Bronwyn Freeman – Module-2 GPH- Digital Health- Prof Dominik Böhler

A Digital Health Visualization: Allostatic Load Index and Perceived Stress

Unit 1: Problem Description

Current health trends would suggest that one of the greatest global health challenges of the current time and the foreseeable future is the escalating incidence and prevalence of Non-Communicable Diseases (NCDs). In response, healthcare system are attempting to shift to a more proactive approach with greater focus on health promotion and prevention. The challenge in adopting these focus areas and prioritizing proactive healthcare lies in the fact that current health systems are designed for curative and disease focused approaches. As such preventive care is not well incentivized and is difficult to see truly realized within current structures. Facing mounting pressure to rapidly mobilize resources to this end, stakeholders are facing unknown territory and also cannot abandon current health systems overnight.

Health systems need simple, cost effective, evidenced based solutions to ease the process of change. Secondary health prevention seeks to timeously identify early signs of disease transition and be followed by early intervention. In the absence of this kind of healthcare poorer health outcomes and quality of life, delayed intervention, disease progression, higher risk of multi-morbidity and complication development and greater health care utilization and costs can be expected. Beyond the impact on the individual, at population scale these issues place vast strain on the healthcare system in terms of costs and resources. Using the concept of allostasis and evaluating individual and population allostatic load enables assessment of health state and early disease transitions. Ideally this knowledge would facilitate higher states of awareness and literacy about preventive health and early and active intervention.

Figure 1: The Body's Tightrope walk: Allostasis vs. Allostatic Load

Allostatic load represents the aggregated toll on the human body due to chronic stressors, both physiological and psychological. ALI-5 which has been previously validated, an evidence-based index score that confers a degree of allostatic load. ALI-5 is based on five key health indicators – heart rate variability, waist circumference, diastolic blood pressure, low density cholesterol and hemoglobulin A1c. A low ALI-5 score (3 or less) denotes less allostatic load and a high score (4 or more) denotes a high allostatic load. Research has found there is but a significant correlation between self-reported stress and the ALI-5 which confirms links between that sustained allostatic load and stress related disease. These simple metrics employed at a individual and population level with the capacity for real time interpretation and the possibility longitudinal tracking, could revolutionize early prevention and health promotion strategies. The ALI-5 uses well known and common health measures that are already in regular clinical use and are relatively inexpensive in terms of administration and data collection, therefore it may be one of the easiest solutions to integrate into existing health systems.

Workflow: Stakeholder Interactions

Whilst there are various overarching arenas that influence health, these stakeholders are identified as being major player in the realm of preventive health care. This infographic illustrates the major and most likely interconnections that stakeholders are likely to have between themselves. Understanding stakeholder dynamics is important as data visualization could be tailored specifically for different groups and would have different impact.

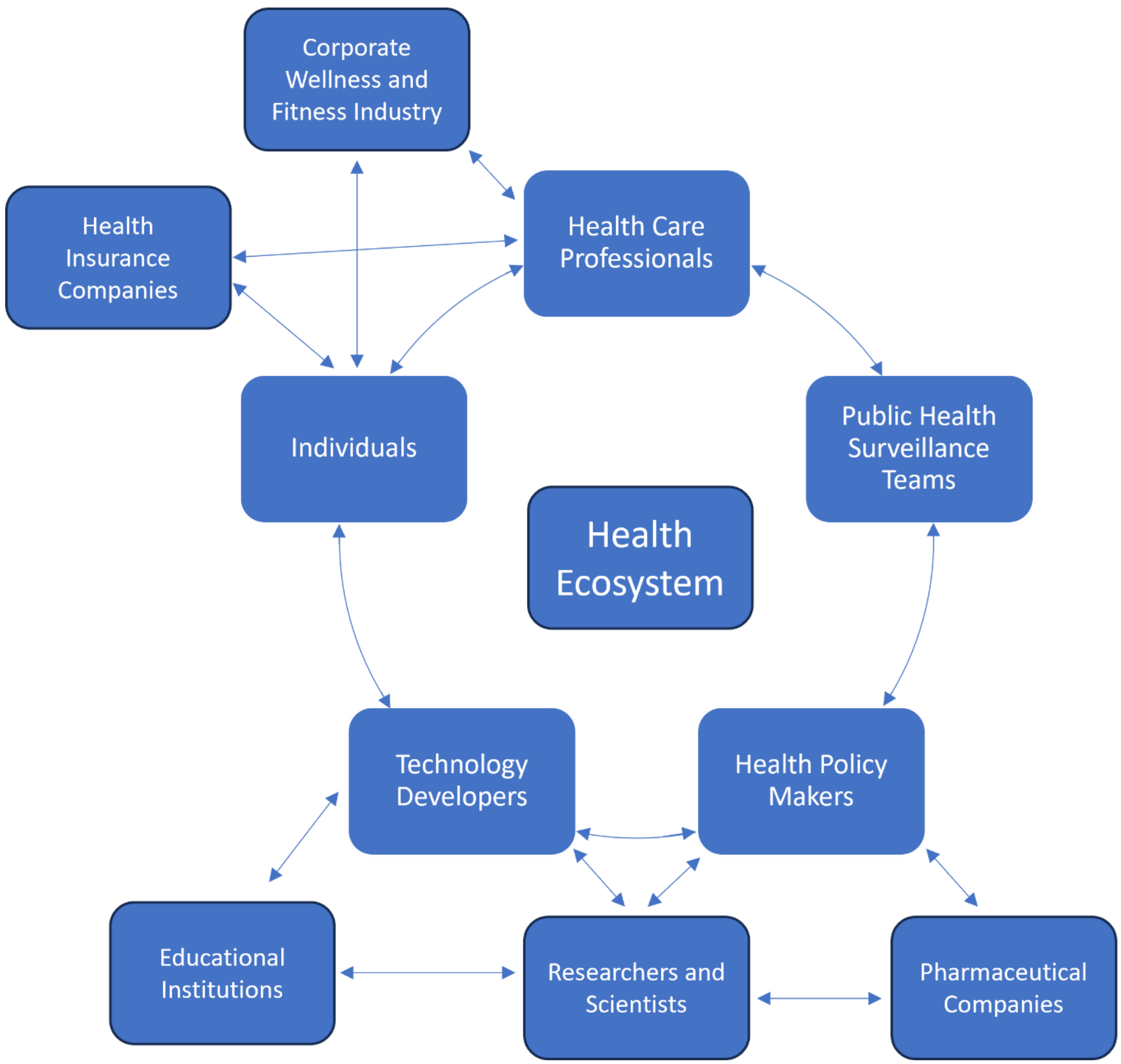

Figure 2: Interconnected stakeholder relationships

Unit 2: Pathway to a Solution: Understanding Stakeholder Challenges and exploring the Benefits of Data Visualization

Stakeholders	Challenges in Preventive Health Care	Impact of Data Visualization
Individuals	The general public may have poor awareness about preventive health care and therefore limited health literacy and difficulty in tracking and goal setting.	Visualizing health data can help individuals understand normative health metric ranges and being able to interpret health data will empower them and facilitate tracking and goal setting.
Health Insurance Companies	Insurance companies may have time constraints in evaluating individual health metrics and lack evidenced based methods with which to evaluate the effectiveness of wellness/prevention overlays offered in a health insurance plan,	Having real time health data allows for insurers to see trends that mark early disease transition and have clear risk assessment. As data collection is relatively inexpensive, insurers can obtain important health metrics in a cost effective way.
Healthcare Professionals (HCPs)	Healthcare professionals may not have the time to collect, evaluate and monitor individual or population health metrics (large shortcoming in capitation based reimbursement). In clinical settings HCPs may have limited resources to do data analysis and interpretation which can hamper assessment of care quality and intervention effectiveness	Using simple data visualization, healthcare professionals can rapidly glean individual & patients' population trends. This allows HCPs to tailor specific interventions for those with high risk. This data visualization is visually intuitive and easy to comprehend making it easier for HCPs to collaborate with each other and their patients, these factors will improve individual and population health outcomes
Researchers and Scientists	The researcher and scientific community do not always have access to diverse and comprehensive datasets. The scientific community often has to transcend many ethical hurdles to conduct studies. Researchers often have difficulty in translating findings into clinical practice	Exploration of aggregated data, trends and correlations in diverse datasets will contribute to scientific understanding of normative and pathological ranges and disease transition. Real de-identified health data removes ethical hurdles for researchers. Researchers have an opportunity to engage in operational research.
Public Health Surveillance Teams	Public health institutions and organizations typically have limited funding for preventive health care as compared to funding for communicable disease. Data often exists in silos and can be fragmented. Public health campaigns may struggle to elicit behavioural and lifestyle changes in the public without compelling yet simple data.	Surveillance teams can observe and map health metric trends so as to assist with priority setting within preventive health care, optimizing limited resources. Data sharing and collaboration can help to break down health care silos. This data visualization provides a tool to facilitate change by visually demonstrating risk and/or impact of interventions undertaken.

Health Policy Makers	Modern day politics and health policy implementation for preventive health care is a challenge if evidence does not exist to support recommendations from the health and science sectors. It is also a challenge to implement needed change where they may be resistance from other stakeholders	Data visualization makes a compelling argument to support the development of policies that address and prioritize the right kind of preventive car and reduce health disparity. Data visualization and evidence of positive changing using preventive health care will foster greater support for more policies that are proactive to healthcare.
Pharmaceutical Companies	Pharmaceuticals may struggle to obtain the relevant information (that proves need) often required to overcome drug development licensing and approval of new medication.	Data visualization may help pharmaceutical companies realize both the areas and degree of need in terms of medication. This data visualization can positively impact medication development, targeting allostasis and constituent risk factors. Trends may also help with medication production planning to avoid under or overproducing.
The Corporate Wellness and Fitness Industry	Many fitness and corporate wellness programs lack objective assessment of health metrics and therefore may struggle with retention of consumers where results and outcomes can't be produced. This may affect adherence for interventions that are known to be effective in the long term. Employee engagement in cooperate and occupational health programmes is low. If employee health is not optimized, employees often have increased incidences of absenteeism and presentism and overall, there are productivity losses.	Increased use of objective assessment and trend tracking by way of data visualisations will assist these industries in the creation of initial buy in and long term engagement and adherence. These industries may be more well position to align their customers to proactive approaches to health. Specific vulnerabilities can be identified in specific working populations and these groups need tailored interventions from an occupational perspective to bolster health and productivity outcomes
Educational Institutions	Schools from a primary to a tertiary level may lack the resources and tools to teach students about preventive health care and the individual and population level impact of undertaking or neglecting preventive health interventions	Students from a variety of ages can begin conceptualizing how disease transition can occur even in the absence of a pathology. Data visualization can highlight to concepts of wellness, the absence of wellness and pathology and how they will differ.
Technology Developers	Developers may lack the medical and scientific background to develop health technologies that are relevant and whose working features actually evidence based.	Access to large datasets, health trends via data visualization can contribute to the development of secure, user-friendly health applications whose features are aligned with science and healthcare

Unit 3: Implementation of the Data Visualization

Step 1: Obtaining a Dataset

Initially, I generated a few synthetic datasets using Synthea and Synthea's Wearable Technology module as I thought that these might generate realistic data for the kind of visualization that I had in mind. I was able to generate 1000 JSON files in the FHIR format standard. I was unable to successfully extract all the variables I was interested in examining (all five health metrics - LDL, HbA1c, HRV, Diastolic Blood Pressure and Waist Circumference simultaneously) from a single JSON file despite a few attempts. I therefore decided to simulate my own dataset in Microsoft Excel as I could comfortably work with and extract data from a csv file in R studio and would be able to (with reasonable clinical background knowledge) be able to generate a reasonably realistic data using a 'determiner' variable with a randomization function applied and extrapolating to 1000 data entries. I aimed to be concise but descriptive in the choice of column headers. After encountering issues in Shiny, which was not able to well handle variable names with spaces(these were the column header names), I opted to change the original variable names to their equivalent variable names with underscores in place of spaces so that Shiny had an easier time of working with the variable names. I saved a 'raw' version of my dataset that retains the randomization functions used for generating the data (called Allostatic Load Dataset Raw.xlsx) and then another version of the file (called ALI.csv) as a csv file type with just static data. This static csv file is the file used in my R Studio Project.

Step 2: Data Cleaning

Once in R Studio, I created a folder called RStudioWork and created a new R project and new R script. Once I set my working directories correctly, I called the ALI.csv file and created a vector which defined the list of column headers in the dataset that require cleaning. For the variables LDL and HbA1c, I created a function so that any commas were converted to periods and converted character values to numeric values. Then a third variable called was established to clean the relevant columns headers defined by the initial function and this was then applied to the specified columns in the dataset and converted all the character values to be recognized as numeric. Lastly missing data was dealt with by being removed, resulting in a cleaned dataset ready for further analysis.

Step 3: Data Manipulation

Once my dataset had been cleaned, I then coded the logic for calculating an allostatic load index score for each data entry in the dataset. ALI scoring considers five health metrics: diastolic blood pressure, HbA1c, LDL, HRV, and waist circumference and is calculated using evidence-based clinical cutoff values outlined by Mauss, Jarczok, & Fischer (2015). These values represent established clinical thresholds and where they are exceeded this represents an elevated health risk. Where the value of each data point falls outside of the defined clinical threshold defined for each health metric or where medication is used to control that particular variable, that data entry receives +1 to the ALI-5 score. The summed index score ranges from zero to five, where a higher score indicates a higher allostatic load which indicates the degree of physiological wear and tear or stress on the body. This logic is applied to each row in the dataset, creating two new variables – the summed ALI-5 score and an ALI risk group. Based on the ALI-5 score a new corresponding ALI Risk group is also created per entry (an ALI-5 score of 3 or less is low risk and a score of 4 or more is high risk). In the code that follows, I created and defined age categories based on the 'Age' variable in the dataset. I created some more user friendly displays names for the select drop down box and labels for the x-axis in the scatterplot and mapped them to the original variable names for better readability from a user perspective and neater graph aesthetics.

Step 4: Creating the Shiny User Interface

I first installed and loaded a package called shiny in my R script. In the first section of my Shiny code, I defined the structure for my user interface. Using the shinyjs package allowed me to use additional Javascript functions to my Shiny app. I also decided to use a Shiny theme called superhero to enhance the aesthetic of my application. I used the prepopulated Side Bar Panel and Main Panel to contain my filtering dropdown selection boxes and my plot as well as some additional sections of explanatory text. I centred the title, changed all text in the user interface to Roboto and adjusted fonts for improved aesthetics. I also created a place holder for the scatter plot within the layout which would be generated based on the user inputs selected. The user interface remains simple but allows users to interactively explore relationships between different health metrics and ALI and filter the scatter plot according to gender and age. The text in the side and main panel clarifies how ALI scoring works, how the plot should be interpreted and how stress and allostatic load are linked to provide extra clarity to users. I rendered the user interface multiple times as an html page to troubleshoot and make edits through the process. In the next section of Shiny code, I setup up the logic for the app to receive the inputs that come from the user interface and define the outputs that will be used to render a plot in the designated output area. With regards to the user inputs the code begins by creating copies of the original data and checks if the user has selected spesific age categories or demographics and if so, then filters the data accordingly. The I use ggplot package to create a function that renders the scatter plot. To do this it pull information about the x-axis variable from the user inputs and some colour coding for the data point based on risk assessment, I applied a small jitter to the data points so that the graph didn't plot superimposed data points and one could get a better visualization. I also did some styling for the plot in terms of text colours, background, legend so that it would aesthetically match the rest of the Shiny user interface and enhance overall user experience. This was done using a package called thematic and once styled could be displayed well as a html web page.

Step 5: Deployment

First, I set up an shinyapps.io account and installed and loaded the rsconnect and shiny packages in my R script. I then retrieved my account details and created and copied the token from the Dashboard in the Shinyapps website. I then configured my account information, Shiny username and my token and secret key. After several attempts to deploy and troubleshooting through failed deployment logs on the Shiny website and adjusting code in my R script I was able to publish and deploy an html webpage with a working URL.

Reflection

Things I did well:

- Generated my own reasonably realistic synthetic dataset using Excel and did learn how to generate synthetic datasets via Synthia/obtain a dataset from a source like Kaggle.
- Was able to troubleshoot variable names between the dataset and Shiny in such a way that the user interface was aesthetically pleasing but so that Shiny could still use original variable names.
- Was able to do basic data set cleaning and preparation for further analysis (replace commas with periods, converting character value to numeric values and account for missing data)

- Was able to identify key stakeholders and their challenges related to this spesific health problem and explore how a data visualization could help address these challenges and enhance intersectoral workflows.
- Had an appropriate choice of data visualization so as to demonstrate the relationship between the variables that give value to the data visualization.
- Was able to write the logic to manipulate data and create new data according to a previously defined algorithm.
- Was able to create and publish a user friendly Shiny web app that had some features of interactivity and was simple to use and understand.
- Was able to apply styling in such a way to create some uniqueness in my Shiny user interface and plot.
- Was able to successfully link and deploy a working Shiny app which has a functioning URL.
- I assisted many classmates on this project by sharing insights, helping with troubleshooting and giving feedback on work. Personally, the process of explaining information actually helped me a lot to clarify and improve my own understanding. I was also helped when I needed it, and it reminded me that in this day and age collaboration is important.

Overall Reflection: I gained proficiency in R studio with basic data cleaning, data manipulation, Shiny app user interface and server logic development and Shiny app deployment. I gained an appreciation for the importance of consistent variable naming and mapping as well as the importance of essential housekeeping in coding (i.e. ensuring that all files are in the same directory and that the app publishing includes all necessary files for proper execution). I was able to apply these skills to develop a data visualization that has meaningful impact on a global health challenge and stakeholders.

Things I did not do well:

- Navigating FHIR standard and extracting data from a JSON file continues to be a challenge and while I opted for a more manageable approach for the sake of time management, I do realize the importance of addressing these challenges in future projects. Electing a simpler approach meant that I could at least keep up with course progression and this justified the creation of my own dataset, as I knew a csv file would be easier for me to continue working on over the duration of the course.
- As a result of creating my own dataset and with the knowledge that none of the data was missing or needed much cleaning, I did not end up doing extensive cleaning or gaining extensive proficiency in things like the tidyverse package in R studio. I also didn't feel I needed much data grouping, and I didn't look very hard into my data types as I felt that I had what I needed to execute.

Overall Reflection: As data cleaning employed was very simple, I am likely to have a very superficial understanding of the processes involved. I am not sure if I really understand data types, choosing what data types ought to be used or how to covert data types. As a result, I might have used data types that are not the most suitable from a technical perspective and therefore my code, may not be the most efficient from a processing perspective. On a small dataset scale with a small dataset, I can essentially get away with this, but I appreciate that in real life settings one is rarely if ever presented with clean data and there will be more difficult types of files that need proper data extraction. In my choice of dataset, to some extent, I have deprived myself of some important learning experiences and thus not really be proficient in extracting real data, understanding data types to enhance coding efficiency and proficient clean data. A more well developed understanding of data types might have improved my project. I used ChatGPT a lot for developing the code and troubleshooting still and often in the beginning I don't understand why code does or doesn't work so it is and may be a crutch for a long time.

Areas for Improvement:

I remain commitment to continued ongoing learning and sharpening of independent troubleshooting and want to undertake the following.

- Take time to understand the importance of understanding data types and data cleaning.
- Consider working with more challenging types of datasets and data files and master extracting, creating, and cleaning data (e.g. JSON files, FHIR standard)
- Strive to understand why code does or doesn't work and try more independent troubleshooting before resorting to ChatGPT.

References

Mauss, D., Li, J., Schmidt, B., Angerer, P., & Jarczok, M. N. (2015). Measuring allostatic load in the workforce: a systematic review. *Industrial health*, *53*(1), 5-20.