson’s (See: http://researchkit.org/), and some efforts implement Mhealth based solutions engage patients and visualize data (AppCore: https://github.com/researchkit/AppCore).

The expanding number of electronic health data streams contributes to an increasing amount of big data in health care, which lays the foundation for many prospective paradigm shifts in modern medicine. Recent reviews on the application of big data to healthcare identify actionability and decentralized data\textsuperscript{17} as key challenges to fulfilling big data’s potential.\textsuperscript{18} 19 A patient’s electronic health data can be aggregated and normalized using data from personal logs, health monitoring or fitness devices, and medical devices (e.g., continuous positive airway pressure pumps for sleep apnea). To get a complete picture of health, the data from hospital administration and operations data can also be aggregated. Information retrieval systems that integrate multiple data streams have been developed or suggested for particular applications including antibiotic clinical decision systems (CDS)\textsuperscript{3}, recording events during surgical operations\textsuperscript{20}, diabetic patient data collection programs\textsuperscript{21}, text-mining, concept extraction from clinical documentation using natural language processing \textsuperscript{19}, and wearable devices using software application programming interfacing (API) services (Human API, Aqua.io, Vivametrica) that allow communication between health monitoring devices and provider EHR databases using a secure programmatic data access protocol. As of yet, there is no system that permits flexible retrieval and interactive visualization of EHR, medical, and patient generated data streams.

**Clinical data visualization:**

Visual descriptions of the health status of patients in clinical settings have been a challenging problem since the introduction of computer programs for care
management. At that time, the principal limitations were resource availability including appropriate graphical engines for rendering and specialized hardware to visualize patient status using computer programs.\textsuperscript{2} Decades later, visualization of clinical information and communicating health, disease or risk status of a patient using visual cues remains as an emerging challenge in the current era of data-driven medicine.\textsuperscript{22, 23} Visualization tools can track the physiological status of patients, biochemical variations, and quantification of biomarkers; such visualization aids were part of particular medical devices designed to monitor one or more specific physiological variables, such as heart rate and pulse rate, both within and external to EHRs. Efficient tools, algorithms, and risk prediction models are now required to for visual communication of clinical information to manage the high volume of biomedical and healthcare data in hospital setting. Integrating such visualization tools with predictive models and risk estimation tools could support accelerated patient stratification for improved care.

**Visualizing healthcare data using clinical dashboards:**

Clinical dashboards offers the healthcare provides with a birds eye view of variety of quality metrics, patient status, progress in cohort aggregation and other various patient safety and care-delivery quality metrics routinely assessed in the clinical setting for performance improvement for physicians, nurses and other healthcare providers. Clinical dashboards are often developed using commercial or custom tools, thus little to no interoperability with tools are available for statistical analysis, machine learning or to integrate with predictive modeling that can aid tasks including acuity prediction, readmission evaluation and to assess risk of readmissions. Designing visual tools to graphically explain risk scores and predictive models would help to accelerate patient risk stratification for improved care. There are a variety of technical challenges for integrated visualization of multiple clinical visualization tools. Interoperability of EHR applications and data feeds from medical devices remains a major challenge. Vendor standards also hinder the integration of diverse data elements to a common platform. Often data feeds need extensive quality control, normalization or other preprocessing
procedures before the utilization in risk scoring engines. Visualization also plays a key role in shared-decision making where a care-provider and patient participate in a discussion regarding therapeutic strategies or clinical care delivery pathways that are supported with a variety of tools including visualization tools. Shared decision making is currently used in the treatment and management of cardiovascular diseases, diabetes, and osteoporosis. Irrespective of the medical specialty, visualization improves the ability to understand trends in a patient’s health and the effects of interventions over time.

**Visualization and forecast of risk estimations in clinical setting:**

Adverse events during hospitalization including hospital acquired infections (HAI) or hospital acquired conditions (HAC) including falls, cardiac arrests and unanticipated intensive care unit (ICU) transfers and death are frequently preceded by several useful and predictors features that can be used for accelerate triaging to improve the care delivery. For example, slow and progressive physiological decompensation was identified in cardiac arrests (79%), unexpected ICU transfers (55%), and death during hospitalization (54%) in a retrospective study that compared cohorts from different countries. Failure to recognize and respond to signs of deterioration includes infrequent or incomplete vital sign assessments, poor design of vital sign charts, and poor accuracy of 'track-and-trigger' systems.

Several single-parameter and multi-parameter risk scoring methods have proposed to implement a “track-and-trigger” method of alerting for patients in clinical wards within 24 hours of an adverse event for accelerated clinical intervention aids. The most established methods are based on vital signs and neurologic status, including Modified Early Warning System (MEWS), Standardized Early Warning Systems (SEWS) and National Early Warning Score (NEWS) that differ with respect to the inclusion of oxygen saturation and supplemental oxygen and the weight of different features. When assessed retrospectively, these vital based systems have area under receiver operator curves (AUROCs) of 0.76-0.83 for cardiac arrests, 0.73-0.77 for ICU transfers, and 0.87-0.88 for mortality, effective for triggering follow-up evaluation. Implementations of these warning systems have required that staff
perform rounds and fill out paper sheets\textsuperscript{36} or electronically enter\textsuperscript{37} the vital signs. Reports of real-time EHR information retrieval based implementations of early warning systems have had some success in reducing adverse events in RCTs\textsuperscript{38} \textsuperscript{39} and cross-over trials\textsuperscript{40}, although the risk models have limitations due to the limited physiological feature space. Algorithms developed using the entire set of discrete health features in the clinical data warehouse have incorporated significantly predictive laboratory values, physician orders, and medications. When assessed retrospectively they consistently outperform the vital-constrained approaches\textsuperscript{41} \textsuperscript{42} \textsuperscript{43}.

In this paper, we propose the design, development and implementation of a clinical dashboard development framework by leveraging open source technologies. The Electronic Health Data Visualization framework (EHDViz) is an interactive and extensible framework implemented using the popular statistical computing language and could use as a clinical decision aid to empowers care patients, biomedical informatics scientists, solution architects, health care executives, healthcare delivery management team and medical professionals to retrieve, integrate, and explore diverse healthcare data streams to assess patient health trends. EHDViz provides an EHR-agnostic visualization framework that can be implemented in real-time to assist in identifying patients with decompensating physiology with a rich, visual aid.

**METHODS**

**Description of EHDViz Framework:**
EHDViz is a software framework designed to generate interactively, web-based healthcare data visualization using various R packages (R language; R version 3.0.2; 2013-09-25). An infographic of the client-server architecture of EHDViz is provided in Figure 1. We compiled various packages to organize a unified software framework for data input/output operations, data management of healthcare data, data cleansing from diverse sources, generation of plots and statistical analyzes. Data cleansing was performed using reshape2 package (https://cran.r-
EHDViz uses the packages `ggplot2` ([http://ggplot2.org](http://ggplot2.org)) and `gridExtra` ([https://cran.r-project.org/web/packages/gridExtra/index.html](https://cran.r-project.org/web/packages/gridExtra/index.html)) for developing plots. A custom, the plot-stitching algorithm is developed to combine separate plots and visualize as a continuous, real-time visualization. R plots can be generated and visualized using PDF viewers, image viewers, and web-browsers. We used the web server implementation using `R/Shiny` to deploy the plots created as part of EHDViz framework. The `Shiny server` architecture ([https://github.com/rstudio/shiny-server](https://github.com/rstudio/shiny-server)) was used as it can be implemented over multiple desktop and server environments and can be distributed as suitable software modules. Data from wearable devices are compiled using the device specific API for Fitbit; wearable-specific API offers a secure way to collect and aggregate data generated by the personal fitness monitoring devices. The package `fitbitScraper` ([https://cran.r-project.org/web/packages/fitbitScraper/index.html](https://cran.r-project.org/web/packages/fitbitScraper/index.html)) was used to extract the data from the wearable device.

**Data handling in EHDViz:**

Various biomedical and healthcare data types can be indexed various clinical dictionaries including International Statistical Classification of Diseases and Related Health Problems (ICD-9: [http://www.who.int/classifications/icd/en/](http://www.who.int/classifications/icd/en/)) codes to define specific disease terms of patients as part of diagnoses. Current Procedural Terminology (CPT: [http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/about-cpt.page](http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/about-cpt.page)) codes are used to identify patients undergoing specific clinical procedures. National Drug Codes (NDC) and RxNorm are used to parse and normalize medication data as part of the data aggregation methods in EHDViz. (NDC: [http://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm](http://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm); RxNorm [https://www.nlm.nih.gov/research/umls/rxnorm/](https://www.nlm.nih.gov/research/umls/rxnorm/)). EHDViz can also handle data from clinical operations and administrative databases including patient-transfer data (from the emergency department to surgery to ward and discharge) to define,
group or aggregate patient cohorts match a given patients in the database to visualize trends.

**Input and output specifications of EHDViz:**

EHDViz can handle data in tab-delimited file format (.tsv) or comma-delimited file format (.csv). Data can also be extracted from various other formats and database using native R packages. For example, EHDViz can extract data from Excel files (xlsx: https://cran.r-project.org/web/packages/xlsx/index.html) or relational database systems that conforms to Open Database Connectivity (ODBC) standards (RODBC: https://cran.r-project.org/web/packages/RODBC/index.html), Java Database Connectivity (JDBC) (RJDBC: https://cran.r-project.org/web/packages/RJDBC/index.html), MySQL (RMySQL: https://cran.r-project.org/web/packages/RMySQL/index.html) or modern, NoSQL database systems such as MongoDB (rmongodb https://cran.r-project.org/web/packages/rmongodb/index.html). Various examples of input formats and sample files are provided at the URL: http://ehdviz.dudleylab.org/help.html#introduction. The data gathered from flat files, or database connections will be used as input for EHDViz dashboards. The diverse set of data from various sources after parsing, quality control, and normalization can be loaded to the visualization templates of an individual project. The output of EHDViz is the customized visualization dashboards. EHDViz output can be rendered using a standard, modern web browser that supports HTML5 and responsive web development standards.

**Clinical dashboards developed using EHDViz:**

To test and evaluate the technical challenges in developing and deploying a dynamic, knowledge-based, real-time visualization tool we designed a prototype web application using R language and the R/Shiny web server architecture as outlined above. Prototype dashboards are developed using three different data sets: 1) data from a single patient (n-of-1) including non-conventional data streams not captured in a clinical setting was used to demonstrate quantified-self data
visualization. 2) Simulated cohorts of outpatients \((n=14,221)\) 3) simulated cohort of inpatients \((n=445)\). The data simulation was performed using a deidentified EHR data compiled at Icahn School of Medicine at Mount Sinai (ISMMS), a hospital of the Mount Sinai Health System in New York City. Data from fitness monitoring devices was aggregated using an API capable of patient specific data from the fitness monitoring device of a user, and a custom web service function was designed to pull and integrate user defined data features in real-time. Dashboards discussed in this manuscript is implemented on a web server

RESULTS

Availability: The source code of EHDViz and various clinical dashboard implementations are available from the URL: http://ehdviz.dudleylab.org/.

Dashboard 1: Visualizing Quantified Self, time series data feeds:

The quantified-self movement involves an increasing interest in individuals and patient communities in tracking many types of biometric data in order to gain insight and awareness of their health.\(^4\) Patients are also increasingly able to access and control their clinically collected health data.\(^45\)\(^46\) Our first demonstration addresses the challenge in quantified-self area of integrating and visualizing time series health data from multiple data sources. The example in Figure 2 demonstrates the integration of an individual patient’s electronic health data sources. For this example, the patient has 3 primary sources of health data: 1) clinical data from outpatient visits, 2) continuous activity data from a wearable device (Fitbit; San Francisco, CA) and 3) a self-recorded blood pressure log. The clinical data from ambulatory visits were simulated by randomly sampling aggregated physiologic and lab values from 14,221 patients in an ISMMS outpatient cohort. The continuous activity data was scraped from one of the author’s (MAB) wearable device using the API at an interval of 15 minutes. The blood pressure log is simulated as weekly measurements from normal distributions \(N(130,15)\) and \(N(85,10)\). The user interface features a main panel with ‘sparklines’ for each health
feature and a sidebar with widgets for the user to select the health features of interest. In this example, a checkbox is provided to group patients for each data source: (1) EMR, (2) data from fitness monitoring device, and (3) personal log. The user can select any combination of health features to be displayed. The main panel displays a stack of sparklines with selected health features sorted according to values selected in the sidebar. Minimums and maximums are highlighted with red and blue dots, respectively. In this application, the data source that updates most frequently was from the wearable device collected in 15-minute intervals; the application was programmed to auto refresh every 15 minutes to retrieve new data.

**Dashboard 2: High-velocity data visualization in clinical setting:**

Next we demonstrate the retrieval of continuous data contained in a collection of patient’s EMRs during an inpatient stay, where data will be much more dynamic than in the previous outpatient example. This implementation was tested with both a simulated cohort of 445 inpatients with clinical labs recorded throughout their encounter as well as with simulated data (Figure 3(a-d) and 4).

For the cohort patient data the user can use the sidebar is used select a patient and the date range of interest, and then relevant information is retrieved from the EHR or data warehouse throughout the encounter (Figure 3(a-d)). Within a single encounter, several hospital units including the emergency department, inpatient units, surgical suites, and ICUs see the patient. This information of admission, transfer and discharge was integrated into EHDViz by coloring the background by location to easily associate health status changes with the patient transfers (Figure 3(b)). For the simulated patients, we randomly retrieved data from the EMR for a simulated cohort of 14,221 patients to populate each of the 375 discrete continuous health features contained in the EMR.

For each of the 7,000 unique diagnoses, we pooled corresponding patient data and found the most frequently measured health features for each ICD-9 class. The simulated patient dashboard (Figure 4) allows the user to select a patient and an ICD9 class from the drop-down menus in the side panel, and then populates the main panel with the most common health features measured for that ICD9 class.
The list of health features corresponding to the selected ICD9 class is additionally displayed as a checkbox group in the side panel so the user can further refine the displayed feature set. This enables the user to rapidly retrieve and assess trends in the most relevant biomarkers. We also provide a demonstration at http://ehdviz.dudleylab.org/providers/full that allows a key word based search and multi-selection of all 375 health features to make customized dashboards. Real-time displays were also designed from the simulated data, demonstrated at http://ehdviz.dudleylab.org/providers/real-time.

**Dashboard 3: Visualization for population health management**

The examples in Figure 3c, 3d, and 5 demonstrate the use of EHDViz for developing visual aids for patient safety and cohort analysis. These dashboards provide risk estimation visualization for users to track all patients simultaneously in a unit, which facilitates identifying atypical and destabilizing features to trigger interventions. Patient vital signs were retrieved from the EMR warehouse from 445 inpatients and processed to calculate the MEWS risk score. Figure 3c and 3d show the dashboard for monitoring these patients MEWS health stability. The user can select the clinical unit of interest with the drop-down menu, and sparklines with MEWS scores are displayed for each patient in the unit with alert triggering thresholds displayed for reference. When there are multiple patients in the unit, MEWS scores are colored by patient (Figure 3d). Data for online demonstrations was simulated as discussed in scenario 2 and the covariates “location” and “patient” was switched from a data coloring covariate to a user filtering covariate and vice versa for use in a cohort application. As shown in Figure 5, a user can select the clinical unit of interest and text search different clinical parameters, and the main panel will display the values of these features for all the patients in the selected unit, colored by patient. This design allows rapid evaluation of various clinical features or predictors. Multiple values relevant to clinical manifestations of patient population can be compiled and new scores (for example MEWS) can be computed for a population of patients. Demonstrations of ICD9 class based feature selection at http://ehdviz.dudleylab.org/visualizations/Population_Management_ICD9/ and a
real time monitoring dashboard implemented using EHDViz is provided at the URL: http://ehdviz.dudleylab.org/visualizations/Population_Management_RealTime/

Discussion:
A physician’s treatment plan for a patient depends on a number of factors that can be collected from a different sources including medications, diagnosis and patients response to therapies or other interventions. Physicians can collect data from the Electronic Health Record (EHR), electronic patient diaries, fitness trackers, and the patient’s recollections of medical history. However, in most presentations this data overwhels physicians instead of guiding decision making. Real-time clinical monitoring and automated alerting provides better tools to that have the capacity to improve patient safety, improve clinical outcomes and quality of healthcare in clinical setting. Tools are available to monitor infection, adverse events and customized tools targeting specific needs of the clinical unit including operating rooms or ICUs. Developing a unified visualization tool that can provide an overview of patient by integrating different vital signs remains as an open challenge. EHDViz offers a solution to this; as an open source data visualization framework capable of real-time data visualization. EHDViz unifies heterogeneous biomedical and healthcare data integration and utilize R, a popular and preferred programming language for big data analytics. R is typically used for desktop or client-cluster based visualization models. Here, we have used a static R visualization package and rendered it as a real time data visualization engine. Close integration with R also enables visualization of advanced analytics and predictive modeling using EHDViz. The base implementation of EHDViz dashboards can be customized for disease specific, division-specific, or institutional applications. EHDViz also offers features for integration with risk prediction algorithms for patient stratification, data mining algorithms to utilize underlying data repositories to refine the user experience and automatically retrieve the most relevant data for a selected context. Integrating various risk assessment algorithms with traditional clinical dashboard style
interface offers a powerful tool for clinicians to use EHDViz for more complex dashboard development projects.

**Application of EHDViz in simulation-based medical education:**

Simulation-based learning is at the core of the pedagogical principles of modern medicine. Medical students, residents, and physicians extensively use EMR at the bedside during care delivery. EHDViz can leverage as a teaching aid capable of generating custom EHR instances and visualizations. Simulated EHR systems can be designed based on use-cases to evaluate a single patient or number of patients that an individual resident is managing on a floor or unit.

**Comparison with related healthcare data visualization applications:**

Multiple visualization tools are currently available for effective integration of actionable information in workflow of clinical care pathways. A systematic review of data visualization tools assessed multiple clinical data visualization tools. Tools like EventFlow\(^47\), LifeLines\(^48\), LifeLines2\(^49\), Visualization of Time-Oriented Records (VISITORS)\(^50\) and Dynamics Icon (DICON)\(^51\) are identified as major tools available for clinical data visualization and dashboard development. Deng et al. used a tag-cloud from radiology reports, pathology reports, and surgical reports to summarize unstructured patient data.\(^52\) Data visualization, such as HARVEST\(^24\) offer web-based infrastructure for integrating, discovering, and reporting data, but are restricted to visualization of data captured in the warehouse. Lifelines and Lifelines2 offers options to align, rank, and summarize temporal visualizations. LifeFlow\(^53\), a tool based on Lifelines and Lifelines2 is capable of visualizing care-related events, including patient transfers. The focus of LifeFlow is temporal clinical event visualization and implemented in Java and deployed as standalone software. Thus, integration with other healthcare delivery or operational data is a challenge, unlike EHDViz that offers various options for customized visualization and integrate with large library of predictive or statistical learning algorithms available as part of R language. CrowdED\(^54\) is another visualization aid that is specific to a clinical; the tool can be used for data visualization in the emergency departments but offers very
limited extensibility. An objective comparison of usability parameters and utilities by implementing various applications in same healthcare or clinical setting would provide a quantitative estimates of the preference, user interface preference and applications. However, several of the existing health care data or information visualization is designed to address a single task, yet lack extensibility. EHDViz address this important challenge by leveraging and demonstrating the use of widely used, scalable technologies to design and develop clinical dashboards for visualizations that would help the care-providers.

CONCLUSIONS

Due to the implementation of the Affordable Care Act (http://www.hhs.gov/healthcare/about-the-law/index.html) and emerging trend of hospitals to rebuild healthcare operations as affordable care organization (ACO), there is a growing need for health information technology solutions to be more practical and sustainable. This need calls for the appropriation of cutting edge health information technologies with minimal expense for the implementation and maintenance. Open source technologies offer an alternative option for healthcare providers to design, develop and deploy highly cost-effective clinical dashboards with no cost for the software license and re-use, thus may reduce the healthcare spending. We developed EHDViz as an open source clinical dashboard and visualization framework to integrate with existing commercial and open source EHRs for heterogeneous biomedical and healthcare data visualization. We envisage that design and development of real-time patient status tools coupled with risk estimation that could enhance the quality of health-care delivery and reduce adverse patient outcomes in clinical settings. EHDViz could also play an important role as a toolkit to emulate EHR environment for improving simulation-based learning. As hospitals and healthcare systems are emerging as learning health systems, in which data capture, smart clinical dashboards, and adaptive data visualization will play an integral role in managing the patient population.
**Contributorship statement:**
MAB, KS, BG, and MST contributed to the data integration, software package implementation, clinical dashboard and web server development. KS, MAL, PM, AK, DLR and JTD formulate visualization strategy and designed illustrative examples. JTD contributed to the overall planning of the project, the development of an extensible software package for clinical dashboard development and the manuscript. All authors have contributed to the writing and compilation of the final manuscript. All authors approve their contributions and the final draft of the manuscript.

**Competing interests:**
MAB has received consulting fees from MetaMed; KS: None declared; BG: None declared; JTD has received consulting fees or honoraria from Janssen Pharmaceuticals, GSK, AstraZeneca, and Hoffman-La Roche. JTD holds equity in NuMedii Inc, Ayasdi, Inc. and Ontomics, Inc. No writing assistance was utilized in the production of this manuscript.

**Funding:**
This work was supported by a grant from the National Institutes of Health, National Center for Advancing Translational Sciences (NCATS), Clinical and Translational Science Awards (UL1TR001433-01) to KS and JTD. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Data sharing statement:**
No additional data available.

**Acknowledgements:**
Authors would like to thank the following members of Mount Sinai Health System – Hospital Big Data initiative: Dr. Adel Bassily-Marcus, Dr. Bruce Darrow, Dr. Alan Moskowitz, Dr. Annette Gelijns, Dr. Patricia Kovatch and Kash Patel. KS and JTD
would like to thank Dr. Carol Horowitz and Janice Gabrilove for useful discussion about the applications of EHDViz in the areas of patient engagement and simulation-based learning. Authors would also like to thank Harris Center for Precision Wellness (http://precisionwellness.org/) and Icahn Institute for Genomics and Multiscale Biology (http://icahn.mssm.edu/research/genomics) for infrastructure support.
References:


Figure 1: Client-server architecture of EHDViz

151x93mm (300 x 300 DPI)
Figure 2: A quantified-self, healthcare data visualization dashboard developed using EHDViz

Different features of the dashboard are highlighted as 1: user management; 2: dynamic selection; 3: integration with data streams; 4: integration with manual data input.

215x166mm (300 x 300 DPI)
Figure 3: Different scenarios of implementing a visual aid for Modified Early Warning Score (MEWS) using EHDViz framework.

a) Visualization of a single patient; b) Visualization of a single patient layered on patient admission, discharge and transfer data; c) Visualizing trends of MEWS in different in-patient units; d) visualizing multiple patients in a same unit

131x62mm (300 x 300 DPI)
Figure 4: A customized, clinical evaluation dashboard developed using EHDViz that illustrates data in emergency department.

Features of this dashboard include selection of specific clinical units using a drop-down menu, controlling for the layout, and selecting patients that are tested for specific biomarkers. Different features of the dashboard are highlighted as 1: selection of individuals; 2: options to control visual layouts 3: integration with ICD-9 codes.

131x61mm (300 x 300 DPI)
Figure 5: A population health management visualization dashboard implemented using EHDViz

Different features of the dashboard are highlighted as 1: visualization of data from floor using admission-discharge-transfer data; 2: dynamic control of visualization; 3: real-time user interaction.
# EHDViz: Clinical Dashboard Development Using Open Source Technologies

<table>
<thead>
<tr>
<th>Journal:</th>
<th>BMJ Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>bmjopen-2015-010579.R2</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Research</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>29-Feb-2016</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Badgeley, Marcus; Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Shameer, Khader; Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Glicksberg, Benjamin; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Tomlinson, Max; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Levin, Matthew; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
McCormick, Patrick; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Kasarskis, Andrew; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Reich, David; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness
Dudley, Joel; Mount Sinai Health System, Icahn Institute for Genomics and Multiscale Biology, Translational Bioinformatics, Genetics, Genomics and Precision Wellness |

### Primary Subject Heading:
Health informatics

### Secondary Subject Heading:
Communication, Health services research, Medical management, Public health

### Keywords:
- Early warning systems, biomedical informatics, data visuzalization, clinical decision systems, clinical dashboard
EHDViz: Clinical Dashboard Development Using Open Source Technologies

Marcus A Badgeley¹,²,#, Khader Shameer¹,²,#, Benjamin S Glicksberg¹,², Max S Tomlinson¹,², Matthew A Levin³, Patrick J. McCormick³, Andrew Kasarskis², David L Reich³, Joel T Dudley¹,²,⁴*

¹ Harris Center for Precision Wellness ² Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology; ³ Department of Anesthesiology; ⁴ Department of Population Health Science and Policy; Icahn School of Medicine at Mount Sinai, Mount Sinai Health System, New York City, New York, USA.

*Corresponding author: joel.dudley@mssm.edu
#Denotes equal contribution

Keywords: clinical dashboard, health informatics, clinical decision support, real-time monitoring, big data, visualization
Abstract

Objective: To design, develop and prototype clinical dashboards to integrate high-frequency health and wellness data streams using interactive and real-time data visualization and analytics modalities.

Materials and Methods: We developed a clinical dashboard development framework called electronic healthcare data visualization toolkit (EHDViz) for generating web-based, real-time clinical dashboards for visualizing heterogeneous biomedical, healthcare and wellness data. The EHDViz is an extensible toolkit that uses R packages for data management, normalization and producing high-quality visualizations over the web using R/Shiny web server architecture. We have developed use cases to illustrate utility of EHDViz in different scenarios of clinical and wellness setting as a visualization aid for improving healthcare delivery.

Results: Using EHDViz, we prototyped clinical dashboards to demonstrate the contextual versatility of EHDViz toolkit. An outpatient cohort was used to visualize population health management tasks (n=14,221), and an inpatient cohort was used to visualize real-time acuity risk in a clinical unit (n=445), and a quantified-self example using wellness data from a fitness activity monitor worn by a single individual was also discussed (n-of-1). The backend system retrieves relevant data from data source, populate the main panel of the application and integrate user-defined data features in real-time and render output using modern web browsers. The visualization elements can be customized using health features, disease names, procedure names or medical codes to populate the visualizations. The source code of EHDViz and various prototypes developed using EHDViz are available in the public domain at http://ehdviz.dudleylab.org.

Conclusion: Collaborative data visualizations, wellness trend predictors, risk estimation algorithms, proactive acuity status monitoring in a clinical setting, and complex disease indicators are essential components of implementing data-driven precision medicine. As an open-source visualization framework capable of
EHDViz aims to be a valuable toolkit for rapid design, development, and implementation of scalable clinical data visualization dashboards.
STRENGTHS AND LIMITATIONS

- Developing scalable and sustainable healthcare information technology (HealthIT) solutions for data management, visual analytics and scalable visualization is critical to improving the quality of affordable healthcare delivery.

- We developed EHDViz as a cost-effective, open-source, extensible toolkit for rapid design, development, and implementation of clinical dashboards to address the need to improve data visualization in different aspects of healthcare delivery including population health management, patient engagement, and simulation-based learning.

- A significant limitation of current version of EHDViz is that developers need skills in R and web-development; also extensive data cleaning and quality control steps are a priority before importing large quantity data for visualization.

- While EHDViz is designed a vendor agnostic framework, importing data from external systems such as sensor devices and fitness monitors need authorizations for access to the data and technical support from the device manufacturers or data integration services.

- To understand the benefits and limitations of the user experience, EHDViz, and other open source or commercial solutions with similar capabilities must be compared.
INTRODUCTION

The subclinical features and symptoms vary for individual patients as diseases progress and can be affected by lifestyle and medical interventions. These variations deviate greatly even among people with similar demographics, clinical profiles, family history and disease burdens. The patient-specific intervention a physician incorporates into a treatment plan relies heavily on the course of the illness.

Electronic Health Record (EHR) software is widely used to capture longitudinal data and record vital signs, medications, laboratory values, diagnostic reports, fluid inputs/outputs, mental states, patient transfers, and other health status parameters. However, EHR software often presents data with tabular views or static text formats, which does not reveal the underlying trends in a patient’s disease progression nor the similarities among patient trends within a given department. EHRs have limited capabilities to integrate biomedical, clinical, and patient-generated data integrated and physicians often have to use multiple tools to gather patient status from heterogeneous databases to get a complete health assessment.

Decades of research have shown that graphical summaries of patient information provide faster and more accurate medical diagnoses, thus improving the healthcare quality. Reproducible studies demonstrate that practicing clinicians are unlikely to adopt any information retrieval task that takes longer than 30 seconds. As patients are becoming more empowered through an increase in patient-generated data, physicians are now being challenged to comprehensively visualize increasingly complex patient histories and associated data streams in a short span of time in the clinical setting. There is an unmet need in the continuum of healthcare delivery to develop better ways of visualizing and interpreting EHR data, upon which physicians can base critical treatment decisions. Electronic Health Data (EHD) includes the acquisition of physiological values, diagnostic reports (radiology reports), laboratory values, pathology reports (biopsy report), physician consults, and clinically actionable genetic information. In recent years, this is further supplemented by information that patients provide directly, such as blood pressure and food logs, and by continuous physiologic data from wearable devices from
patient portals\textsuperscript{10} or mobile phones\textsuperscript{3}. Some institutions have implemented remote monitoring of patients using implanted devices, like the implantable cardioverter-defibrillator, as well as augmented clinical management using data streams from health monitoring devices\textsuperscript{11}, leading to improved outcomes\textsuperscript{12, 13}, cost savings\textsuperscript{12, 14}, and earlier identification of device malfunction\textsuperscript{15}. A subset of modern clinical trials are also incorporating remote monitoring devices, including ones capable of collecting physiological data and cloud computing electrophysiological data\textsuperscript{16}. As an example, several ongoing clinical trials make use of Apple ResearchKit for evaluating patients with asthma\textsuperscript{17}, cardiovascular disease, diabetes, and Parkinson’s disease (See: http://researchkit.org/), and some efforts implement mHealth based solutions to engage patients and visualize data (AppCore: https://github.com/researchkit/AppCore). This trend is growing and generating a n influx of data that patients or physicians typically not handle in the clinical setting. Better tools are now required to integrate such data streams and provide detailed summary of a patients for improving patient engagement. See Khader and Marcus \textit{et.al.} for an extensive discussion on real-time data streams in healthcare.

The expanding number of data streams that can integrate with EHR contributes to the increase in the volume of big data in health care, which lays the foundation for paradigm shifts in modern medicine. Furthermore, acceleration in massive data influx is expected with the wide-adoption and maturity of internet of healthcare things (IoHT; See https://en.wikipedia.org/wiki/Internet_of_Things), where health sensors, fitness monitors, and implantables will be able to upload directly physiological data to patient authorized and secure systems, which can provide a data-rich portrait of a patient at the point of care. Recent reviews on the application of big data to healthcare identify actionability and decentralized data\textsuperscript{18} as key challenges to fulfill the potential of big data in healthcare.\textsuperscript{19, 20} The EHD of a patient can be aggregated and normalized using data from personal logs, health monitoring or fitness devices, and medical devices (e.g., continuous positive airway pressure pumps for sleep apnea). To get a comprehensive picture of wellness or illness state of patients, the data from hospital administration and operations data
can also be aggregated. Information retrieval systems that integrate multiple data streams have been developed or suggested for particular applications including antibiotic clinical decision systems\(^3\), recording events during surgical operations\(^{21}\), diabetic patient data collection programs\(^{22}\), text-mining, concept extraction from clinical documentation using natural language processing\(^{19}\), and wearable devices using software application programming interfacing (API) services (e.g. Human API; See: https://www.humanapi.co/) that allow communication between health monitoring devices and provider databases using a secure, programmatic data access protocol. As of yet, there is no system that permits flexible retrieval and interactive visualization of EHR, medical, and patient-generated data streams and provide tools for real-time visualization.

**Clinical data visualization:**
Visual descriptions of the health status of patients in clinical settings have been a challenging problem since the introduction of computer programs for care management. At that time, the principal limitations were resource availability including appropriate graphical engines for rendering and specialized hardware to visualize patient status using computer programs.\(^2\) Decades later, visualization of clinical information and communication of wellness trajectories or disease risk trajectories of a patient using visual cues remains as an emerging challenge in the current era of data-driven medicine.\(^{23,24}\) Visualization tools can track the biochemical variation and physiological status of patients, as well as quantify biomarkers. Such visualization aids were originally part of particular medical devices designed to monitor one or more specific physiological variables, such as heart rate and pulse rate, both within and external to EHRs. Efficient tools, algorithms, and risk prediction models are now required to for visual communication of clinical information to manage the high volume of biomedical and healthcare data in the hospital setting. Integrating such visualization tools with predictive models and risk estimation tools could support accelerated patient stratification for improved care.
Visualizing healthcare data using clinical dashboards:

Clinical dashboards are tools that can visually capture the cross-sectional view of a variety of quality metrics, including patient statuses, progress in cohort aggregations, patient safety and healthcare delivery measures, performance improvement for care providers and aid in understanding the key features of the overall patient satisfaction and improved outcomes. Clinical dashboards are often developed using commercial or custom-built tools internally developed by hospitals or health systems, thus little to no interoperability with tools are available for statistical analysis, machine learning, or integration with predictive modeling that can aid in tasks including acuity prediction and readmission evaluation. Designing visual tools to graphically explain risk scores and predictive models would help to accelerate patient risk stratification for improved care. While there are a variety of technical challenges for integrated visualization of multiple clinical visualization tools, the interoperability of EHR applications and data feeds from medical devices remains a significant challenge. Vendor standards also hinder the integration of diverse data elements into a common platform. Data feeds often need extensive quality control, normalization, or other preprocessing procedures before the utilization in risk scoring engines. Visualization also plays a crucial role in shared-decision making (SDM), where a care provider and patient participate in a discussion regarding therapeutic strategies or clinical care delivery pathways that are supported by a variety of tools including those with visualization. Shared decision making is currently used in the treatment and management of cardiovascular diseases, diabetes, and osteoporosis. Irrespective of the medical specialty, visualization improves the ability to understand trends in a patient’s health and the effects of interventions over time.

Visualization and forecast of risk estimations in clinical setting:

Adverse events during hospitalization such as hospital acquired infections (HAI) or hospital acquired conditions (HAC) including falls, cardiac arrests and unanticipated intensive care unit (ICU) transfers, and death are frequently preceded by several useful and predictive features that can be used for accelerate triaging to improve the...
care delivery. For example, slow and progressive physiological decompensation was identified in cardiac arrests (79%), unexpected ICU transfers (55%), and death during hospitalization (54%) in a retrospective study that compared cohorts from different countries. Failure to recognize and respond to signs of deterioration includes infrequent or incomplete vital sign assessments, poor design of vital sign charts, and reduced accuracy of ‘track-and-trigger’ systems. Several single parameter and multi-parameter risk scoring methods have been proposed to implement a “track-and-trigger” method of alerting for patients in clinical wards within 24 hours of an adverse event for accelerated clinical intervention aids. The most established methods are based on vital signs and neurologic status, including Modified Early Warning System (MEWS), Standardized Early Warning Systems (SEWS), and National Early Warning Score (NEWS) that differ on the inclusion of oxygen saturation and supplemental oxygen and the weight of different features. When assessed retrospectively, these vital based systems have an area under receiver operator curves (AUROCs) of 0.76-0.83 for cardiac arrests, 0.73-0.77 for ICU transfers, and 0.87-0.88 for mortality, effective for triggering follow-up evaluation. Implementations of these warning systems have required that staff perform rounds and fill out paper sheets or electronically enter the vital signs. Reports of real-time EHR information retrieval based implementations of early warning systems have had some success in reducing adverse events in randomized control trials and cross-over trials, although the risk models have restrictions due to the limited physiological feature space. Algorithms developed using the entire set of discrete health characteristics in the clinical data warehouse have incorporated significantly predictive laboratory values, physician orders, and medications. When assessed retrospectively these features consistently outperform the vital-constrained approaches.

In this paper, we propose the design, development, and implementation of an extensible clinical dashboard development framework by leveraging open source technologies. The Electronic Health Data Visualization framework (EHDViz) is an interactive and extensible framework implemented using a modern statistical computing language. Biomedical informatics scientists, solution architects can use
EHDViz to develop clinical decision aids to empower patients. EHDViz provides an EHR-agnostic visualization framework that can be implemented in real-time to assist healthcare providers in identifying patients with decompensating physiology via a visual aid. Thus, healthcare delivery management teams, healthcare executives, and medical professionals can use the dashboards developed using EHDViz to retrieve, integrate, and explore diverse healthcare data streams to assess patient health trends in a clinical unit, hospital or health system.

METHODS

Description of EHDViz Framework:

EHDViz is a software framework designed to interactively generate web-based healthcare data visualization using various R packages (R language; R version 3.0.2; 2013-09-25). We provide an infographic of the client-server architecture of EHDViz in Figure 1. We compiled various packages to organize a unified software framework for data input/output operations, data management of healthcare data, data cleaning and normalization from diverse sources, generation of plots, and statistical analysis. Data cleaning and quality control steps including the removal of outliers were performed using reshape2 package (https://cran.r-project.org/web/packages/reshape2/index.html). EHDViz uses the packages ggplot2 (http://ggplot2.org) and gridExtra (https://cran.r-project.org/web/packages/gridExtra/index.html) for developing plots. Native R plots can be generated and visualized using PDF viewers, generic image viewers, and web-browsers, but base R offers limited options for visualizing real-time data streams. We developed a custom algorithm to combine individual R plots and visualize as a continuous, real-time data stream. We implemented the web server implementation using R/Shiny to deploy the plots created as part of EHDViz framework. We used the Shiny server architecture (https://github.com/rstudio/shiny-server) because it can be implemented over multiple desktop and server environments and can be distributed as suitable software modules. Data from wearable devices are compiled using the device
specific API for Fitbit. Wearable-specific APIs offer a secure way to collect and aggregate data generated by personal fitness monitoring devices. The package fitbitScraper (https://cran.r-project.org/web/packages/fitbitScraper/index.html) was used to extract the data from the wearable device.

Data handling in EHDViz:
Various biomedical and healthcare data types, including disease and procedure indexes, clinical dictionaries and ontologies, namely the International Statistical Classification of Diseases and Related Health Problems (ICD-9: http://www.who.int/classifications/icd/en/) codes are indexed in the current implementation to define specific disease terms pertaining to patients as part of diagnoses. Patients undergoing specific clinical procedures can also be retrieved and aggregated using Current Procedural Terminology codes (CPT: http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/about-cpt.page) or SNOMED-CT (Systematized Nomenclature of Medicine -- Clinical Terms) codes. EHDViz can also parse and normalize medication data using National Drug Codes (NDC) and RxNorm and use medication data as part of the data aggregation methods in EHDViz. (NDC: http://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm; RxNorm https://www.nlm.nih.gov/research/umls/rxnorm/). EHDViz can also handle data from operational and administrative datasets generated as part of healthcare delivery, including patient-transfer data (i.e. from the emergency department to surgery to ward and discharge) to query or aggregate patient cohorts in an adaptive fashion and to precisely visualize their health trends.

Input and output specifications of EHDViz:
EHDViz can handle data in tab-delimited file format (.tsv) or comma-delimited file format (.csv). Data can also be extracted from various other formats and database using native R packages. For example, EHDViz can extract data from Excel files (.xlsx: https://cran.r-project.org/web/packages/xlsx/index.html) or relational database systems that conforms to Open Database Connectivity (ODBC) standards (RODBC:...
Java Database Connectivity (JDBC) (RDBC: https://cran.r-project.org/web/packages/RJDBC/index.html), MySQL (RMySQL: https://cran.r-project.org/web/packages/RMySQL/index.html) or modern, NoSQL database systems such as MongoDB (rmongodb https://cran.r-project.org/web/packages/rmongodb/index.html). Various examples of input formats and sample files are provided at the URL: http://ehdviz.dudleylab.org/help.html#introduction. The data gathered from sources including EHRs, flat files, data warehouse, or database connections will use as the input for EHDViz dashboards and output customized visualization. The diverse set of data from various sources after parsing, quality control, and normalization can be integrated into the visualization templates of an individual project. The output of EHDViz is in the format of customized visualization dashboards rendered using a standard, modern web browser that supports HTML5 and responsive web development standards. Current version of EHDViz dashboards were successfully Chrome Browser, Safari.

**Clinical dashboards developed using EHDViz:**

To evaluate the technical challenges in developing and deploying a real-time biomedical, clinical, and patient-generated data visualization dashboards, we created multiple prototype web applications using R language in the backend and the R/Shiny web server architecture as the front-end as outlined above. Prototype dashboards are developed using three different data sets: 1) data from a single patient (n-of-1) with data streams not captured in a clinical setting demonstrate quantified-self visualization; 2) simulated cohort of inpatients (n=445) and 3) simulated cohorts of outpatients (n=14,221). The data simulation was performed using a deidentified EHR compiled at Icahn School of Medicine at Mount Sinai (ISMMS), a hospital of the Mount Sinai Health System in New York City. Data from fitness monitoring devices was aggregated using an API capable of secure retrieval of data from the fitness monitoring device of a user, and a custom web service function was designed to pull and integrate user defined data features in real-time.
Dashboards discussed in this manuscript are implemented on a web server with nginx (http://nginx.org/) on a secure, cloud-based virtual private server running on Ubuntu. The web interface is implemented using HTML, CSS, and JavaScript, and visualization dashboards are rendered using R/Shiny architecture.

RESULTS

Availability: The source code of EHDViz and various clinical dashboard implementations are available from the URL: http://ehdviz.dudleylab.org/.

Clinical dashboards developed using EHDViz:

Collaborative data visualizations, wellness trend predictors, risk estimation algorithms, proactive acuity status monitoring in a clinical setting, and complex disease indicators are essential components of implementing data-driven precision medicine. In following section, we discuss various dashboards developed using EHDViz. Briefly, we parsed the source data and removed the outliers as part of the data cleaning step. A custom web service function was designed to pull and integrate user-defined data features in real-time from simulations of the clinical cohorts and fitness monitors using normalized data. The final dashboards were designed to show specific visualizations.

Dashboard 1: Visualizing time series health data (quantified self):

The quantified-self movement involves an increasing interest in individuals and patient communities in tracking many types of biometric data in order to gain insight into their health. Increasingly, patients are able to access and control their clinically collected health data. Our first demonstration addresses the challenge in quantified-self area of integrating and visualizing time series health data from multiple data sources. The example in Figure 2 demonstrates the integration of an individual patient’s EHD sources. For this example, the patient has 3 primary sources of health data: 1) clinical data from outpatient visits, 2) continuous activity data from a wearable device (Fitbit; San Francisco, CA) and 3) a self-recorded blood
pressure log. The clinical data from ambulatory visits were simulated by randomly sampling aggregated physiologic and lab values from 14,221 patients in an ISMMS outpatient cohort. The continuous activity data was scraped from one of the author’s (MAB) wearable device using the API at an interval of 15 minutes. The blood pressure log is simulated as weekly measurements from normal distributions $N(130,15)$ and $N(85,10)$. The user interface features a main panel with ‘sparklines’ for each health feature and a sidebar with widgets for the user to select the health features of interest. In this example, a checkbox is provided to group patients for each data source: (1) EHR, (2) data from fitness monitoring device, and (3) personal log. The user can select any combination of health features to be displayed. The main panel displays a stack of sparklines with selected health features sorted according to values selected in the sidebar. Minimums and maximums are highlighted with red and blue dots, respectively. In this application, the data source that updates most frequently was from the wearable device collected in 15-minute intervals; the application was programmed to auto refresh every 15 minutes to retrieve new data.

**Dashboard 2: Visual analytics of data streams in clinical setting:**

Next we demonstrate the retrieval of continuous data contained in a collection of patient’s EHRs during an inpatient stay, where data will be much more dynamic than in the previous outpatient example. This implementation was tested with both a simulated cohort of 445 inpatients with clinical labs recorded throughout their encounter as well as with simulated data (Figure 3a-d and 4).

User of this particular dashboard user can use the sidebar to select a patient and the date range of interest. The relevant information is then retrieved from the EHR or data warehouse throughout the encounter (Figure 3a-d). Within a single hospital visit, a patient could be go through different hospital units including the emergency department, ICUs, inpatient units, surgical suite or ambulatory wards depending on the clinical status of the patient. In this example, patient transfers including admission, transfers and discharge were color-coded by location to intuitively show the dynamic trends in the health status (Figure 3b). For the
simulated data, we randomly retrieved data from the EHR for a age and gender matched cohort with 14,221 patients to populate each of the 375 discrete continuous health features contained in the EHR.

For each of the 7,000 unique diagnoses, we pooled corresponding patient data and found the most frequently measured health features for each ICD-9 class. The simulated patient dashboard (Figure 4) allows the user to select a patient and an ICD9 class from the drop-down menus in the side panel, which then populates the main panel with the most common health features measured for that ICD9 class.

The list of health features corresponding to the selected ICD9 class is additionally displayed as a checkbox group in the side panel so the user can further refine the displayed feature set. This enables the user to rapidly retrieve and assess trends in the most relevant biomarkers. We also provide a demonstration at http://ehdviz.dudleylab.org/providers/full that allows a key word based search and multi-selection of all 375 health features to make customized dashboards. Real-time displays were also designed from the simulated data, demonstrated at http://ehdviz.dudleylab.org/providers/real-time.

**Dashboard 3: High-velocity patient acuity status monitoring and data visualization in the clinical setting:**

The examples in Figure 3c, 3d, and 5 demonstrate the use of EHDViz for developing visual aids for patient safety and cohort analysis. These dashboards provide risk estimation visualization for users to track all patients simultaneously in a unit, which facilitates the identification of atypical and destabilizing features to trigger interventions. Patient vital signs were retrieved from the EHR warehouse from 445 inpatients and processed to calculate the MEWS risk score. Figure 3c and 3d show the dashboard for monitoring these patients MEWS and shows the clinical stability trends. The user can select the clinical unit of interest with the drop-down menu, and sparklines with MEWS scores are displayed for each patient in the unit with alert triggering thresholds displayed for reference. When there are multiple patients in the unit, MEWS scores are colored by patient (Figure 3d). Data for online demonstrations were simulated as discussed in scenario 2 and the “location” and
"patient" covariates were switched from a data coloring covariate to a user filtering covariate and vice versa for use in a cohort application. As shown in Figure 5, a user can select the clinical unit of interest and text search different clinical parameters, and the main panel will display the values of these features for all the patients in the selected unit, colored by patient. This design allows rapid evaluation of various clinical features or predictors. Multiple values relevant to clinical manifestations of patient population can be compiled and new scores (e.g. MEWS) can be computed for a population of patients. Demonstrations of ICD9 class based feature selection at http://ehdviz.dudleylab.org/visualizations/Population_Management_ICD9/ and a real time monitoring dashboard implemented using EHDViz is provided at the URL: http://ehdviz.dudleylab.org/visualizations/Population_Management_RealTime/

Discussion:
The treatment pathway for a patient depends on a number of factors that can be collected from different sources including patient generated data, medications, vital signs, diagnoses, and responses to therapies or other interventions. Physicians can collect data from the EHRs, patient health records, patient portals electronic patient diaries, fitness trackers, and the patient’s recollections of medical history. In most presentations, however, this data overwhelms physicians instead of guiding them to informed decision making. Real-time clinical monitoring and automated alerting provide better tools to improve patient safety, clinical outcomes, and quality of healthcare delivery. Tools are currently available to monitor patient acuity, infectious diseases, and adverse events. Specifically, there are customized tools that target specific needs of the clinical unit including operating rooms or ICUs. Developing a unified visualization tool that can provide an overview of a patient by integrating different healthcare, biomedical, or clinical data streams remains an open challenge. EHDViz, an open source data visualization framework capable of real-time data visualization, can be used to address many of these issues. EHDViz aims to unify heterogeneous biomedical and healthcare data integration through R language, a popular and preferred programming language for scientific computing,
predictive analytics and machine learning. R language is typically used for desktop or client-cluster based visualization models. Here, we have improvised an R visualization package designed to generate static plots and rendered it as a real-time data visualization engine. Real-time displays can also be implemented and deployed over the web-browsers using other programming languages including Python and JavaScript, and future releases of EHDViz could extend to these languages. Close integration with R also enables visual analytics and predictive modeling using the large library of R packages to run seamlessly within EHDViz. Users can customize the different levels of implementation of EHDViz dashboards for disease-, division-, or institutional-specific applications. EHDViz offers features to integrate risk prediction algorithms for patient stratification with data mining algorithms to utilize underlying data repositories to refine the user experience and automatically retrieve the most relevant data for a selected context. Integrating various risk assessment algorithms with the traditional clinical dashboard style interface offers a powerful toolkit for clinicians. EHDViz could aid in designing dashboard development projects that combine visualization, analytics, and predictive modeling in healthcare and wellcare.

Application of EHDViz in simulation-based medical education:
Simulation-based learning is at the core of the pedagogical principles of modern medicine. Medical students, residents, and physicians extensively use EHR at the bedside during care delivery. EHDViz is an EHR and vendor-agnostic dashboard development toolkit that users can leverage as a teaching aid capable of generating custom EHR instances and visualizations. Simulated EHR systems can be designed based on single-use cases to evaluate an individual patient or number of patients that a resident is managing on a floor or unit.

Comparison with related healthcare data visualization applications:
Multiple visualization tools are currently available for effective integration of actionable information in the workflow of clinical care pathways. A systematic review of data visualization tools assessed multiple clinical data visualization tools:
tools like EventFlow⁴⁸, LifeLines⁴⁹, LifeLines2⁵⁰, Visualization of Time-Oriented Records (VISITORS),⁵¹ and Dynamics Icon (DICON)⁵² were listed as tools capable of clinical data visualization and dashboard development. Deng Y et al used a tag-cloud from radiology reports, pathology reports, and surgical reports to summarize unstructured patient data.⁵³ Data visualization tools, such as HARVEST,²⁴ offer web-based infrastructure for integrating, discovering, and reporting data, but are restricted to the data captured in a data warehouse. The design philosophy of EHDViz is to provide a tool that can integrate and visualize data from different sources in addition to data warehouses. Lifelines and Lifelines2 offer options to align, rank, and summarize temporal visualizations. LifeFlow⁵⁴, a tool based on Lifelines and Lifelines2, is capable of visualizing care-related events, including patient transfers. The focus of LifeFlow is temporal clinical event visualization and implemented in Java and is deployed as standalone software. Thus, integration of different healthcare delivery or operational data is a challenge for LifeFlow. EHDViz, on the other hand, offers various options for customized visualization and integration with a large library of predictive or statistical learning algorithms available as part of R language. CrowdED⁵⁵ is another visualization aid that is specific to the specific clinical locations; the tool can be used for data visualization in the emergency departments but offers very limited extensibility. An objective comparison of user experiences, usability parameters and utilities by implementing various applications in same healthcare or clinical setting would provide quantitative estimates of the preference of data sources and user interface preferences. Several of the existing health care data visualization tools, however, are designed to address a single task and lack extensibility. EHDViz addresses this important challenge by leveraging widely used, scalable technologies to create clinical data visualization dashboards to aid care-providers.

**CONCLUSIONS**

Due to the implementation of the Affordable Care Act (http://www.hhs.gov/healthcare/about-the-law/index.html) and the emerging trend of hospitals to rebuilding healthcare operations as affordable care
organization (ACO), there is a growing need for health information technology solutions to be more agile and sustainable across different levels of hospitals and health systems. The need for delivering high-quality care by leveraging biomedical and healthcare data calls for the appropriation of health information technologies capable of handling and managing healthcare big data. Open source technologies offer a complementary option for health information technology (HealthIT) developers to design, develop, and deploy cost-effective clinical dashboards with no cost for the software license and re-use. Adoption of these technologies HealthIT may thus reduce overall healthcare spending. We developed EHDViz to integrate data from diverse sources including for biomedical and healthcare data visualization for integrated health assessment. Further, EHDViz could also play a significant role as a toolkit to emulate EHR environment to improve simulation-based learning. Hospitals and healthcare systems are emerging as learning health systems, and as such, data capture, smart clinical dashboards, and adaptive visual analytics could play an integral role in managing the patient population. We envisage that design and development of real-time patient status assessment tools coupled with risk estimation using heterogeneous data could enhance the quality of health-care delivery and reduce adverse patient outcomes.

**Contributorship statement:**

MAB, KS, BG, and MST contributed to the data integration, software package implementation, clinical dashboards, and web server development. KS, MAL, PM, AK, DLR, and JTD formulated visualization strategy and designed illustrative examples. DLR and JTD contributed to the overall planning of the project, the development of an extensible software package for clinical dashboard development, and the manuscript. All authors have contributed to the writing and compilation of the final manuscript. All authors approve of their contributions and the final draft of the manuscript.

**Competing interests:**
KS: None declared; MAB: None declared; BG: None declared; MST: None declared; MAL: None declared; PM: None declared; AK: has received consulting fees from Mars, Inc.; DLR: None declared; JTD: has received consulting fees or honoraria from Janssen Pharmaceuticals, GSK, AstraZeneca, and Hoffman-La Roche, and holds equity in NuMedii Inc, Ayasdi, Inc., and Ontomics, Inc. No writing assistance was utilized in the production of this manuscript.

Funding:
This work was supported by a grant from the National Institutes of Health, National Center for Advancing Translational Sciences (NCATS), Clinical and Translational Science Awards (UL1TR001433-01) to KS and JTD. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data sharing statement:
No additional data available.

Acknowledgements:
The authors would like to thank the following members of Mount Sinai Health System – Hospital Big Data initiative: Dr. Adel Bassily-Marcus, Dr. Bruce Darrow, Dr. Alan Moskowitz, Dr. Annetine Gelijns, Dr. Patricia Kovatch and Kash Patel. KS and JTD would like to thank Drs. Carol Horowitz and Janice Gabrilove for useful discussion about the applications of EHDViz in the areas of patient engagement and simulation-based learning. The authors would also like to thank Harris Center for Precision Wellness (http://precisionwellness.org/) and Icahn Institute for Genomics and Multiscale Biology (http://icahn.mssm.edu/research/genomics) for infrastructure support.

References:


Figure 1: Client-server architecture of EHDViz
151x93mm (300 x 300 DPI)
Figure 2: A quantified-self, healthcare data visualization dashboard developed using EHDViz

Different features of the dashboard are highlighted as 1: user management; 2: dynamic selection; 3: integration with data streams; 4: integration with manual data input.

215x166mm (300 x 300 DPI)
Figure 3: Different scenarios of implementing a visual aid for Modified Early Warning Score (MEWS) using EHDViz framework.

a) Visualization of a single patient; b) Visualization of a single patient layered on patient admission, discharge and transfer data; c) Visualizing trends of MEWS in different in-patient units; d) visualizing multiple patients in a same unit

131x62mm (300 x 300 DPI)
Figure 4: A customized, clinical evaluation dashboard developed using EHDViz that illustrates data in emergency department.

Features of this dashboard include selection of specific clinical units using a drop-down menu, controlling for the layout, and selecting patients that are tested for specific biomarkers. Different features of the dashboard are highlighted as 1: selection of individuals; 2: options to control visual layouts 3: integration with ICD-9 codes.

131x61mm (300 x 300 DPI)
Figure 5: A population health management visualization dashboard implemented using EHDViz

Different features of the dashboard are highlighted as 1: visualization of data from floor using admission-discharge-transfer data; 2: dynamic control of visualization; 3: real-time user interaction