MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220177

AuguR: A Scalable Open-Source Interactive Web Application for Routinely Collected Data

Kieran Zucker [a], Millie Wagstaff [a], Cathy Tomson[a,b], Roger Beecham [b], Geoff Hall [a]

[a] *Leeds Institute for Medical Research, University of Leeds, Leeds, West Yorkshire, UK*

[b] *School of Geography, University of Leeds, Leeds, West Yorkshire, UK*

Abstract

Most data collected by hospitals as a consequence of the delivery of routine care is not utilised for analytics or organisational intelligence. This project aims to develop tools to enhance the utilisation of routinely collected cancer data within hospitals across England. This was achieved by developing a web application using open source tools to provide health care professionals and hospital managers with easy to use, interactive analytics for cancer data. The application uses data items hospitals in England are mandated to collect as part of the Cancer Outcomes and Services Dataset (COSD), to provide clinical insight into survival outcomes, population distributions, service demands, waiting times, geographical case distributions and treatment information in real-time or near real-time. Development was guided by end user needs through the use of panels of clinical and non-clinical end users.

Keywords:
Neoplasms, Informatics, Data Science

Introduction

Cancer data such as survival outcomes and waiting times are routinely collected across hospitals in the UK, as part of legally mandated national datasets [1,2]. This information has significant utility for clinical decision making, service evaluation and assessing the quality of care. Whilst organisations such as NHS Digital and Public Health England analyse these data, their outputs can take years to produce and provide limited information at a local or regional level. This limits the utility of these analyses for clinicians as they are not timely enough to facilitate meaningful changes in care delivery and often fail to address the information needs of individual healthcare workers, managers or healthcare providers.

An alternative approach would be to conduct local analysis on local data. This would enable the possibility of near real-time insight and focusing analysis on areas of local need and interest. This is however rarely possible, due to a lack of expertise and infrastructure within healthcare providers [3].

This project aims to overcome these issues by developing software that allows health providers and their staff to obtain immediate insights from their own cancer data without the need for data analytics expertise. To fulfil this aim and be scalable, a number of upfront requirements were defined including 1) The use of open source tools, 2) An ability to provide interactive, on demand user driven analysis outputs 3) To be suitable for clinical and non-clinical end users 4) To be based on data items hospitals already collect as part of national datasets in England.

Methods

An interactive web application and an automated analytics report were developed using open-source R packages. A Shiny framework was used to develop the tool which enables the building of interactive web applications directly from R code [4]. Further customisation and extensions were added using HTML, CSS, and JavaScript. The application is designed to enable visualisation through interactive dashboards, based on data items collected as part of the English Cancer Outcome and Services Dataset, and The National Cancer Waiting Times Monitoring Dataset.

The automated analytics report was created using R markdown. The report, which users can subscribe to, displays Cancer Waiting Times data which is automatically updated and emailed to subscribed key decision makers with a frequency specified by end users.

The web application and automated report rely on several packages :

- Web App: shinydashboard, shinyWidgets, shinyBS and shinycssloaders
- Visualisations and Tables: , plotly, ggplot2, GGally, GGallyx ggfortify, survminer, sunburstR, treemap, d3r, leaflet , UpsetR, DT, data.table, and kableExtra.
- Data Wrangling: scales, grid, sf, rgdal, xlsx, rlang, plyr, dplyr, lubridate, tidyr, tidyverse, tibble, forcats, car, formattable, formatR, rio, zoo and stringr.
- Survival Models: Survive

Development was informed through the inclusion of oncology clinicians and data visualisation experts within the development team. A panel of clinical and management end users was formed and used to guide the development of functionality and design, based on end user requirements. All visualisations presented are based on synthetic data.

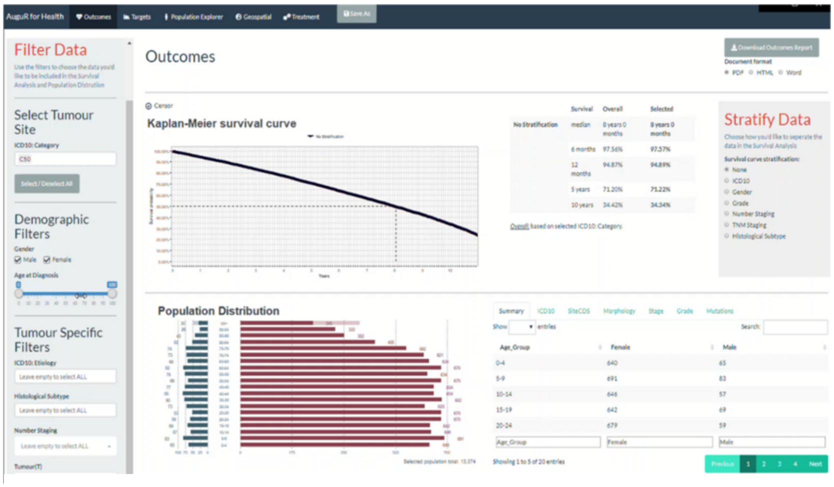

Figure 1: Screen shot of clinical outcomes tab for breast cancer patient data (C50). Image includes data filters, KM curve, population pyramid, case number summary table and comparative survival by time

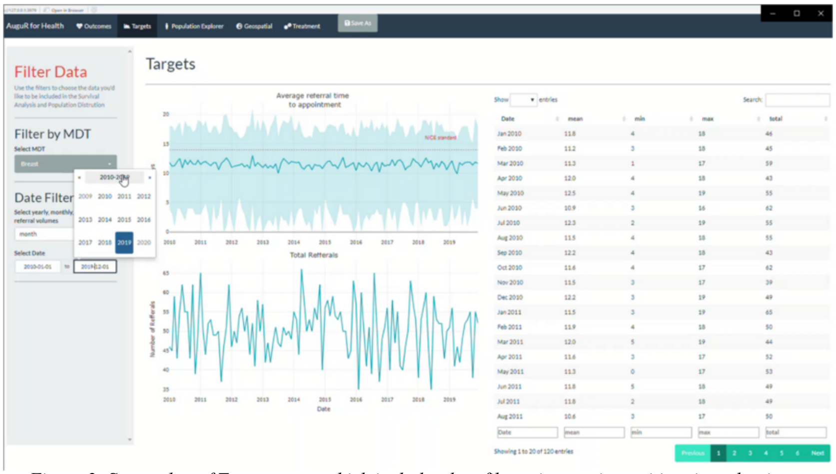

Figure 2: Screenshot of Targets page which includes data filters, interactive waiting time plot, interactive referral volume plot and summary data table

Results

End User Requirements

Across a total of 3 hours of meetings with clinicians and managers, a number of key functions were highlighted as being of particular use. These included; survival outcomes, referral volumes, waiting times, trial feasibility, accessing data for freedom of information act requests and treatment information. Other features requested included being able to share results from the tool with collaborators.

Global Structure of the Web Application

Based on end user requirements the app was developed around five pages: Outcomes, Targets, Population Explorer, Geospatial and Treatment. The pages can be selected using the navigation bar at the top of the application. The center of each page houses the relevant visualisations and a grey filter panel on the left-hand side allows the user to filter the data which updates the visualisations without needing to refresh the page.

Two key functionalities have been developed to enable users to return to or share the application's state i) Bookmark: Ability to store the state on the server and encodes this in a URL. ii) Save As: Allows users to save the state of the application as a rds file and re-upload the settings using the Upload Saved Settings button.

Outcomes

To meet user requirements for survival analysis the app was developed to allow users to apply data filters using a combination of drop down selections, sliders and toggle buttons (Figure 1). Text drop downs are text searchable, for example, to select a cancer site, users may either type the name e.g. “breast” or the ICD-10 code C50. These filters include the cancer site, demographics (age and gender), tumour characteristics (TNM or other clinically relevant stage, grade, morphology, hormone status and tumour molecular profile) and patient genetics (e.g. BRCA status). Options presented within drop down menus are limited to those relevant to the previously applied filters, for example selecting breast cancer will prevent non-breast cancer morphology options being present.

The main area of the tab includes a Kaplan Meier survival curve, population pyramid, table comparing the survival probability of the total population of the selected cancer sites to the filtered population at 6 months, 12 months, 5 years and 10 years, and an interactive table of patient numbers. Changes made to the filters result in these elements being updated to reflect the new selection. The survival curve includes a median survival indicator. The population pyramid includes the entire population of the selected cancer sites with fill of the bars indicating the subsequently selected population through the application of filters. The patient numbers summary table is laid out in tabbed format and is text searchable. Stratification by cancer site, gender, grade, stage and histology can also be applied.

Targets

The Targets tab (Figure 2) was implemented to meet the need to monitor referral volume and waiting times. As with the survival tab, users can apply filters using the left panel which includes the tumour site, multidisciplinary team[5] and time period. Users can also select the time increment by which the data is broken down (annual, monthly or daily). The data selected is used to generate two interactive graphs and a table of the average wait time and total number of referrals. The interactive plots allow users to pan and zoom into areas of interest and display further data labels when users hover their curser over the plot. The data table provided is also text searchable.

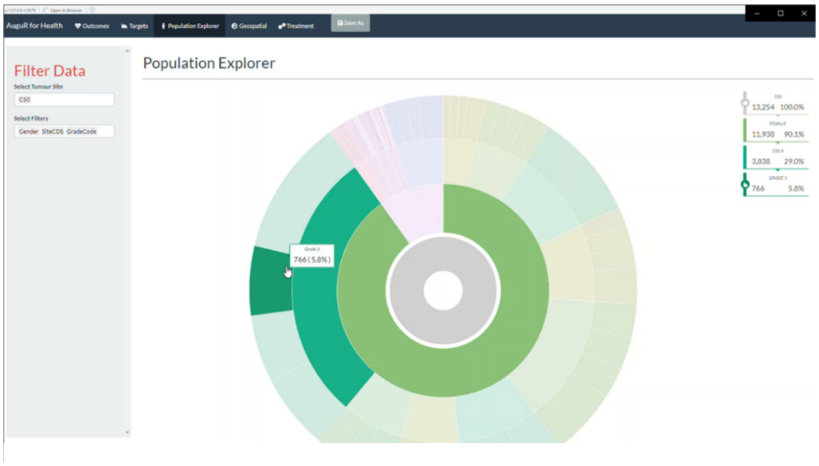

Figure 3: Screenshot of population explorer showing population of breast cancer patients broken down by gender, site and grade

Figure 4: Treatment page showing the number of patients who have received Chemotherapy, Chemoradiotherapy, Radiotherapy and Surgery and combinations of these treatments. The Chemotherapy but-ton has been selected to break this treatment down into intent

Geospatial Data

To allow future comparisons across wide geographies case numbers adjusted for population size were plotted onto interactive maps for the Geospatial tab. Here users can select the cancer population of interest and the fill scale optimized for the range of the data or set to show the full possible range of theoretical values

Treatment Data

The Treatment tab (Figure 4) displays the number of patients who have received chemotherapy, chemoradiotherapy, radiotherapy, surgery and combinations of these treatments. An UpsetR plot was implemented to display this data with a panel of action buttons at the top of the page to allow more granular treatment information broken down by intent (e.g. palliative or curative). This tab houses the same filters as the clinical outcomes tab and can be used to select a specific category of patient to look at.

Automated report

The automated report contains graphs and summary statistics outlining the number patients seen that week, how many patients were seen within the 14-day target, number of referrals made and the average waiting time. A side panel allows the user to navigate to the relevant part of the document which is broken down by multidisciplinary team.

Population Explorer

The Population Explorer tab (Figure 3) was developed to meet the needs for assessing trial feasibility and extracting information for freedom of information act requests. It allows the user to start with any population of cancer patients and apply any order of filtering sequentially to the data. The user is first required to select a starting population of cancer patients by searching for and selecting an ICD-10[6] code using the Tumour Site Filter. The starting population of cancer patients is represented by the grey center of the plot. The user can then divide the data by gender, histological subtype, number staging, TNM stage, grade, and molecular status in any order. Each level of filtering that is added is represented by an additional ring to the plot. Summary information is displayed in the top right corner or can be gained by hovering over the different segments of each arc.

Discussion

The meetings with clinicians and managers confirmed that access to real world, real-time, local data was a significant current unmet need. Despite often having different intended use cases, clinical users of all levels of seniority and hospital managers often highlighted similar sorts of data and information requirements. The exception for this was survival outcomes which was a clearer priority for clinical rather than non-clinical end users.

The provision of local and up to date survival data (Figure 1) allows the users to examine survival trends within their own hospital, along with the flexibility to look at specific sub-populations where there may be a lack of information in the published literature. It also provides the ability to assess current outcomes where clinical trial and other published data may be outdated. This has potential value in service evaluation, auditing performance and also providing data that can be used as part of explanations to patients in clinical consultations. The ability to export and share analyses via web address should further support information sharing and collaboration between clinical and management users.

The targets page provides data that was requested by all end users. This may be used to identify, in near real time, delays in referral and treatment pathways from the average baseline, as well as identifying spikes in service demand. When applied to real world data locally in Leeds, this was used to

drops in the number of referrals to cancer services through the Covid-19 pandemic[7].

The population explorer was developed to present information required for three separate use cases. Clinicians were keen to be able to identify the volumes of particular subtypes of patients treated in their center, to identify whether conducting clinical trials in that group was feasible. The managers needed the same functionality but for the development of business cases and providing information requested as part of freedom of information act requests. By having a single approach to these separate issues, we maximize the utility of the tool for all end users whilst minimising the number of tabs, improving ease of use.

In its current form the geospatial mapping information is of limited use due to the relatively small populations from a single centre. Its utility will improve if adoption occurs across multiple centers allowing the use of larger geographical subdivisions for comparison.

The treatment tab was developed as clinicians wanted to have a more detailed understanding of how tumour specific and demographic variables relate to the choice of treatment for a patient. Further iterations of this tab will allow filtering by time period after treatment and further distinguishing treatment types e.g. immunotherapy, targeted treatment and chemotherapy.

Despite the potential for this tool, there are a number of potential limitations. Firstly, the tool is reliant on the quality of the data on which it operates. Where data is missing or inaccurate then the resulting analyses will reflect this. The tool has however been designed to show users missing data. Once data such as this provides day to day utility for those responsible for collecting data, this may serve as an incentive to improve data capture and coding practices.

AuguR, by design, uses simple analysis approaches which are designed for descriptive and exploratory purposes only. Many clinicians mentioned a desire to combine treatment information into the outcomes tab. Due to the app being designed for users with a range of background knowledge, some concern was raised by end user groups and developers that the inclusion of this and more advanced multivariable survival analysis risked users drawing inappropriate causal conclusions, which could in turn lead to inappropriate changes in clinical services and treatment delivery.

Future work on AuguR will focus on proof of scalability and value. This will be achieved via deployment into a secure cloud environment for use across The Yorkshire and Humber Region and assessing uptake and utilization amongst target end users. This aspect of the project is already underway as part of the Population Health Management work of the Yorkshire and Humber Local Health and Care Record Exemplar. As AuguR is based on mandatory routinely collected data, it could be deployed in other geographies or even applied to the English national dataset.

Although AuguR focusses on oncology data the approach taken could be applied to any number of clinical problems. As such, in addition to adding further oncology functionality, such as mortality within 30 days of oncological treatment, the tool could be adapted to provide similar but tailored information for other areas of clinical practice.

Conclusion

This project aimed to develop tools to enhance the utilisation of routinely collected clinical data within hospitals. This has been achieved by developing an automated report and web application which provides healthcare professionals and hospital managers with easy to use, interactive analytics for cancer outcomes, treatments, waiting times and population make up. The current web application is made up of five pages each containing features and functionalities identified as important by a team of hospital clinicians and managers. As the app is based on nationally collected data items and uses open-source tools there is potential for scaling deployment nationally and internationally, which could ultimately lead to the improvement of patient care and service delivery[8].

Acknowledgments

AuguR is supported as part of the Health Foundation's Advancing Applied Analytics funding programme. The Health Foundation is an independent charity committed to bringing about better health and health care for people in the UK.

Previous funding for AuguR has been provided through the Connected Yorkshire - Connected Health Cities programme.

This work uses data provided by patients and collected by the NHS as part of their care and support

References

[1] NHS Digital, Hospital Episode Statistics, (2020). https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics (accessed April 28, 2021).

[2] Public Health England, Guidance: National Cancer Registration and Analysis Service (NCRAS), (2016). https://www.gov.uk/guidance/national-cancer-registration-and-analysis-service-ncras (accessed November 21, 2019).

[3] A.L. Beam, and I.S. Kohane, Big Data and Machine Learning in Health Care, *Jama*. **02115** (2018). doi:10.1001/jama.2017.18391.

[4] R Studio, Shiny, *Shiny from R Studio*. (2020). https://shiny.rstudio.com/ (accessed April 28, 2021).

[5] NHS Digital, Multidisciplinary Team, (2021). https://datadictionary.nhs.uk/nhs_business_definitions/multidisciplinary_team.html (accessed April 22, 2021).

[6] World Health Organisation, International Disease Classification, (2020). http://www.who.int/classifications/icd/en/ (accessed April 28, 2021).

[7] A.G. Lai, L. Pasea, A. Banerjee, G. Hall, S. Denaxas, W.H. Chang, M. Katsoulis, B. Williams, D. Pillay, M. Noursadeghi, D. Linch, D. Hughes, M.D. Forster, C. Turnbull, N.K. Fitzpatrick, K. Boyd, G.R. Foster, T. Enver, V. Nafilyan, B. Humberstone, R.D. Neal, M. Cooper, M. Jones, K. Pritchard-Jones, R. Sullivan, C. Davie, M. Lawler, and H. Hemingway, Estimated impact of the COVID-19 pandemic on cancer services and excess 1-year mortality in people with cancer and multimorbidity: near real-time data on cancer care, cancer deaths and a population-based cohort study, *BMJ Open*. **10** (2020) e043828. doi:10.1136/bmjopen-2020-043828.

[8] Dashboards for improving patient care: Review of the literature, *International Journal of Medical Informatics*. **84** (2015) 87–100. doi:10.1016/j.ijmedinf.2014.10.001.