

Infectious Diseases Data Analysis Program: A Proof of Concept for User-Friendly Patient Data Analysis

A Capstone Report
presented to the faculty of the
School of Engineering and Applied Science
University of Virginia

by

Jessica E. Mahoney

with

Michael Hughes

Elnaz Ghajar-Rahimi

Laura Dunphy

Dr. Jason Papin

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signature _____ Date _____
Jessica Mahoney

Signature _____ Date _____
Dr. Jason Papin

Infectious Diseases Data Analysis Program: A Proof of Concept for User-Friendly Patient Data Analysis

Michael M. Hughes^a, Elnaz Ghajar-Rahimi^a, Jessica E. Mahoney^a, Laura J. Dunphy¹, Jason A. Papin²

^a Fourth Year Biomedical Engineering Undergraduate at the University of Virginia

¹ Correspondence: ljd6ab@virginia.edu, 415 Lane Rd. Charlottesville, VA. 22908, (802)338-5769

² Correspondence: jap8r@virginia.edu, Box 800759, Health System University of Virginia Charlottesville, VA. 22908, (434)924-8195

^{1,2}Correspondence Affiliation: Department of Biomedical Engineering, University of Virginia

Abstract

Storing patient data within Electronic Health Record (EHR) systems has become a standard practice in contemporary medicine. Although 86% of all office-based physicians across the United States incorporate EHRs into their practices, clinicians have cited even the most popular EHR system in the country, EPIC, as being overly complex and difficult to use. Interviews revealed that systems such as Epic have yet to successfully develop data filtering and data visualization packages that are easy to use. Clinicians also expressed the desire for a program dedicated to aiding large patient data analysis processes. In order to provide an appropriate scope, this project focuses on data from the Infectious Diseases Department at the University of Virginia. This project aims to design a program that easily sorts and presents patient data according to user commands by focusing on time taken to complete data analysis and program ease of use. Here, we present the Infectious Disease Data Analysis Program (IDDAP), that serves as a proof-of-concept for third-party, user-friendly data analysis software designed specifically for use by clinicians. Through user performance assessment and surveys, we illustrate that IDDAP is more efficient than Excel for completing data visualization tasks, and that IDDAP's layout is intuitive and organized. The development of a standalone data analysis program that easily sorts and displays clinicians with patient data can improve data analysis for medical research in all fields and thus improve outcomes for patients in the future. IDDAP can be deployed via html for Windows and Mac. IDDAP's source code is publicly available through the web-hosted app (<https://iddap.shinyapps.io/iddap/>) or on Github (<https://github.com/jessmahoney12/IDDAP>).

Keywords: patient data analysis, patient data sorting, Electronic Health Record systems, Epic, IDDAP, Excel, infectious disease

Introduction

Due to the advances in technology in the healthcare sector, each patient now comes into the doctor's office not only with an illness or injury, but also a host of data. Healthcare organizations have seen an explosion of data in recent years, climbing to 8.41 petabytes in 2018.¹ To put that in perspective, a 64 gigabyte smartphone is 0.0000064 petabytes. Data, carried everywhere, are growing rapidly and few areas are growing more rapidly than healthcare.¹ This growth is due in part to the passing of the Health Information Technology for Economic and Clinical Health (HITECH) Act in February 2009 which mandated that all

healthcare institutions across the United States use an Electronic Medical Record (EMR) system in a "meaningful way".² The "meaningful use" of EMR (also referred to as Electronic Health Records (EHR) systems) was not clearly defined by the Obama administration when the HITECH act was passed; however, has since included things from storing or consolidating patients' protected health information (PHI) to electronically prescribing medication.^{3,4} Due to the passing of this legislation, nearly every hospital in the US now uses an EHR system. As of 2017, approximately 96% of U.S. Non-federal Acute Care hospitals, and 86% of all office-based physicians have adopted a certified EHR

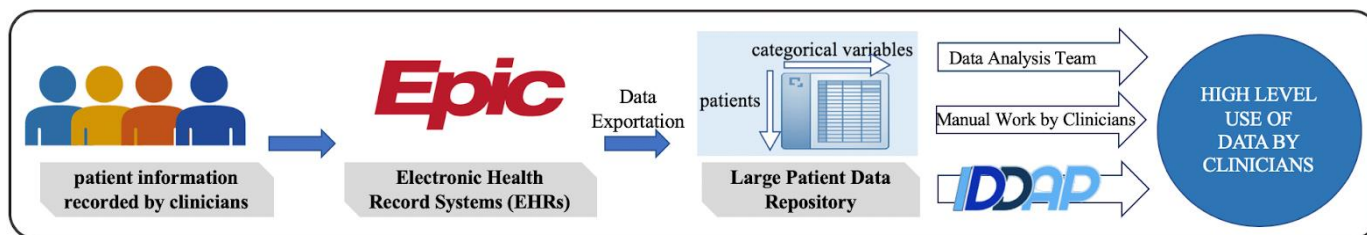


Fig. 1. IDDAP in the Medical Field. *IDDAP was made to fill the gap between patient data in Epic and the medical research clinicians wish to perform. Patient information recorded by clinicians is uploaded into EPIC, where it stays until grouped together and extracted to Large Patient Data Repositories such as Excel sheets. Once there, clinicians can either use IDDAP alone, or in combination with manual methods and help from data analysts, to aid high level patient data analysis. Once there, clinicians have the ability to perform analysis that is often slow and tedious.*

system^{5,6}. The purpose of such legislation was to improve the manner in which healthcare is delivered and patients are served, and EHR systems have certainly done that to an extent.^{2,7}

One of the most popular EHR systems in the United States is a system called Epic. Epic claims a majority of the market share with 58% of large hospitals, 28% of all acute care hospitals, and over 250 million patient records.^{8,9} The research done in this paper focuses on the professionals using Epic at the University of Virginia's (UVa) Medical center. Despite its widespread use, many clinicians have reported difficulties using Epic. Previous studies conducted at UVa's Health System as well as the data collected for the purpose of this research have shown that the system can be overly complex and result in data inconsistencies.³ While shadowing medical students in the UVa hospital, it was apparent that clinicians conducting patient research manually filtered thousands of patient data entries in Microsoft Excel rather than using Epic. This was due in part to Epic's complexity, inconsistency, and mostly to its lack of ability to isolate large amounts of information for analysis. Doing so allows clinicians to discover the role different variables have in treatment outcomes and other measurable factors such as admission rates. Currently, clinicians have this ability in a limited scope. Clinicians can either go through the aforementioned process in excel or reach out to the data analytics team.¹⁰ The flow of patient data from patient to medical research is shown in Figure 1. Though it is typically a lengthy process, data analysts can extract the data the clinicians require from Epic, perform the requested analysis and provide the clinician with a dashboard most often formed in a software called Tableau.¹¹ This process takes place between Epic and Large Patient Data Repositories. (Figure 1) These dashboards are specific to the initial question asked by the clinician and have little to know user modification abilities, so if the clinician

wishes to learn of a new insight he or she must request a new analysis dashboard.¹⁰

Through continued one-on-one interviews with clinicians and analysts in the Infectious Diseases department it became apparent that a gap exists between the tools available in the Epic EHR system and the goals that clinicians wish to accomplish. Our program aims to bridge this gap in order to improve the efficiency of data analysis in the Infectious Diseases department at UVa. (Figure 1) A program that quickly and easily generates visualizations of patient data would help clinicians perform the previously mentioned analysis, while also helping them gain insight into ways in which their work as caretakers affects patients' outcomes.

Results

Identification of Program Requirements

To identify the data analysis tasks that medical researchers typically perform, clinicians in the Department of Infectious Disease at the UVa hospital were interviewed in person and surveyed. Survey respondents consisted of researchers across several disciplines of infectious diseases: drug resistance, disease epidemiology, pathogenesis, biotechnology and treatment delivery, immunology, and alternative pharmacology. Clinicians indicated a need for filtering large amounts of patient data specifically. Although 67% of clinicians noted that they analyze data on a weekly basis, a majority of them had either not heard of, or had heard of but not used the pre-existing Slicer Dicer tool in Epic that is intended for aiding data filtering processes. Furthermore, of the clinician responses, 83% recorded that they would find a data filtering application for patient data to be helpful. Survey responses are summarized in Figure 2. The full survey is included in the Appendix.

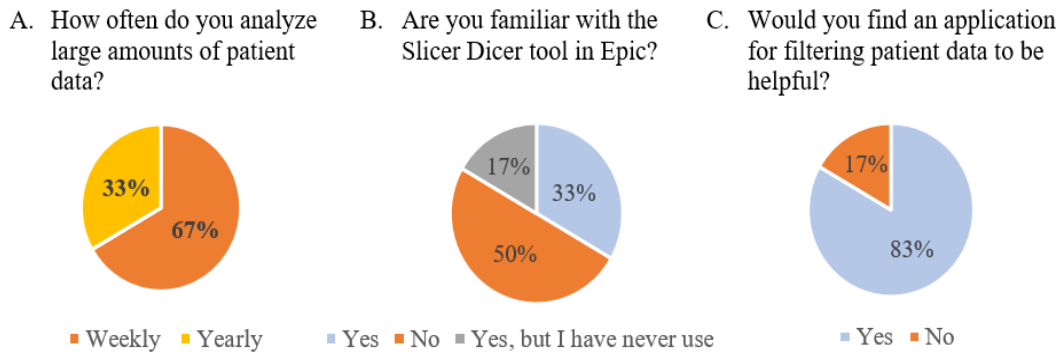


Fig. 2. Motivation for Creating IDDAP. Survey results from clinicians at the University of Virginia Hospital. **A)** The majority of clinicians analyze large amounts of patient data on a weekly basis **B)** Although a Slicer Dicer for data-filtering exists within EPIC, 50% of surveyed clinicians have never heard of it, and only 17% of them have actually used it. **C)** 83% of clinicians would find an application for patient data filtering to be helpful.

It can be noted that one medical research student at the UVa hospital was seen manually sorting through hundreds of patients in Microsoft Excel to isolate patients with a specific criterion. When asked the method they currently use for filtering data, the majority of participants listed manual search and Excel.

Major limitations of current medical research data analysis tools were identified as the inability to perform real-time changes in data displayed, the inability to collect large amounts of data, and the inability to quickly sort through data files. In addition to the limited filtering abilities of pre-existing software, the layout of the software itself is hard to navigate and non-intuitive. More specifically, the Slicer Dicer tool in Epic is limited in its filtering options and forces the user to export a new file each time they wish to change filtering criteria. After completing interviews and conducting the survey, three main program requirements were concluded. The program must possess:

- I. The ability to filter large amounts of data and output results in real time relative to user specified filters
- II. The ability to quickly conduct basic data visualization that changes in real time relative user specified filters
- III. A user-friendly interface that is clean, organized, and intuitive

Program Characteristics

IDDAP features two main sections hosted under two dashboard tabs: data manipulation and data analytics. The File Manipulation tab displays the user-uploaded patient file, which the user can then sort with the filters on the left side of the interface. The Data Analytics tab includes three sections, each associated with either a graph or a set of

statistical measures a researcher might like to have displayed. Adjustments to the filters are reflected in real-time within the data file under the File Manipulation tab, and within the graphs and statistical measures of the Data Analytics tab.

Providing multiple formats for data visualization allows for a wider variety of data analysis and expands the target audience. Although the filter requirements were pre-set to specify those used in the Outpatient Antibiotic Therapy

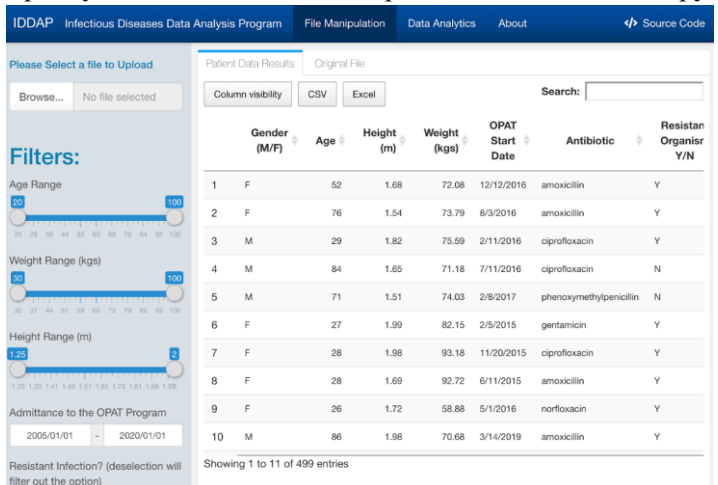


Fig. 3. IDDAP Interface: File Manipulation. The interface has a variety of functions centered around the patient data displayed under the Patient Data Results tab. Patient results can be filtered by the user with the panel on the left side of the interface. Filters provide real-time changes to the Patient Data Results and the Data Analytics tabs. Individual columns can be organized in either ascending/descending order or alphabetical/reverse alphabetical order by clicking the grey arrows next to the column header. The filtered data file can be exported as either a .csv or .xlsx file by clicking CSV or Excel, respectively.

(OPAT) division of the Infectious Disease department, they can easily be altered in future iterations to change relative to the user uploaded file. The IDDAP interface applied to mock data can be seen in Figure 3. Due to the nature of medical research, the program is intended to be run locally on a processor that is certified protected in order to guard sensitive patient data.

Data Manipulation Capabilities

IDDAP’s interactive layout allows users to manipulate data and receive results in real time. The ability to upload data files allows for a wide variety of analyses to occur. This feature allows for the user to select comma-separated values (.csv), tab-separated values (.tsv), text (.txt), and excel workbooks (.xlsx) from anywhere on his or her device and upload it to the program. Once uploaded, the data file appears as a table under the “Patient Data Results” and “Original File” section of the Data Manipulation tab.

The data filters are based off of and specific to the columns of the mock data sheet, one filter for each of the eight columns. There are several different types of filters including slider filters, selection filters, and filter by text. Age, height, and weight are continuous; and therefore, represented as sliders for their associated filters to allow users to quickly assign and adjust a desired range of values. Gender, antibiotic resistance, and risk of readmission are discrete values, and are represented with selection filters, so

that users can easily eliminate or add all of the values next to the check box with one click. A date range input was assigned for the date of readmission, for which users can specify a start and end date from a drop-down menu. As the user modifies the filter requirements, he/she can see the data displayed in the Patient Data File tab change in real time. The data presented under the Original File tab will remain unchanged, allowing users to quickly make comparisons and reference to the original data set without resetting his/her filters

In addition to the filter requirements, users can interact with the table itself. The Column Visibility option above the data table specifies which columns will be shown on the Patient Data Results tab. Furthermore, users are also able to click and drag entire columns of data from left to right in order to better spatially orient categories of data they wish to compare. A set of arrows next to each column header allows for numerical data to be sorted in ascending or descending order; and for categorical data to be sorted in alphabetical or reverse alphabetical order. The ability to organize data within a column allows users to sort the data file by a given category. There is also a search bar above and to the right of the data table to allow for the user to search for a specific data entry or criteria by name. At any point while using the program, users can export the data table under the Patient Data File tab, with all of the changes made during filtering, in either comma separated values (.csv) or Excel workbook

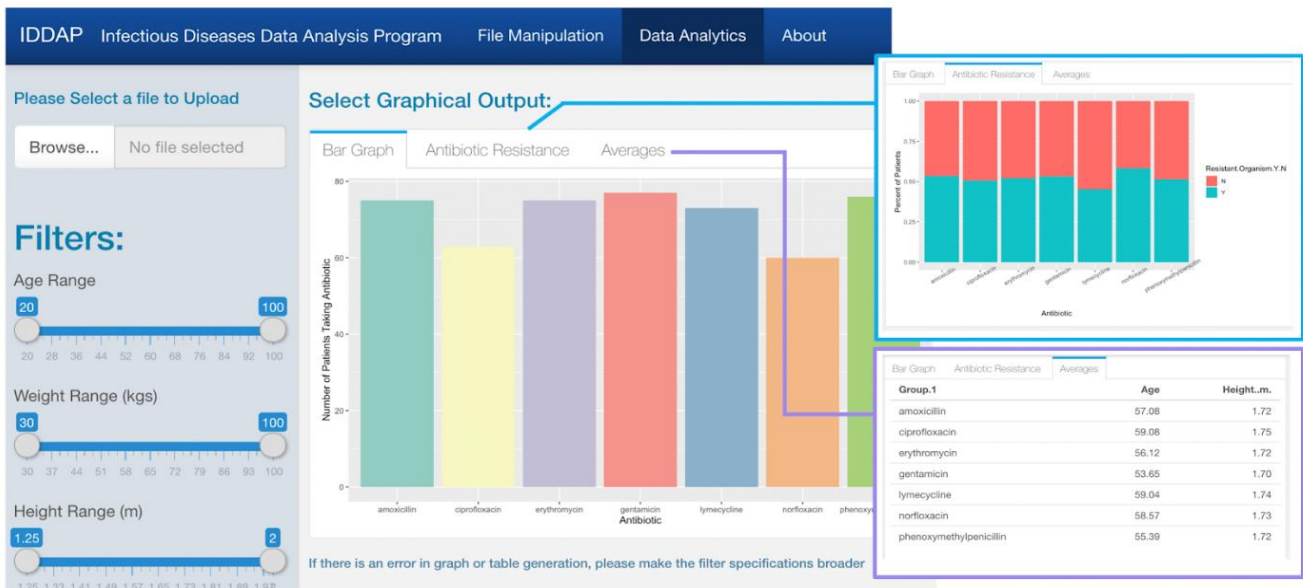


Fig. 4. IDDAP Interface: Data Analytics. The Data Analytics tab of the interface displays different graphical outputs based on the filtering criteria in the sidebar. Graphical output responds in real time to filtering changes. Three tabs are available for data analytics: Bar Graph (left), Antibiotic Resistances, and Averages (right). The latter displays statistical metrics of averages of patient height and age given a certain antibiotic in the patient data set. The two former tabs provide graphical outputs of antibiotic count, and fraction of patients resistant to each antibiotic.

(.xlsx) format by selecting the appropriate button above the data table. The ability to quickly save filtered data files aids in reducing data analysis times and aids medical researchers in their clinical needs.

Data Visualization and Analytics Components

Data analytics and analysis are vital for drawing conclusions and forming a valid argument. The various features encompassed in the data analytics components of the program aid in drawing appropriate conclusions. The data analytics components can be broken up into two sets: data visualization and statistics. The user can access the data visualization and statistics features through the data analytics tab. (Figure 4) If the user wants to focus on data visualization, they are able to do so in the first two tabs above the initial graph. The user can easily switch between tabs to view different types of graphical outputs shown in the various callouts in Figure 4. The user is also able to manipulate these visual outputs through the filters described in the previous section. The graph that is presented will change in real time as the filters on the left of the interface are manipulated. This allows the user to view data trends and comparisons while specifying a set of the data. In the current version of the program there are two ways in which the user can visualize his/her data. The preliminary graph shown is a bar chart that plots the number of patients versus the antibiotic they are taking. By selecting the Antibiotic Resistance tab, the user can also view the ratio of patients that presently have a disease that is resistant to the traditional antibiotics. This graph is displayed in the blue callout in Figure 4. These visualizations were selected as preliminary figures based on the data collected on common analysis clinicians in the Infectious Diseases department at UVa usually perform. The Averages tab displays the average age and height for all patients taking a specified antibiotic, and changes relative to the selected filters.

IDDAP Evaluation

User Performance Assessment- Data Processing Efficiency

A time comparison test was conducted to determine whether a significant difference exists between filtering and graphing data in Excel vs IDDAP. As previously mentioned, clinicians cited time as one of the largest hurdles they had to overcome when using their current method of data analysis. Clinicians indicated time was one of their most precious resources, so it was necessary for clinicians to be able to perform data analysis faster in the prototype IDDAP.

Several tasks were selected based on the typical analysis clinicians and medical researchers perform. The range of times taken (min) to complete each task is larger in Excel (IQR=0.575, 1.80, 4.42 for T1, T2, T3) than in IDDAP (IQR=0.202,0.210,0.518, for T1, T2, T3) for all three tasks. This suggests there is a greater variability in user skills when using Excel as compared to IDDAP, despite only having used IDDAP once. The median times (min) for T1, T2, and T3 respectively were 1.04, 0.835, and 5.13 for Excel and 0.25, 0.33, and 0.50 for IDDAP. (Figure 5) It was shown that the more complex the task, the larger the difference in time between IDDAP and traditional methods. The tasks assigned increased in complexity with Task 3 being the most complex. Task 3 showed the largest difference between the average time to complete the task in Excel versus IDDAP. The average time to complete the third task was significantly faster in IDDAP than in Excel ($p < 0.05$). (Figure 5)

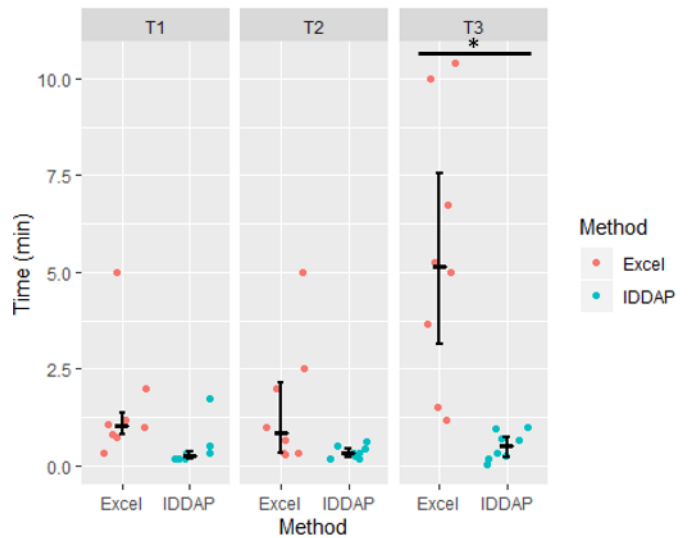


Fig. 5. Task Completion Time in Excel vs IDDAP. The median time taken to complete each task is higher in Excel than in IDDAP for all three tasks; and the mean time difference is statistically significant for task 3. Significance level determined by two-tailed Student's t-test (* indicates p -value < 0.05). Crossbar denotes median, and error bars show the interquartile range of each set of task completion times.

Evaluation of Program Usability

While improving the efficiency of data analysis is the primary goal in creating IDDAP, it is important to ensure that the program is also user friendly. To have a program that accomplishes tasks faster, but is not easy to use would create another problem. Users were asked to assess IDDAP

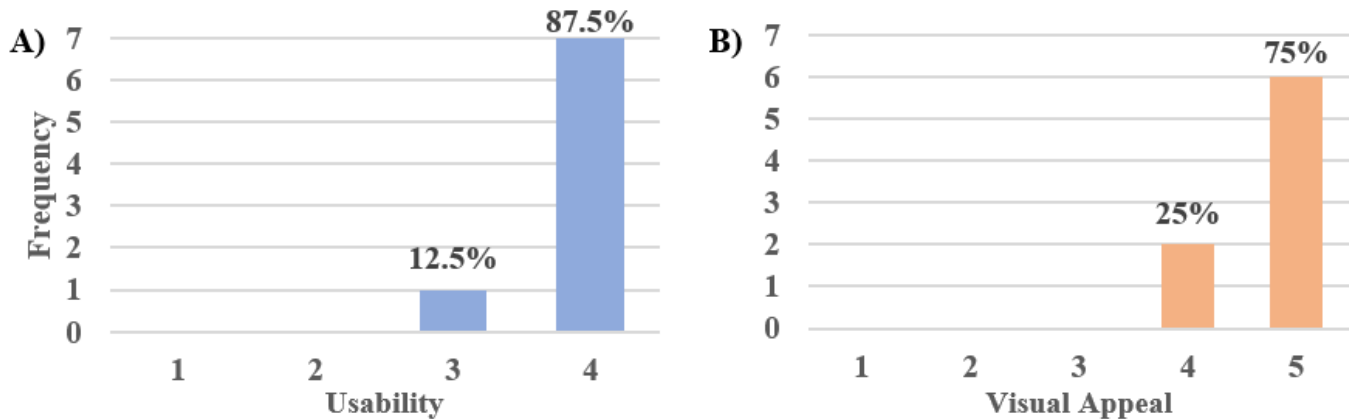


Fig. 6. IDDAP Usability and Visual Appeal. A) User responses regarding the usability of IDDAP. The scores for usability ranged from 1, indicating the program was not user friendly, to 4, indicating a very user-friendly program. 87.5% of users rated IDDAP as very user friendly. B) User responses regarding the visual appeal of the program. The scores for visual appeal ranged from 1, indicating the program was not visually appealing, to 4, the program was very visually appealing. 75% of users rated IDDAP as very clear/organized.

based on two key criteria for creating an overall user-friendly program: usability or ease of use and visual appeal. Both qualitative and quantitative data was collected to ensure these criteria were met. Figure 6 shows that IDDAP performs well in both the usability and visual appeal categories. 87.5% of users surveyed gave IDDAP the highest mark for usability with an average overall usability score of 3.875 out of 4. Users also indicated that IDDAP was visually appealing. Out of 5, users gave IDDAP a 4.75 for visual appeal with 75% of the users indicating IDDAP deserved the highest marks. (Figure 6B) Qualitative data collected indicated that users thought the program was very clear and had an interface that was presented pleasantly. In several survey responses users also noted that the program was very user friendly and organized.

Discussion

The volume of data in healthcare is projected to grow faster than it will in other sectors. Healthcare data is expected to experience a compound annual growth rate of 36% through the year 2025.¹² An emphasis must be placed on managing extremely large data assets in order to keep up with this unrelenting increase in healthcare data. Because EHR systems are effective at compartmentalizing patient data, those systems will continue to be used around the world. However, if medical researchers wish to perform any types of analysis on this data, they will continue to find it difficult to do so with current methods.

The novel program described in this paper effectively organizes and visualizes patient data. This allows for the continued use of programs such as Epic at the UVa medical center while also allowing medical researchers to quickly

and easily perform critical analysis on their patients. The ability to visualize data helps researchers draw conclusions from data that may otherwise go unnoticed. Data visualizations allow for researchers to easily compare sets of data as well as view trends in certain data sets. The ability to identify trends in medical data is crucial for making predictions about prescribed treatments, and thus, aids in improving patient care. In the case of IDDAP, clinicians in the Infectious Disease department will be able to view how variables involving a person's demographic, type of antibiotic, and type of disease affect the rate of readmission in order to reduce readmission for future patients. Charts such as the one shown on the Antibiotic Resistance tab will help clinicians determine whether traditional treatment methods will or will not be effective for a given set of patients. Conclusions such as these are essential for improving the outcomes and quality of care for future patients.

Program Limitations and Future Work

Presently, IDDAP has several limitations related to the narrow focus of the program. However, as a proof of concept IDDAP highlights the ways in which data visualization and analysis in medical research can be improved. Although providing a general sense of the efficiency of IDDAP as compared to Excel, the sample population for the time assessment analysis does not fully encompass the target audience. The studies were conducted during the COVID-19 pandemic in the year 2020, and due to this, access to the clinicians and medical researchers in the Infectious Diseases department at UVa was very limited. In order to accomplish the testing and validation of

the program, IDDAP was tested with biomedical engineering graduate and undergraduate students. Engineers and medical researchers interact with data differently and have differing degrees of familiarity with Excel. Therefore, had they participated; medical researchers may have reported task completion times different from those of the participants in this study.

We recommend all future work regarding this project be centered around broadening the application. The application can be expanded within infectious diseases and broadened beyond infectious diseases. We recommend that the program be modified to allow for more diverse patient data to be uploaded. A wider range of inputs would be crucial for pushing the program to other areas of research. A critical part of this would be automating the collection of filter criteria. This would allow future iterations of IDDAP to have filter options that better coincide with the inputs from the file. We also recommend the program be expanded to include additional statistical functions. This would allow for use by more clinicians performing a wider variety of research than is currently accepted by the program. Statistical analysis is vital for drawing conclusions and forming valid arguments. Expanding on this feature would encourage more researchers to use the program, aid in drawing appropriate conclusions, and thus further improve patient outcomes and quality of care. Providing additional data modelling capabilities would accomplish a similar goal. Features such as data extrapolation would also allow clinicians to improve on their predictions for future patient outcomes ensuring a higher level of patient care.

Materials and Methods

Program Overview

IDDAP, an open source application written in RStudio, focuses on acting as a proof of concept for patient data analysis software that aids clinical data research. The application can be run locally on Microsoft Windows and macOS through RStudio, as well as through a public html link. The program is currently hosted on an RStudio server through shinyapps.io. Using this server allowed for IDDAP to be hosted and distributed online immediately, and eliminated the need for a custom-built server. The app can be managed through the shinyapps.io, and we are able to track usage, capacity, and number of connections. IDDAP's source code is publicly available through the app at <https://iddap.shinyapps.io/iddap/> or on Github <https://github.com/jessmahoney12/IDDAP>.

Designing IDDAP in RStudio

Within RStudio, IDDAP was constructed in an Rmarkdown file, specifically because of a package exclusive to Rmarkdown called "flexdashboard" that achieves both the user-friendly and positive visual appeal characteristics of our program.¹³ More specifically, the flexdashboard package was used to create the sidebar that governs all created dashboards (i.e. "File Manipulation" and "Data Analytics" in the top menu bar of the interface). The "shiny" package was implemented to allow users to make changes to the program in real-time. The following open source packages were used throughout the entire script: flexdashboard, ggplot2, mgcv, shiny, dplyr, readr, RColorBrewer, and DT.

Data Filtering

Flexdashboard comes equipped with a variety of input features that were incorporated into the sidebar of our program, such as fileInput(), sliderInput(), dateRangeInput(), and checkboxInput(). The readr package allows for the fileInput() feature to read csv, excel, and tsv file types that users may wish to upload (Figure 7). Filters were included for variables such as age, weight, height, and admittance date. Specification checkboxes were included for variables such as sex, risk of readmission, and infection resistance. The list of filters and specifications is capable of expanding beyond those provided, especially with application beyond the Infectious Diseases Department, but these were deemed best fit for the purposes of the IDDAP prototype.

User selections within the IDDAP interface are processed primarily through a series of conditional statements within the code. A duplicate copy of the uploaded file is made



Fig. 7. IDDAP Structure. The three essential functionalities that IDDAP achieves. Use of the program starts with uploading a file that contains rows with 1:n patients. Both the file input and features and user specifications include a variety of input statements and allow for immediate reflection of changes to the program output. Program output requires the processing of input selections and rendering of either a statistical or visual representation of the data.

within the code once uploaded, allowing for the alteration of the file based on filtering while still storing the file's original form elsewhere. Both the manipulated file and original file are displayed on IDDAP's interface in tabs. Both files are displayed within these tabs as data tables, using the function `renderDataTable()`. This function paired with the "DT" package enables file exportation, reordering of columns, scrolling in both x and y directions, and removing data by columns, respectively.

Data Visualization

The `ggplot2` package and preexisting functions in RStudio were combined with the manipulated file in order to create the graphs and tables in the Data Analytics tab of IDDAP. `RenderPlot()` and `ggplot()` were used together to create the Bar Graph and Antibiotic Resistance graph, while `renderTable()` was used to generate the table displaying average height and age for patients treated with each antibiotic. `RenderPlot()` and `renderTable()` are reactive elements in that they change visual outputs relative to user specifications.

Performance Assessment- Data Processing Efficiency

Quantitative Assessment of IDDAP

A user performance assessment was conducted to quantify the data processing efficiency of IDDAP. Data was collected from eight individuals from the biomedical engineering department at UVa. Each respondent was given a copy of the mock patient data file and access to IDDAP via an html. Respondents were asked to time themselves completing three tasks (T) with the mock data sheet in IDDAP and in Microsoft Excel, and to report their task completion times. The first two tasks focused on filtering data, while the third focused on simple data visualization:

- T1. Find and isolate all patients with the following criteria:
female and above age 30
- T2. Find and isolate all patients taking erythromycin and have a high risk of readmission.
- T3. Create a bar graph of for the number of female patients above age 30 taking each antibiotic (task 1 patients)

Statistical comparisons of average task completion time were made using the two-tailed Student's t-test. A statistically significant difference in average task completion time was considered at the $\alpha = 0.05$ level.

Qualitative Assessment of IDDAP

Participants were also asked about the usability and visual appeal of IDDAP. Usability was presented on a scale of 1-4 to force a definitive, non-neutral, rating: either a positive (score ≥ 3) or negative (score ≤ 2) user experience. This scoring system allowed for the program developers to classify the usability of IDDAP as either a success or a failure. Beyond organization, visual appeal is not critical for the program's functionality and overall effectiveness. Due to this, the scoring for visual appeal on a scale of 1-5 allowed for users to rank neutrally, thus indicating score of 3, meaning they neither liked nor disliked the programs aesthetic. Participants additionally were given the option to provide comments and feedback on IDDAP which was collected as qualitative performance assessment data.

Experimental Data Set: Mock Data Sheet

The mock patient data sheet used for developing IDDAP was created alongside Dr. Joshua Eby. A team member met with Dr. Eby in order to create an accurate representation of the data typically analyzed in the Infectious Diseases department. This included asking key questions to isolate the vital aspects of patient data files that Dr. Eby and his colleagues use. These questions included but were not limited to: What trends do clinicians look for? Which variables are typically assessed against one another? What is the range of demographics that is seen in patient data files? Secure firewall protection and health system emails were obtained to gain access to patient data-files from the department of infectious disease. A file with appropriate column headers and variables was created in Microsoft Excel. Functions such as the `RANDBETWEEN()` in Microsoft Excel were used to randomly generate values for each variable in order to provide a significant diversity of pseudo-patients. Physiological parameters that affect one another, such as gender and height or age and weight were taken into account while generating values. This ensured that the physiological values of the mock patient data sheet were somewhat representative of a real patient data file.

End Matter

Author Contributions and Notes

M.H., E.G.R., and J.M. designed research, M.H., E.G.R., and J.M. performed research, M.H., E.G.R., and J.M. built the program, M.H., E.G.R., and J.M. analyzed the data, M.H., E.G.R., and J.M. wrote the paper.

The authors declare no conflict of interest.

Acknowledgements

We would like to thank Dr. Joshua Eby and Rena Morse providing for helping us understand the inner workings of Epic and the OPAT program. We would also like to thank Matt Jenoir for his teachings on GitHub and GitKraken use. Finally, we would like to thank Dr. Timothy Allen and Dr. Shannon Barker for their guidance throughout the year.

References

1. Donovan, F. Organizations See 878% Health Data Growth Rate Since 2016. HITInfrastructure <https://hitinfrastructure.com/news/organizations-see-878-health-data-growth-rate-since-2016> (2019).
2. The HITECH Act and Electronic Health Records. Health IS Technology Blog /is/blog/2018/02/13/The-HITECH-Act-and-Electronic-Health-Records.
3. Davis, Z. & Khansa, L. Evaluating the epic electronic medical record system: A dichotomy in perspectives and solution recommendations. Health Policy and Technology 5, 65–73 (2016).
4. Introduction | Meaningful Use | CDC. Public Health and Promoting Interoperability Programs <https://www.cdc.gov/ehrmeaningfuluse/introduction.html> (2020).
5. Health IT Quick Stats. The Office of the National Coordinator for Health Information Technology <https://dashboard.healthit.gov/quickstats/quickstats.php> (2019).
6. Health IT Data Summaries. The Office of the National Coordinator for Health Information Technology <https://dashboard.healthit.gov/apps/health-information-technology-data-summaries.php?state=National&cat9=all+data&cat1=ehr+adoption>.
7. DesRoches, C. M. et al. Electronic Health Records in Ambulatory Care — A National Survey of Physicians. N Engl J Med 359, 50–60 (2008).
8. Bryant, M. Epic, Cerner control 85% of large hospital EHR space, KLAS reports. Healthcare Dive <https://www.healthcaredive.com/news/epic-cerner-control-85-of-large-hospital-ehr-space-klas-reports/553906/>.
9. About Us | Epic. <https://www.epic.com/about>.
10. Eby, J. Clinician Interview: Dr. Joshua Eby. (2019).
11. Morse, R. Data Analyst Interview: Rena Morse. (2019).
12. Kent, J. Big Data to See Explosive Growth, Challenging Healthcare Organizations. HealthITAnalytics <https://healthitanalytics.com/news/big-data-to-see-explosive-growth-challenging-healthcare-organizations> (2018).
13. flexdashboard: Easy interactive dashboards for R. <https://rmarkdown.rstudio.com/flexdashboard/index.html>.
14. Using DT in Shiny. <https://rstudio.github.io/DT/shiny.html>

Appendix

I. Clinician Survey to Department of Infectious Diseases: To understand the clinicians' day-to-day interactions and relationship with clinical data, a survey was sent to clinicians in the Department of Infectious Diseases at the University of Virginia hospital. The full survey, as well as summarized responses are below. Direct quotations were removed and summarized to protect respondent identity

Questions

1. What method do you typically use to sort and filter a given data file? (manual search, apps, help from external party, etc.)
2. Would you find an application for filtering patient data to be helpful? (Mark one)
 - Yes
 - No
3. If you answered "Yes" in the question above, what features would be helpful in an application for filtering patient data? (sorting by admission date, listing patients taking a certain medication, ability to quickly generate graphs, etc.)
4. Are you familiar with the Slicer Dicer tool in EPIC? (Mark only one)
 - Yes
 - No
 - Yes, but I have never used it
5. How often do you analyze large amounts of patient data? (Mark only one)
 - Never
 - Occasionally
 - Every time I need to analyze patient data
6. How often do you request the data analytics team to provide a workup on patient information?

Responses

1. Manual, manual search, Excel, Epic, SQL
2. Yes: 83.3%, No: 16.7%
3. Disease state, lab results, age, gender, race, diagnoses, medications, affiliation with other comorbidities, length of stay, sort/filter by relevant variables
4. Yes: 16.7%, No: 50%, Yes, but have never used: 33.3%

Occasionally: 50%, Never: 50%