

## Appendix SCM-A: Snowball Study Social Mixing Analysis – Example Code and Files

The Snowball Toolkit has been made publicly available on the website:

<https://sites.duke.edu/dnac/resources/snowballtoolkit/>.

Page | 45

The site, which briefly introduces the Snowball Toolkit, includes the R code and example input files for the Snowball Study social mixing analyses.

Three zip files are linked on the site for these analyses. The first includes example input files (.xlsx and .csv) and R code for the main analyses used to understand social mixing patterns and assess the Snowball Study outcomes, which included its effectiveness in enrolling a representative cohort.

- The social\_mixing\_analysis.R file includes the code to compare the study cohort to the target population; compare mixing patterns among cohort members to what would be expected if people mixed in the community without any assortativity along both categorical (i.e., ethnicity/race) or continuous (i.e., age) attributes of the people; the graphs to compare percent positivity of the study or sampling program being tested to other local testing programs; and the calculation of secondary attack rates within a household or defined setting, both overall and by selected traits.
- contacts.xlsx is the input attributes file that is the basis of the social mixing analyses. It has the attributes of everyone represented in the network: the enrolled participants and the contacts described by that person. For each enrolled participant (SnoID) who took the survey (the respondent), this long-format dataset has one row for the respondent and a row for each contact described by that respondent. Each row has columns to track the referring (CASE), household (HH), and familial (FAM) connections; relationship type between the respondent and cohabitant (Recommend, cohab, cont, fam); enrollment status of each person (Seed, Peer); demographic attributes of the contact being described (gender, age, ethnicity/race, employment status, education level, marital status); and the respondent's assessment of the contact's SARS-CoV-2 status and reason for that assessment. This is test data.
- demographics.csv is the demographic proportions of the background / target population. The file has columns for gender/sex, age band, and combined ethnicity and race. This is real data for Durham County, NC, as estimated for July 2019.
- household\_data.csv contains the pertinent variables for the secondary attack rate calculations. This wide-format dataset has one row for each enrolled participant (sno\_id) who had at least one other cohabitant (up to 7 cohabitants). The dataset

includes a diagnosis date for the referring participant (`target_date`); indicators used for the stratified secondary attack rate calculations (`household_members:predom.var`); and a set of columns for each cohabitant (series of indicator columns for demographics and SARS-CoV-2 status, including a date used for the days between infections [hierarchical symptom onset to diagnosis]); household indicators used for the stratified analyses (`household_response`, `mask_practice`); and the days between the respondent's and cohabitant's SARS-CoV-2 dates (`cohab_covid_daysdiff_1`: `cohab_covid_daysdiff_7`). This is test data.

- [PercentPositive.xlsx](#) has the information used for the percent positivity comparisons. The file has the ID number of the person referred to the study for testing, the dates of sampling, and the results. This is test data.
- [strongWeak.xlsx](#) contains the indicators for whether the contact described was referred by a strong or weak contact. This file has the ID number of the contact described (Record ID) and 3 different examples of ways to stratify strong vs weak contacts (as the *a priori* measure commonly used in the social sciences; by coworker status; and whether the contact was described as a contact in the referring participant's survey). This is test data.

The last two zip files contain the data made available by CDC disclosing number of SARS-CoV-2 tests and positive results by FIPS code by date. These were the basis of our comparison of the Snowball Study's percent positivity against the target population, which was Durham County (FIPS code 37063).

All study data provided in the input files are test data and do not represent actual study data collected.

[Back to top](#)