# **Code guide: Creating area level B4SC profiles**

# A Better Start National Science Challenge

This document descries the code files used in the “Creating area level B4SC profiles” project in October 2020.

Two additional files are required to run the code:

* ‘Annual areas 2018.txt’: the standard area geography file for 2018 downloaded from the Stats NZ website
* ‘list of var levels.csv’: a manually created file containing each level of each predictor variable with the reference level for each variable flagged.
* ‘dhb\_map’.csv: a manually created file containing the mapping of TALB areas to DHBs (see analytical report for details about how this was done).

These files have been provided alongside the code.

# Programs to be run inside the datalab environment

Author: Sheree Gibb

## “1. Create denominator pop.sas”

This file creates the cohort of children who are eligible for a B4 School Check. A separate output file is created for each fiscal year. A total of eight output files are created for fiscal years 2010/2011 to 2017/2018. Files are labelled with the end of the fiscal year eg “denominator\_pop\_2011\_junyr”.

To run the code you need access to IDI standard tables and dia\_clean

## “2. Create b4sc dataset.sas”

This file takes the output files containing the eligible population that were produced in the previous step, and adds on B4SC completion information and the sociodemographic variables. A separate output file is created for each fiscal year. A total of eight output files are created for fiscal years 2011/2012 to 2018/2019. Note that the year labels in the file names denote the year of eligible population (most of the B4SC checks were completed in the following year). Files are named eg “b4sc\_cov\_2011\_junyr\_full” for the eligible population from 2010/11 (who completed most of their checks in 2011/12).

To run this code you need access to the IDI standard datasets plus dia\_clean, cen\_clean (both 2013 and 2018), the SOCRATES collection (clean\_read\_MOH\_SOCRATES in IDI\_AdHoc), and the following tables in moh\_clean: B4 School Check; publically funded hospital discharges; and PHO enrolment.

## “3. Combine years.sas”

This file takes the eight output files from the previous step and combines them into a single file, labelled with fiscal year. Some additional data preparation is done in this file, including combining data from 2013 and 2018 censuses. The output file from this is called “b4sc\_allyr\_FINAL”.

No dataset access is required to run this code.

## “4.Create datalab output data files.sas”

This file takes the output file from the previous step (“b4sc\_allyr\_FINAL”) and uses it to create aggregate counts for output. Several output files are produced containing counts at dhb level and TALB level.

No dataset access is required to run this code.

# Code to be run outside the datalab environment

Author: Eileen Li

The following R code formats the ouput files and makes the graphs and tables outside of the datalab. It uses the files that were output from step 4 above.

## “Analysis\_Predictor\_Rank.R “

This program determines the order of predictor variables in forest plots and composition plots. To reproduce the plots and tables using current data (2011-2018), there is no need to run this file. In case the order of predictor variables changes when future data becomes available, run this file to obtain updated version of variables list.

## “Data\_and\_Functions.R “

This program imports all raw data files into R and creates functions to generate plots and tables. This code file starts by loading required packages and importing relevant data files, including concordance tables and raw data files. Following that, the section labelled “attendance by module and year” creates functions for attendance plots for each module, at territorial authority (TA) level in general, and at local board (LB) level for Auckland area. There are six functions in this section, namely: myplot\_vht, myplot\_nurse, myplot\_sdqt, myplot\_vht\_lb, myplot\_nurse\_lb and myplot\_sdqt\_lb. The first three are for TA level attendance plots and the last three are for LB level.

The third section “forest plots for each DHB”, creates individual forest plots by module at DHB level. Similarly, three functions are created under this section: pred\_forest\_vht, pred\_forest\_nurse and pred\_forest\_sdqt.

The fourth section labelled “composition of predictors for each DHB”, creates composition plots showing the differences between DHB and national average. One function is created here: comp\_top\_pred.

Note that rate ratios (forest plot), attendance rates (attendance plot) and percentage differences (composition plot) vary across DHBs and hence the x/y limits differed across DHBs. This was incorporated in the functions using if, else if statements.

The last section labelled “characteristics tables” generates one table per DHB, and demonstrates the characteristics of children in each TALB area within the DHB. The functions are called chartable and chartable\_lb.

## “Produce\_Outputs.R”

Produce\_Outputs.R runs the functions created in step 1 in order to output plots and tables. It is comprised of four sections: forest plots, attendance plots, composition plots and characteristics tables.

Note that, for the three in one forest plots, individual plots will be generated first and can then be imported to Adobe Illustrator for manual setting of plot layout and colouring.