ImpactR Documentation

R Shiny tool to assess the impact of climate change on waterborne illness

This documentation will enable the user to examine acute gastrointestinal illness risk with properly formatted illness counts. A sample data set is provided to show the capability of R Shiny tool. NOTE: Model coefficients and covariates selected in Stage 1 may require new calibration should the model be applied in climate regimes dissimilar to Metro Vancouver in order to account for contextual variations in local precipitation, case counts, and turbidity.

The operational tool can be found at : [https://vetepir.shinyapps.io/ImpactR/](https://www.r-project.org)

Submitted to the Public Health Agency of Canada as a deliverable for the project “ The Impact of Climate Change on Drinking Water and Health in Vulnerable Drinking Water Systems”

Project MOA No: 4500305199

March 11, 2016

Table of Contents

1. Background 3

2. Instructions for Use 5

Step 1: Installing the required software 5

Step 2: Preparing necessary files 5

a. Input data file 5

b. Procedure to obtain historical precipitation data from PCIC 6

c. Download and process future climate data from PCIC for use with the ImpactR tool. 7

Step 3: Using the features of ImpactR 8

Panel 1. Data Upload 8

Panel 2. Historical Data Patterns 10

Panel 3. Outcome-Exposure Relationships 11

Panel 4. Future Climate Data Patterns 13

Panel 5. Future Impacts 15

Panel 6. More information 16

2. Limitations 16

3. Acknowledgements 16

## 1. Background

This document outlines the features and instructions to use ImpactR, a Graphic User Interface (GUI) enabled statistical tool that enables the user to:

1. perform exploratory analysis of acute gastro-intestinal illness (AGI), drinking water and climate data using tables, maps and graphs,
2. assess the relationship between extreme precipitation, drinking water and AGI; and,
3. estimate the impact of future climate on AGI.

ImpactR has been developed as a part of a scientific study of AGI due to cryptosporidiosis and giardiasis reported over a 13-year period (1997 – 2009) from areas served by municipal drinking water system with surface water source in British Columbia (BC). We have found extreme precipitation to be significantly associated with AGI with time-delayed effects (Chhetri et al. 2017). Further, we used our model for this association to examine how the rate of AGI in the drinking water system’s service area will increase in the future based on future climate projections if no interventions were to be put in place. This research serves as a proof of concept but we believe our tool could be adapted to work in similar drinking water systems elsewhere. However, the current tool is calibrated for the temperate rainforest environment which surrounds Metro-Vancouver. For use in other ecological settings the models must be calibrated independently to account for contextual variations in local precipitation, case counts, and turbidity.

Development: The details of the statistical model development and limitations are described in Chhetri et al. 2017 and the manuscript associated with this website. Basically, we propose a statistical model to describe the relationship between extreme precipitation events and weekly waterborne GI illness using disease and precipitation data for 13 years. We evaluated the ability of the model to predict cases using graphical methods, and give estimates of the mean absolute error based on the fit of the model to our hold out samples. We then used our model to predict future cases based on projected precipitation data available for the entire country on a 300-arc seconds grid (roughly 10 X 10 kms) (PCIC 2015). Annual average estimated future cases for four periods were determined (2020-2039, 2040-2059, 2060-2079, 2080-2099) and then compared to historical average annual number of cases to derive the projected percent change in annual burden of cases.

Applications: The GUI contained with this tool serve the following purpose:

* 1. Explore the relationship between weekly precipitation, water quality and illness data.
  2. Use future downscaled precipitation data from 10 global circulation models to determine general trends (external script, providing links to PCICS for climate model projection data)
  3. Estimate the future burden of illness based on climate model projections.

## 2. Instructions for Use

ImpactR is provided to the user as a folder named ImpactR with R scripts and example dataset. The contents are as follows (with directory structure):

|🡪 ImpactR

|🡪 ImpactR.proj

|🡪 install\_pack.R

|🡪 PCIC\_data\_preprocess.R

|🡪 PCIC\_data\_preprocess\_hist.R

|🡪 sample\_data.csv

|🡪 server.R

|🡪 ui.R

|🡪 RCP\_4.5

|🡪 USER WILL CREATE RCP\_4.5 FILES IN THIS FOLDER

|🡪 RCP\_6.0

|🡪 USER WILL CREATE RCP\_6.0 FILES IN THIS FOLDER

|🡪 RCP\_8.5

|🡪 USER WILL CREATE RCP\_8.5 FILES IN THIS FOLDER

|🡪 HISTORICAL\_CLIMATE

|🡪 USER WILL PUT HISTORICAL PRECIPITATION DATA OBTAINED FROM PCIC IN THIS FOLDER

|🡪 www

### Step 1: Installing the required software

1. Install R for Windows or Mac from

[https://www.r-project.org](https://www.rstudio.com/products/rstudio/download)

1. Install R studio for Windows or Mac from [https://www.rstudio.com/products/rstudio/download](mailto:ssobie@uvic.ca)
2. Open the Impact R folder and click ImpactR.proj
3. From the file menu open install\_pack.R: File>Open File> install\_pack.R. >Open. This will install all necessary R packages required to run the App

### Step 2: Preparing necessary files

ImpactR will use data provided by the user to explore patterns in the data as well as model the relationships between waterborne disease and extreme precipitation. Although not required, the user can also provide water quality information (e.g., weekly turbidity summaries). The input data files must follow the layouts specified below. We encourage users to work with the sample dataset before preparing their own data.

#### a. Input data file

This will be a comma separated file (e.g., foofile.csv) with one row for each week of data (for each disease) and the following columns:

Week: Numeric field given week number e.g., 1 to 52

Disease\_name: Cryptosporidiosis, Giardiasis or Cryptosporidiosis+Giardiasis

Reporting\_date : Date marking the first day of the week yyyy-mm-dd; e.g., 2016-03-18

Cases: Number of cases reported for the week

Precipitation : Precipitation for the week (mm)

*See STEP 2-b below on how to obtain precipitation data.*

Dry.Days : Number of days in the previous two months with rainfall <0.1 mm/day

Turbidity [optional] : Turbidity value for the week. e.g. mean turbidity across 7 days

Week\_no : Week sequence number; e.g., 1, 2, …, 520 for 10 years worth of weekly data.

If turbidity data need not be provided; if not provided, this column should be omitted.

ImpactR will automatically append the following columns to the table based on the reporting date:

Year: Calendar year

Month: Calendar Month

Holiday\_week: Value 1 if a major holiday falls on the week otherwise 0. Major holidays considered are: New Years day, Good Friday, Easter, Victoria day, Canada Day, Labour Day, Thanksgiving, Remembrance Day, Christmas Day, Boxing Day.

Season: Dry or Rainy. Dry if week falls in the calendar months of April, May, June, July, August and September. Rainy otherwise.

Note: Missing values are allowed but data for corresponding rows of the missing values will be excluded from the analysis.

#### b. Procedure to obtain historical precipitation data from PCIC

i. Contact PCIC at [**ssobie@uvic.ca**](http://tools.pacificclimate.org/dataportal/downscaled_gcms/map/)

ii. Provide the latitude and longitude of the location you are interested in to get data for

iii. Provide the start and the end date for your historical period of interest

iv. A download link will be provided to download data in netCDF format

v. Download the data into folder named HISTORICAL\_CLIMATE

vi. Open and run R script – PCIC\_data\_preprocess\_hist.R. This script will process data to a form that Impact R can use. The resulting file will be named as historical\_precip.csv. Use this data to prepare the precipitation column in your input file

#### c. Download and process future climate data from PCIC for use with the ImpactR tool.

At this time ImpactR does not allow for automatic future climate data download. Future enhancements should allow user to click on a location of interest in a map to down future climate data. Below are the steps that will enable the user to fetch and process data so that ImpactR can readily use it.

i. Open the following URL in your browser

[http://tools.pacificclimate.org/dataportal/downscaled\_gcms/map/](https://vetepir.shinyapps.io/ImpactR/)

ii. On the map select the grid that corresponds to the drinking water system source.

iii. Select the appropriate dataset from the dataset selection box. Make sure you download all BCSD model outputs or BCAQ model outputs

iv. Select date range from 2020/01/01 to 2099/12/31

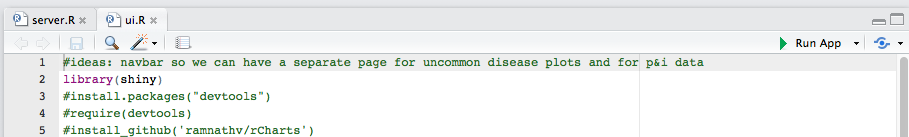
v. Select NetCDF as the Output Format

vi. Download the data into folder with respective RCP names e.g. all RCP 4.5 model outputs should go into folder names RCP\_4.5

vii. Open and run R script – PCIC\_data\_preprocess.R. This script will process data to a form that Impact R can use. The resulting file will be named by RCP scenario.e.g. RCP\_4.5.csv. Move the prepared file to project folder ImpactR from the subfolder

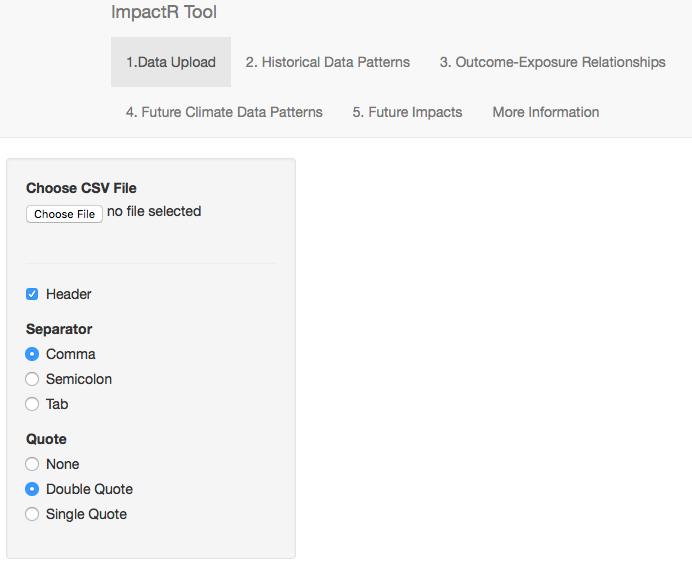
### Step 3: Using the features of ImpactR

The tool is ready to be used and can be deployed by clicking Run App command in R-studio (make sure you have server.R and ui.R open).

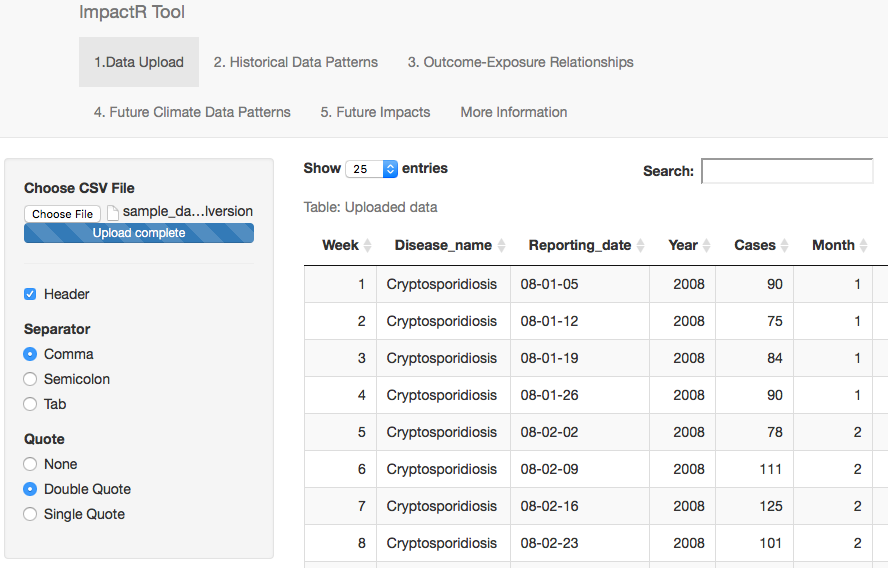


User can upload own data or sample data packaged with the tool to explore the features of ImpactR. Below is a description of each Panel in the tool and what the outputs mean.

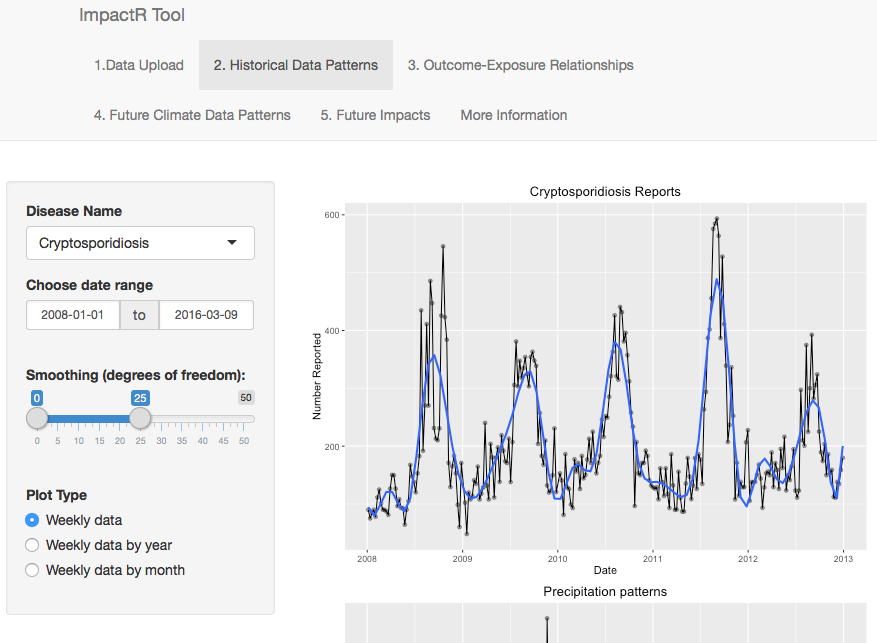
#### Panel 1. Data Upload

****

This panel lets user upload a Microsoft Excel Comma Separated values (csv) file or semicolon/tab delimited files and displays the data. The data should be formatted as specified in STEP 2a. The user is able to interact with the data to display and search specific attributes e.g. display only Cryptosporidiosis.



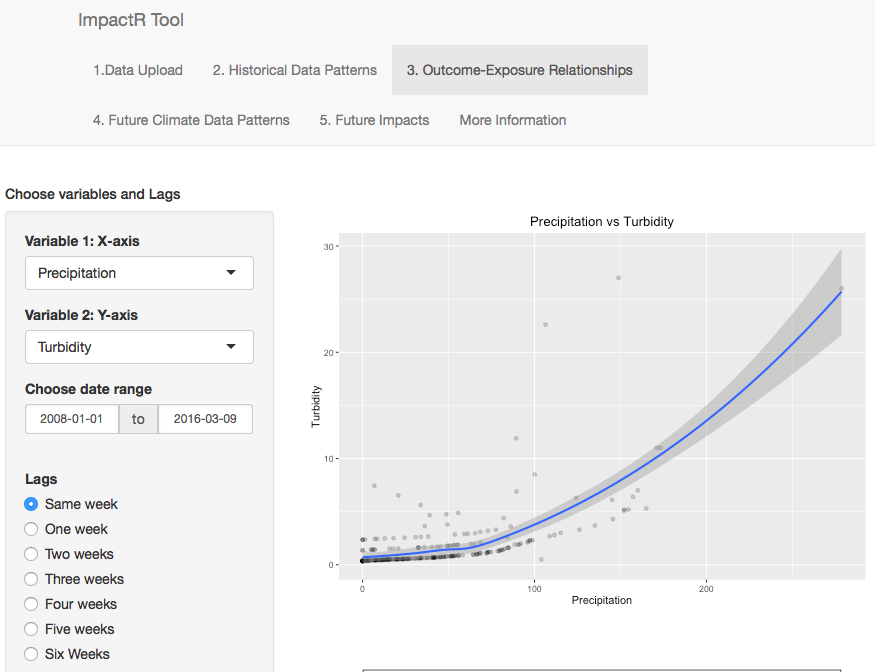
#### Panel 2. Historical Data Patterns



Via point and click, the user is able to generate weekly time series of disease counts, weekly precipitation, turbidity measurements, and previous dry period. These series can further be displayed by year (each 52 or 53 weeks of the year ) or months (boxplots of weekly values by month).

The disease plots will change according to the disease name chosen in the drop down menu. A smoother can be added to the plot to aid in visualising patterns. The user can vary the degrees of freedom used for the smoother from 1-50, with smaller degrees of freedom leading to greater smoothing. The time series display described above also changes according to the user’s selection to display the weekly series, the weekly series stratified by year, or monthly boxplots summarizing the distribution of values for each week across all years in the dataset.

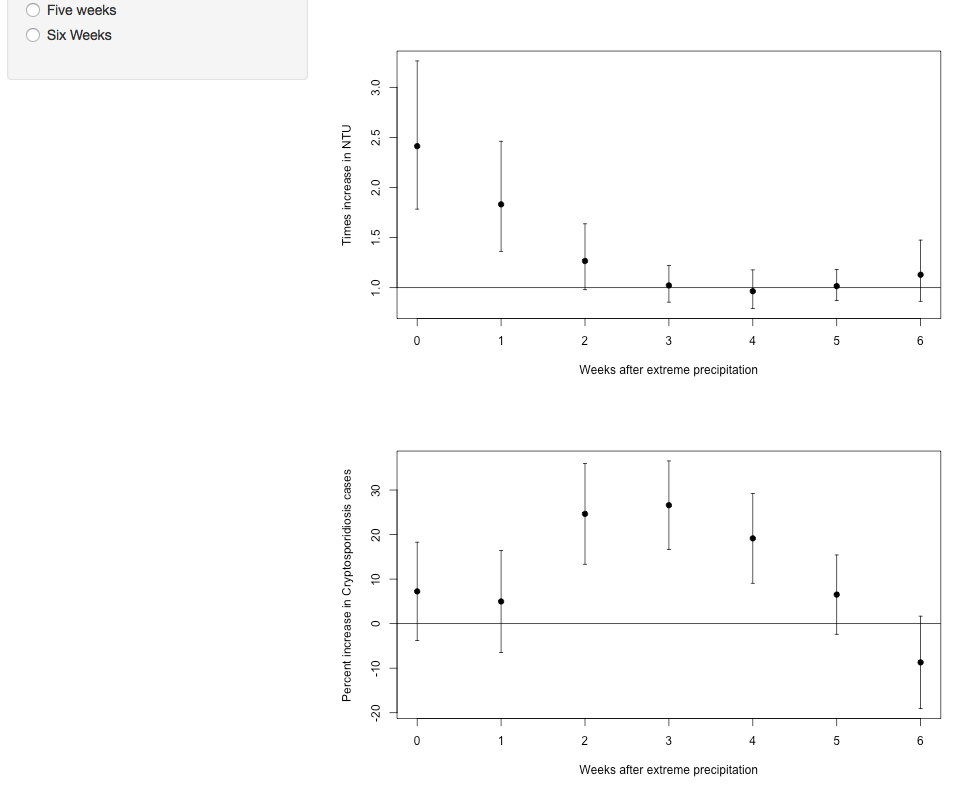
#### Panel 3. Outcome-Exposure Relationships

****

This panel displays the scatterplots for two variables selected by the user; for example, the plot shown in the figure displays the relationship between precipitation and turbidity data. The x-variable is lagged according to the choice selected, from lag 0 (the x-variable is from the same week as the y-variable) to lag 6 (the x-variable plotted is the value from 6 weeks prior to the y-variable). A default smoother smoother is shown in blue, along with a grey band displaying point-wise 95% confidence intervals for the mean value of the y-variable at any given value of the x-variable.

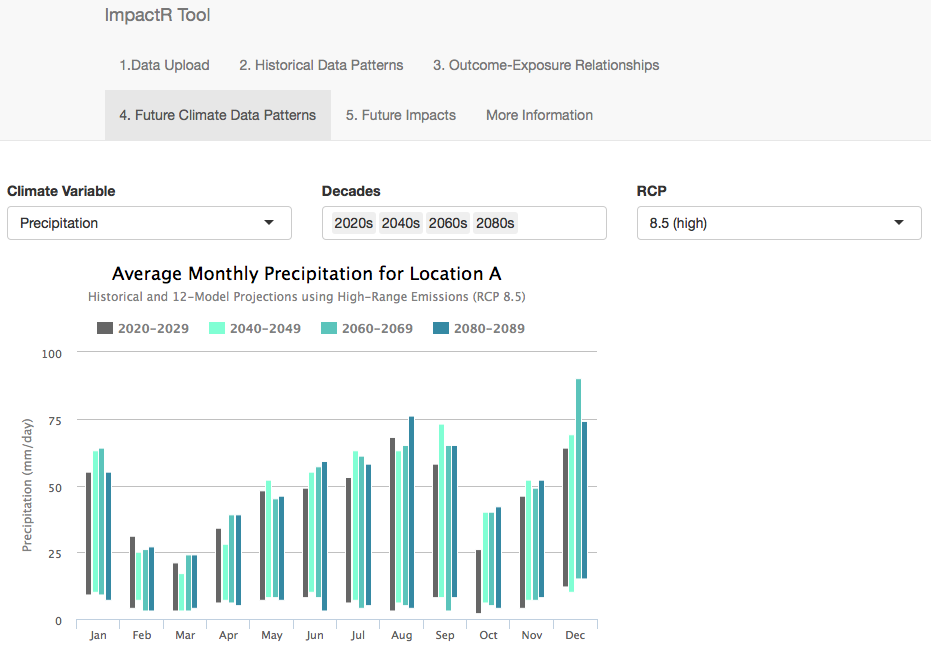
The lower two plots summarize the estimated effects of an extreme precipitation event on drinking water turbidity (upper plot) and on disease counts (lower plot). The error bars represent 95% confidence intervals. The effect estimates are extracted from a distributed lag non-linear model linking the reported weekly turbidity (upper plot) or disease counts (lower plot) to the reported weekly precipitation, with the effects of precipitation being accumulated through up to six weeks. These estimates represent the relative turbidity, or the percentage increase in disease cases, that can be expected given a one-week precipitation value equal to the 90th percentile weekly precipitation relative to a week with zero precipitation. The plot for turbidity will not presented if the user does not provide turbidity data.

Interpretation: In the figures shown here, turbidity values are expected to increase by almost 2.5-fold in a week with precipitation at the 90th percentile of precipitation (i.e., in the week following an extreme precipitation event) compared to a week with no precipitation. Turbidity values remain significantly elevated for at least one week, and they return to near normal values 2 weeks later (Chhetri et al 2017 and this manuscript).

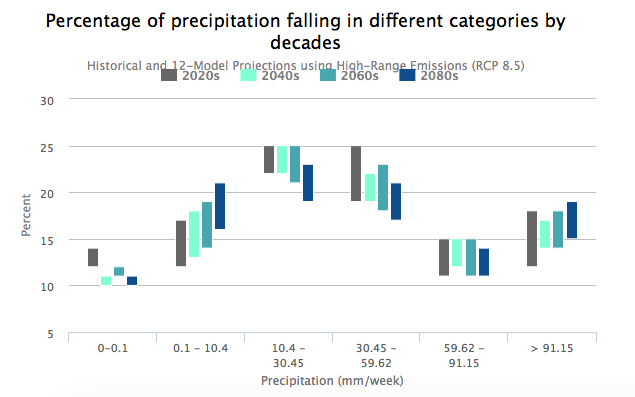


The top figure on this panel shows a scatterplot display of disease, precipitation or turbidity based on the variables chosen by the user. To facilitate the analysis of lagged effects, the user can choose to display the lagged effect of x variable for six weeks. e.g explore effect of precipitation last weeks precipitation on current week’s turbidity. User can also focus on specific date range of interest from the date range selection. However, for plots displaying the impacts of extreme precipitation on drinking water quality (Turbidity) and the impact of extreme precipitation on disease counts, the user cannot make choices as the plots are generated from a statistical model using full time period, precipitation, turbidity and both diseases. For statistical methodology please refer to (Chhetri et al 2017 and this manuscript).

#### Panel 4. Future Climate Data Patterns

****

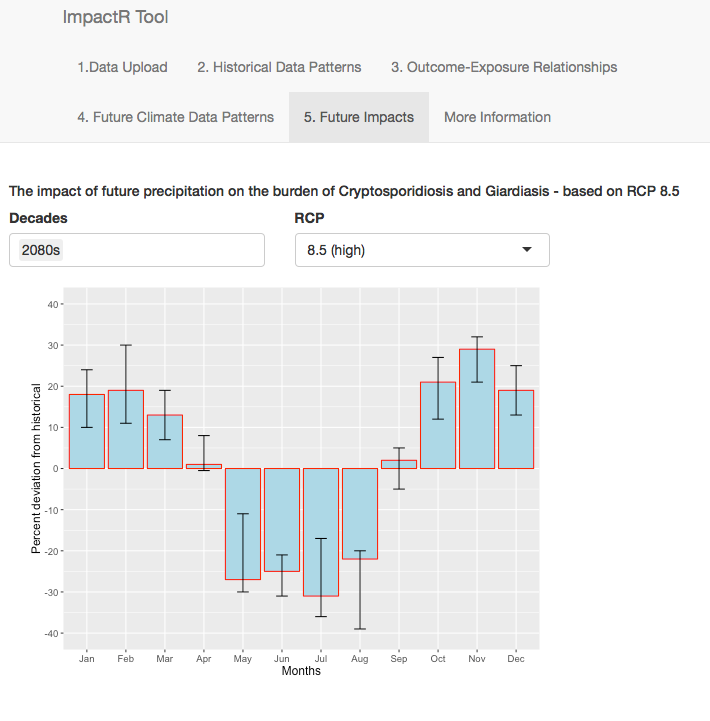
This panel display two plots for the data provided in STEP 2b. The first plots display the average monthly projected precipitation by chosen decades (2020s, 2040s, 2060s, 2080s) and RCP scenario (4.5,6.0,8.5). This plot portrays how average precipitation trends can be expected to in the future under climate change.



The second plot displays the histograms of the percent of projected precipitation falling in different categories (quintiles of historical precipitation) by chosen decades (2020s, 2040s, 2060s, 2080s) and RCP scenario (4.5,6.0,8.5). This helps to understand how extreme precipitation can be expected change in the future under climate change. If extreme precipitation become more common, we can expect to see a shift in distribution patterns with percentage of precipitation falling as extreme will increase.

The user can choose to display data for specific decade(s) – 2020s, 2040s, 2060s, 2080s or specific RCP scenario – 4.5, 6.0, 8.5.

#### Panel 5. Future Impacts

****

This panel displays the percent change in number of disease counts by month compared to historical period i.e. user data input. The bars indicate the ensemble mean of the projections from the 12-climate model. The error bars indicate the range projected by 12 climate models.

#### Panel 6. More information

This panel provides links to the source codes, information on statistical methodology and limitations relevant to the tool, relevant publications, project funders and collaborators.

## Limitations

The tool is flexible enough to model the non-linear relationship between precipitation and illness as well as account for the delayed effect of precipitation on illness. However, the tool in its current form may not be flexible enough to enable the user to choose a different structure of relationship between precipitation and illness. For example, precipitation extremes are defined as 90th percentile; seasonality can only be adjusted by use of second-degree harmonics. Further, we have evaluated only one type of drinking water system; extrapolations to other systems would require validation. The parameters within the tool may be influenced by data characteristics that differ by location because it uses the precipitation-illness relationship at local level using aggregate data. Further, the relationship between environment, host and agent may be influenced by numerous factors that vary geographically. It is prudent to test and validate this tool in a different geographic location or controlling for further factors that influence the precipitation-disease relationship.

The estimation of the future burden of waterborne illness presents considerable challenges. Several factors that may influence how climate change will affect disease risk can be difficult to model due to the lack of data and our limited understanding of future biological (host-agent-environment) interactions. These include changes in drinking water system characteristics (e.g., improvement in water quality and infrastructure), increased adaptation efforts to climate change (e.g., better watershed management, health services) or changes in behavior influencing the risk of disease (e.g., change in the proportion of the population drinking tap water). Additionally, the long-term illness projections (2020-2080) are based upon a relatively short historical period (1997-2009). Our findings need confirmation in other populations using unfiltered surface water systems with longer historical periods. Examining the range of possible outcomes could assist municipalities in prioritizing building resilience to extreme rain events into their water systems.

Our final model, built on historical data, did not include temperature since this was not a significant factor controlling cryptosporidiosis and giardiasis in the historical period (Chhetri et al. 2017). While temperature has consistently been associated with bacterial acute gastrointestinal illness (Fleury et al. 2006), such a link with cryptosporidiosis and giardiasis is less clear (Levy et al. 2016). The model does however account for seasonal variation by including month as a factor in the model which acts as a proxy for temperature. Summer drought may increase the risk of waterborne diseases due to concentration of pathogens that are then washed into a DWS (Charron et al. 2004) a phenomenon also observed in the lagged response of disease to precipitation in the present study. In the future, very high temperatures and the risk of drought in the summer may lead to a muted summer decrease compared with the model we developed.

## 3. Acknowledgements

Funding - Public Health Agency of Canada

Partners - British Columbia Centre for Disease Control, Simon Fraser University

Fraser Health Authority, Vancouver Coastal Health Authority, Pacific Climate Impacts Consortium, R Development Core Team

Software - R-Studio

**4. Reference**

Charron D, Thomas M, Waltner-Toews D, et al. Vulnerability of waterborne diseases to climate change in Canada: a review. *J Toxicol Environ Health A* 2004; 67(20-22): 1667-77.

Chhetri BK, TakaroTK, Balshaw R, Otterstatter M, Mak S, Lem M, Zubel M, Lysyshyn M, Clarkson L, Edwards J Henderson S and Galanis E. Associations between extreme precipitation and acute gastro-intestinal illness due to cryptosporidiosis and giardiasis in a Canadian surface drinking water system (1997-2009). Water and Health. 15:898-907. DOI: 10.2166/wh.2017.100. 2017.

Fleury M, Charron DF, Holt JD, Allen OB, Maarouf AR. A time series analysis of the relationship of ambient temperature and common bacterial enteric infections in two Canadian provinces. *Int J Biometeorol* 2006; 50(6): 385-91.

Levy, K., Woster, A. P., Goldstein, R. S., & Carlton, E. J. . Untangling the impacts of climate change on waterborne diseases: A systematic review of relationships between diarrheal diseases and temperature, rainfall, flooding, and drought. *Environmental Science & Technology*, 2016; 50: 4905–4922.

PCIC. 2015. Statistically Downscaled Climate Scenarios. Available: https://www.pacificclimate.org/data/statistically-downscaled-climate-scenarios [accessed January 12 2015].