**Guide to Formatting Dataset for VERSE methods:**

Formatting:

* Ensure input dataset is in .csv file format

Naming & Structuring Variables:

* The uploaded dataset should include
  + Age of child/vaccine recipient (in months). This variable should be renamed: **hw1**
  + Sex of the child/vaccine recipient (1 = Female, 0 = Male). This variable should be renamed: **b4**
  + Maternal Education Level (typically in years but can be any unit, lower numbers must correspond to lower values and higher values to higher education). This variable should be renamed: **v106**
  + Rural indicator (= 1 if rural, 0 if Urban). This variable should be renamed: **v025**
  + Region indicator(consecutive number from 1 to N, order does not matter but numbering must be consecutive). This variable should be renamed: **v101**
  + Region Names This variable should be called: **geo\_names**
  + Wealth (must be wealth quintile with 1 being lowest and 5 being the highest). This variable should be renamed: **v190**
  + Covered by health insurance (1 = has health insurance, 0 = does not have health insurance).This variable should be renamed: **v481**
  + Survey weights to scale individuals to the nationally-representative level should be named: **v005** if survey weights are not relevant the variable v005 should still be created as variable that takes the value 1 for every observation (a column of only 1s in the .csv file).
  + All missing observations should be blank cell
  + Vaccine outcomes. All vaccine outcomes should take the value 1 if the individual is vaccinated 0 if not vaccinated and a blank cell if missing. Naming of these variables should be:
    - **BCG** = Bacille Calmette-Guerin (Tuberculosis)
    - **DTP1** = Diphtheria, Tetanus, Pertussis (First Dose)
    - **DTP2** = Diphtheria, Tetanus, Pertussis (Second Dose)
    - **DTP3** = Diphtheria, Tetanus, Pertussis (Third Dose)
    - **OPV1** = Oral Polio Vaccine (First Dose)
    - **OPV2** = Oral Polio Vaccine (Second Dose)
    - **OPV3** = Oral Polio Vaccine (Third Dose)
    - **MCV1** = Measles Containing Vaccine (e.g. MMR1) (First Dose)
    - **MCV2** = Measles Containing Vaccine (e.g. MMR2) (Second Dose)
    - **PolioBD** = Polio Vaccine (Birth Dose)
    - **HIB1** = Haemophilus Influenzae Type b (First Dose)
    - **HIB2** = Haemophilus Influenzae Type b (Second Dose)
    - **HIB3** = Haemophilus Influenzae Type b (Third Dose)
    - **PCV1** = Pneumococcal Conjugate Vaccine (First Dose)
    - **PCV2** = Pneumococcal Conjugate Vaccine (Second Dose)
    - **PCV3** = Pneumococcal Conjugate Vaccine (Third Dose)
    - **ROTA1** = Rotavirus Vaccine (First Dose)
    - **ROTA2** = Rotavirus Vaccine (Second Dose)
    - **ROTA3** = Rotavirus Vaccine (Third Dose)
    - **HEPB1** = Hepatitis B Vaccine (First Dose)
    - **HEPB2** = Hepatitis B Vaccine (Second Dose)
    - **HEPB3** = Hepatitis B Vaccine (Third Dose)
    - **HEPB4** = Hepatitis B Vaccine (Fourth Dose)
    - **HPV1** = Human Papillomavirus Vaccine (First Dose)
    - **HPV2** = Human Papillomavirus Vaccine (Second Dose)
    - **JE1** = Japanese Encephalitis Vaccine (First Dose)
    - **JE2** = Japanese Encephalitis Vaccine (Second Dose)
    - **TCV** = Typhoid Conjugate Vaccine
    - **CHOLERA1** = Cholera Vaccine (First Dose)
    - **CHOLERA2** = Cholera Vaccine (Second Dose)
    - **CHOLERA3** = Cholera Vaccine (Third Dose)
    - **MENAC1** = Meningococcal Conjugate Vaccine (First Dose)
    - **MENAC2** = Meningococcal Conjugate Vaccine (Second Dose)
    - **MENB1** = Meningococcal B Vaccine (First Dose)
    - **MENB2** = Meningococcal B Vaccine (Second Dose)
    - **HEPA1** = Hepatitis A Vaccine (First Dose)
    - **HEPA2** = Hepatitis A Vaccine (Second Dose)
    - **VAR1** = Varicella Vaccine (First Dose)
    - **VAR2** = Varicella Vaccine (Second Dose)
    - **EV711** = Enterovirus 71 Vaccine (First Dose)
    - **EV712** = Enterovirus 71 Vaccine (Second Dose)
  + Naming & Creating Additional Composite Outcome Variables
    - **ZERO** = Zero-Dose (Defined as never having received a single dose of DTP, OPV, and MCV by 12 months of age). Coded as 1 if “zero-dose”, 0 otherwise.
    - **FULL** = Fully-immunized for age (Defined as having received all recommends national immunization schedule vaccines appropriate for the age of the child observed). Coded as 1 if “fully-immunized”, 0 otherwise.
    - **Note: the FULL variable (Fully-Immunized for age) must be included in the input data as this is used to define reference levels.**

**Interacting with R-Code:**

* **Line 42**: Set working directory where data file is stored
* **Line 45**: Input the vaccine outcomes desired according to the names of the variables and recommend formatting above. For example if MCV1, DTP3, Zero-Dose, and Fully-Immunized for age were the desired outcomes the VACCINES vector would be entered in as:
  + VACCINES <- c("MCV1","DTP3", “ZERO”, “FULL”)
* **Line 46**: Input the vaccine schedule for the desired outcomes in the same order as they appear in the VACCINES() vector. The SCHEDULE() vector should be numeric and contain the age in months, rounded down to nearest quarter month (if fractional month). Assume 4 weeks = 1 month. For example, a vaccine given at 6 weeks of age would be 1.5 months, a vaccine given at 7 weeks of age would be 1.75 months. Fully immunized has no age requirement so a value of 0 should be given in the SCHEDULE() for FULL. Zero-dose is defined for children 1-year of age and older so a schedule value of 12 should be given for the ZERO outcome. Assuming MCV1 has a schedule age of 12 months and DTP3 has a schedule age of 4 months then the corresponding SCHEDULE vector for the Above vaccine vector containing:
  + VACCINES <- c("MCV1","DTP3", “ZERO”, “FULL”)

Would be:

* + SCHEDULE <- c(12,4,12,0)
* Line 47: GEO should be input as the name of the geographic unit desired e.g. for district
  + GEO = “District”

For Province:

* + GEO = “Province”
* Line 48: COUNTRY = name of the country for the analysis
* Line 49: FACTORS is a vector containing the names of the factors to be included in the analysis. By default this should be input as:
  + FACTORS <- c("region","rural","education","wealth","sex","insurance")
* Line 50: DATA is the dataframe of input data formatted in .csv format with the coding consistent with what is outlined above. The name of the file should be input and correspond with the file’s location in the working directory specified in Line 42.

Running the code:

* Copy and run lines 53-497 to store the VERSE function
* Run lines 44-50 to reset input parameters
* Run line 500 to run the function

Outputs will be stored in the **results** object. To call specific results type:

* **results$NAME**
* The **NAME** of the stored output can be specified with the names below to call the specific outputs:
  + **pie\_MCV1** = pie graph for decomposition of MCV1 (replace MCV1 with any vaccine in the VACCINES vector for corresponding pie graph).
  + **efficiency\_MCV1** = coverage-equity tradeoff plane for MCV1(replace MCV1 with any vaccine in the VACCINES vector for corresponding coverage-equity plane).
  + **VACCINES** = Vaccines examined in the model.
  + **Coverage\_Results** = Coverage of each vaccine at the national level as a % (in order of values in the VACCINES object).
  + **CI\_Results** = Standard concentration index for VERSE Composite Indictor at the national level for each vaccine (in order of values in the VACCINES object).
  + **CI\_E\_Results** = Erreygers corrected concentration index for VERSE Composite Indictor at the national level for each vaccine (in order of values in the VACCINES object).
  + **AEG\_Composite\_Results** = Absolute Equity Gap for VERSE Composite Indictor at the national level for each vaccine (in order of values in the VACCINES object ***(This should be the primary reported value***).
  + **Coverage\_Results\_GEO\_Output** = Coverage of each vaccine at the sub-national level as a % (in order of values in the VACCINES object) for each vaccine and each geographic subunits (in order of values used to define geographic subunits.
  + **CI\_Results\_GEO\_Output** = Standard concentration index for VERSE Composite Indictor at the sub-national level for each vaccine (in order of values in the VACCINES object) for each vaccine and each geographic subunits (in order of values used to define geographic subunits.
  + **CI\_E\_Results\_GEO\_Output** = Erreygers corrected concentration index for VERSE Composite Indictor at the sub-national level for each vaccine (in order of values in the VACCINES object) for each vaccine and each geographic subunits (in order of values used to define geographic subunits.
  + **AEG\_Composite\_Results\_GEO\_Output** = Absolute Equity Gap for VERSE Composite Indictor for each vaccine and each geographic subunits (in order of values used to define geographic subunits (This should be the primary reported value).
  + **FACTORS** = List of factors included
  + **REF** = Reference categories for each Factor