Codebook for the BCFG Simulated Megastudy

**File 1: Patient\_Vaccination\_Status.csv**

This file contains individual-level data on all of the participants who were included in the simulated megastudy. The following variables are included:

* id: a unique identifier for each patient
* condition: the condition to which pharmacy patients were randomly assigned. They were assigned to one of the following 8 conditions:

1. A business as usual control in which patients received no intervention messages. They only received the standard messages from their pharmacy
2. Patients received a text message telling them that the vaccination is “new and improved!” with an image of a vial with the words “new and improved” on it
3. Patients received a text message highlighting the dangers of not getting vaccinated
4. Patients received a text message with a link to a map showing them where their nearest pharmacy is located
5. Patients received a text message telling them that getting vaccinated is “good for society”
6. Patients received a text message telling them that their friends and family will be “healthier and safer” if they get a vaccination
7. Patients received a text message telling them that the vaccination is “reserved” for them with an image of a vial with their name on it
8. Patients received a text message telling them that the vaccination is “easy, quick, and painless”

* gender: patients gender:
  + female, male, or unknown
* race: patient’s race:
  + African American/Black, Asian, White, other, or unknown
* ethnicity: patient’s ethnicity:
  + Hispanic or non-Hispanic
* age: patient’s age, in years, at the time of the intervention
* vaccination\_status: the outcome measure for this megastudy.
  + 0 = not vaccinated
  + 1 = received a vaccination within one month of the intervention.
* store\_id: an identifier of the patient’s primary pharmacy. This maps on to the store-level information included in the second file

**File 2: Store\_Data.csv**

This file contains store-level data for each of the pharmacies from which patients were included. The following variables are included:

* store\_id: an identifier of the patient’s primary pharmacy. This matches the store\_id that is included in the patient-level dataset
* median\_income: the median income of the county in which the pharmacy is located
* region: the US region in which the pharmacy is located
  + Northeast, Midwest, Southeast, West, Southwest
* diversity: a diversity index for the county in which the pharmacy is located
  + Diversity scores can range from 0-1. This index shows the probability that two people chosen at random will be from different race or ethnic groups.