BC Chronic Disease Capstone Proposal

Jennifer Hoang, Jessie Wong, Mahmoodur Rahman, Irene Yan

May 13, 2022

Contents

1 Executive Summary

The BC Chronic Disease Registry (CDR) is a data product that captures information about the rates of new and persistent cases of chronic diseases across the province. Age-standardized rates of disease are studied as per different regions, including HAs (Health Authorities) and CHSAs (Community Health Service Areas), as well as for demographic variables such as sex. In this project we aim to create an interactive dashboard that will allow users of all technical expertise to explore and visualize temporal information of the disease rates in the data, and to develop an analysis pipeline that will describe the temporal trends in the data. This proposal will outline the approach we will take to tackle this problem and achieve the project goals.

2 Introduction

Millions of people in BC live with a chronic disease, so it's important to understand and interpret the distribution of disease prevalence throughout the province for a variety of reasons. We may want to know how to best allocate healthcare resources, or to identify if a specific region is experiencing rapid growth of a disease. The dashboard is a tool that will allow healthcare professionals and eventually the general public to access the disease information and answer these questions.

CDR has 3 different types of rates that we will be incorporating into our data product. Incidence Rate is the rate at which new cases occur in a specified population during a specified time period; Lifetime Prevalence is the proportion of individuals who have had the condition for at least part of their lives, and Active Healthcare Contact Prevalence are the cases for which a patient seeks healthcare services for relapsing - remitting conditions. For each disease rate metric, the data is stratified by region at multiple tiers. In this project we will focus on the 5 Health Authorities (HA) and the 195 Community Health Service Areas (CHSA).

2.1 Dashboard

To visualize the spatial and temporal trends of disease rates in the province, we will build an interactive dashboard that will allow users to compare the rates of various diseases in one specific health region over time, as well as to compare how the rates of one disease has varied across several health regions over time. The users will be able to compare between HAs or between CHSAs. We also plan to create an information page in the dashboard containing definitions and descriptions of variables and diseases to increase usability for less technical users. Lastly, we will have a page displaying the data table with filters applied, with the option to download the data as a report. A sketch of the proposed dashboard design is shown below in Figure 1.

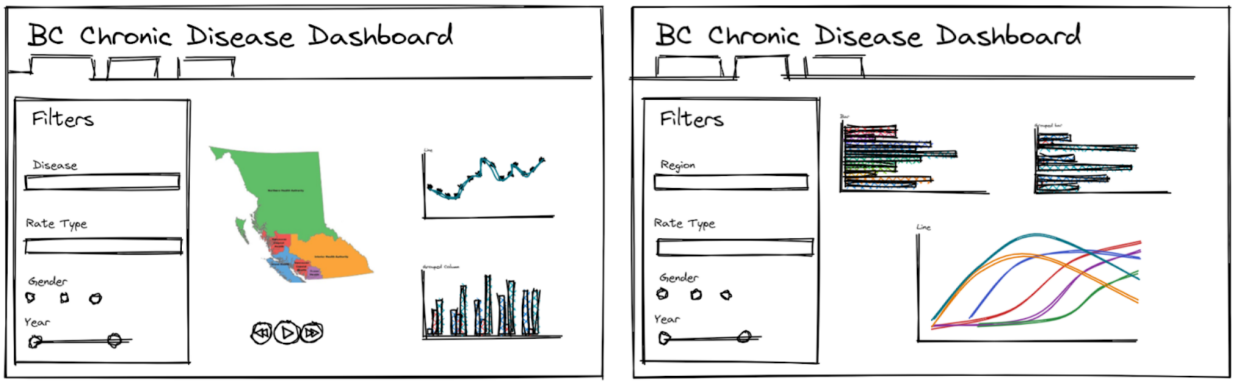

Figure 1: Proposed Dashboard Design Layout. The dashboard will consist of 4 tabs: a disease-specific tab, a region-specific tab, an information tab, and a download tab. The left panel in the image shows the disease-specific tab with map animation and temporal graphs. The right panel in the image shows the region-specific tab with temporal graphs for multiple diseases.

2.2 Temporal Modeling

The Office of BC Provincial Health Officer curates data of ongoing surveillance of 25 Chronic Diseases throughout the whole region (British Columbia province). Currently the dataset has data on Incidence Rate, Prevalence Rates, Geographical and administrative geo-tiers and age-groups. To observe the trend of changes throughout the years of such surveillance data, temporal analysis is the standard choice (Wang and

Wang (2020)). We will explore various methods as per nature of data, established evidences, stakeholder discussions and variables of interest. Eventually, the outputs will be incorporated with the primary R Shiny app.

3 Data Science Techniques

Several different data science tools and techniques will be used throughout the project to accomplish the project deliverable. Some tools are familiar to the team, while others tools we will need to learn. The tools and techniques for each aspect of the project are described in this section.

3.1 Dashboard

We will be using R Shiny to build the framework of the interactive dashboard, and we will be utilizing the leaflet and ggplot packages to assist in generating visually appealing graphs and maps. For data wrangling we will use the tidyverse set of packages to tidy the data in preparation for plotting. We will also be performing descriptive statistical analysis to create informative and explanatory graphs within the dashboard.

3.2 Temporal Modeling

We aim to explore a Bayesian temporal smoothing model for the inferential analysis of chronic disease incidence and prevalence over time. The goal of this model is to generate smoothed estimates with smaller intervals for CHSAs with small populations. A Bayesian approach was selected in order to integrate prior knowledge about the disease rates and to generate 95% credible intervals from the posterior distribution. We propose to use the INLA package for Bayesian analysis for computational efficiency compared to traditional MCMC approaches. (Gómez-Rubio (2021))

Temporal smoothing is appropriate for our data due to the autocorrelation that is observed between disease incidence and prevalence rates from sequential time points in a given region. For each disease, we will compare 2 models for each of the 3 standardized disease rates: a first-order random walk model (RW1) and a second-order random walk model (RW2). A RW1 model will smooth towards the previous value, whereas a RW2 model will smooth towards the previous slope. Therefore, a RW2 model will produce a greater smoothing effect than an RW1 model. (Wakefield (2021 [Online]))

As a baseline model for comparison, we will use the local polynomial regression model (LOESS). The LOESS model is flexible model that acts like a weighted moving average, taking into account neighboring time points within a given span. We will also compare these models to the 95% confidence intervals created using frequentist approaches provided in our original dataset. If the Bayesian models demonstrate an improvement compared to the 95% confidence intervals of the LOESS approach, we will select a final model among the RW1/RW2 approaches.

Model selection will be based on the Deviance Information Criterion (DIC) and the Widely Applicable Information Criterion (WAIC), where a smaller DIC/WAIC indicates a better fit. The final model chosen for a given disease and metric will then be fit on all CHSA regions. Time series plots will then be generated with both observed and modeled disease incidence/prevalence rates for the Shiny app.

4 Timeline

We will form two sub-teams to work in parallel to develop the two final deliverable. The milestones descriptions and respective target week of completion for each sub-team are showed below in Figure 2.

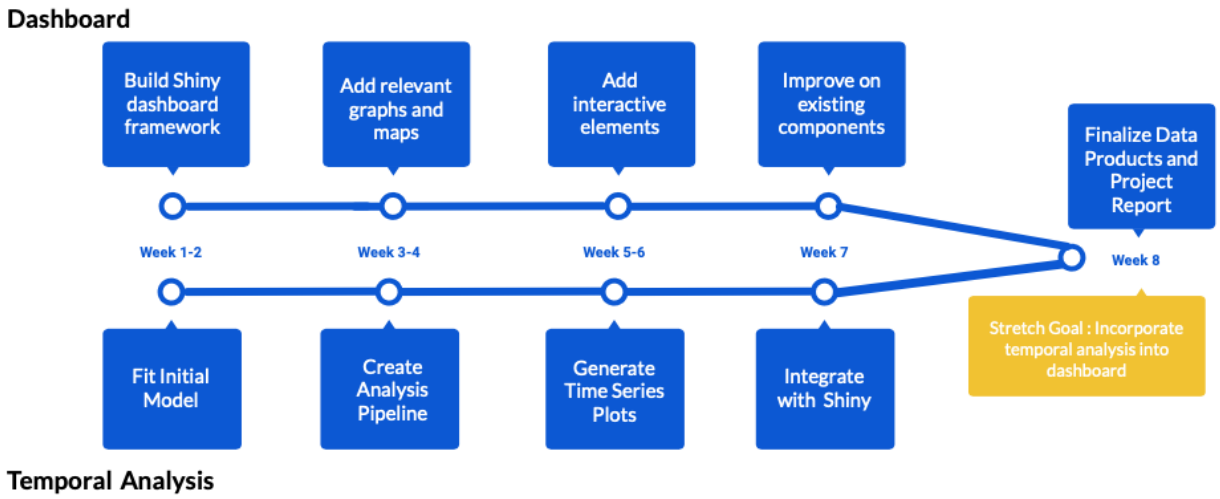

Figure 2: Timeline of Capstone project milestones. Two teams will work in parallel to develop the Shiny dashboard and temporal models over 8 weeks and produce the final data products and reports by June 29.

5 Conclusion

Each year, the Office of British Columbia Provincial Health Officer summarizes statistics on 25 chronic diseases in the region. This document has proposed methods to accommodate the possible inquiries about the prevalence of chronic conditions and their temporal trends. The proposed methods include: (1) an interactive and informative R Shiny dashboard for users to visualize the distribution of diseases with filters for measure, disease, sex, and health authority region, and (2) a data analysis pipeline using the Bayesian temporal smoothing model to infer and display temporal trends of diseases on the dashboard. Spatial trends are also likely to be present in our data, however, due to the limited time frame of this project, we will primarily focus on the modeling of temporal trends. We will develop the final products based on the partner's needs and constant feedback in 8 weeks (Figure 2). The project will include a formal presentation on June 16 and a final report on June 29. We hope the dashboard can better inform Ministers, public officials, and non-expert users on public health issues and chronic diseases.

References

Gómez-Rubio, Virgilio. 2021. *Bayesian Inference with INLA*. Chapman & Hall/CRC Press.

Wakefield, Jon. 2021 [Online]. "Bayesian Subnational Estimation Using Complex Survey Data: Bayesian Inference and Smoothing Model." University of Washington. http://faculty.washington.edu/jonno/AV-Smoothing.pdf.

Wang, Yang, and Jinfeng Wang. 2020. "Modelling and Prediction of Global Non-Communicable Diseases." *BMC Public Health* 20 (1): 1–13.