

Social Mixing Dataset of Contacts and Referrals

Duke RDS²: Respondent-Driven Sampling for Respiratory Disease Surveillance, the Snowball Sampling Study (<https://sites.duke.edu/snowball/>)

An open access, peer-reviewed publication describes the enrollment procedures for this study (doi: [10.1097/PHH.0000000000001780](https://doi.org/10.1097/PHH.0000000000001780)). Briefly, 384 people with newly diagnosed SARS-CoV-2 were enrolled through an academic medical center in Durham County, NC, over 19 months (Dec 2020-July 2022). These seed cases enrolled 125 peer participants from among their social networks. These 509 participants (384 seeds + 125 peers) described a total of 2,199 contacts in their social networks. This dataset contains the contacts and referrals. Peers did not have to be part of the seed social network survey to be referred for participation. The unit of analysis is the pair (respondent->contact or ->enrolled peer). We retained the referrer's ID across multiple records. To protect privacy, did not retain the ID for any person who was a contact or referral of multiple respondents or who was a contact or referral and then chose to enroll in the study as a peer. Therefore, any peers who were described by a respondent have a different ID as a referral than as a referrer.

This pairwise dataset was set up in that way so that the sociodemographic variables repeat for the referrer (person 1) and the contact or referral (person 2). Enrolled participants who did not describe any contacts or make any referrals appear at the end of the dataset as a person 1 with NAs in the person 2 columns. The purpose of this is so that the tables and figures can be reproduced.

Data Dictionary

VARIABLE NAME	DESCRIPTION	CODING	NOTES
studyid.1	Participant ID		[variables].1 are for the enrolled participant who responded to the survey
partype.1	Participant Type	1 = Seed 2 = Peer 3 = contact	studyid.1 should only be a seed or a peer, as only someone who enrolled should be able to recruit peers or describe contacts Note: peer participants could have been referred / recruited and enrolled without being described as a contact in the referrer's (coupon distributor's) survey.
age.1	Age in years at the time of survey	Age in years; ages 80 or above re-coded to 80 to protect privacy	
gender.1	Gender	1 = Male 2 = Female 3 = Non-binary / Other	

raceth.1	Ethnicity/Race (combined into a single variable)	1 = Hispanic ethnicity and White or no race selected 2 = non-Hispanic American Indian or Alaska Native 3 = non-Hispanic Asian 4 = non-Hispanic Black 6 = non-Hispanic Native Hawaiian or Other Pacific Islander 7 = non-Hispanic White 8 = Multiracial / multi-ethnic 99 = Not reported / unknown	Combined ethnicity and race variables
educ.1	Education level	1 = Never attended school 2 = K-8th grade 3 = Some high school 4 = High school or equivalence (GED) 5 = Associate's / 2-year college 6 = 4-year college 7 = Graduate school 99=unknown	
inshelter	Was this pair enrolled in a shelter?	1=Yes, 0=No	Shelter participants were excluded from the community mixing calculations, but not the peer recruitment matrices
studyid.2	Participant ID		[variables].2 are for the enrolled participant who responded to the survey Note: rows where there is a studyid.1 but no studyid.2 information are people who did not describe any contacts or refer any peers
parttype.2	Participant Type	1 = Seed 2 = Peer 3 = contact	studyid.2 should only be a peer or contact, as this pairwise dataset was adjusted to remove the ability to recreate the multiple waves / steps in the network Note: peer participants could have been referred / recruited and enrolled without being described as a contact in the referrer's (coupon distributor's) survey.
age.2	Age in years at the time of survey	Age in years; ages 80 or above re-coded to 80 to protect privacy	[variables].2 are for the person who was described as a contact or referred by the respondent as a peer participant (coupon recipient)
gender.2	Gender	1 = Male 2 = Female 3 = Non-binary / Other	

raceth.2	Ethnicity/Race (combined into a single variable)	1 = Hispanic ethnicity and White or no race selected 2 = non-Hispanic American Indian or Alaska Native 3 = non-Hispanic Asian 4 = non-Hispanic Black 6 = non-Hispanic Native Hawaiian or Other Pacific Islander 7 = non-Hispanic White 8 = Multiracial / multi-ethnic 99 = Not reported / unknown	Combined ethnicity and race variables
educ.2	Education level	1 = Never attended school 2 = K-8th grade 3 = Some high school 4 = High school or equivalence (GED) 5 = Associate's / 2-year college 6 = 4-year college 7 = Graduate school 99=unknown	

The project described was supported by Grant/Cooperative Agreement Number 75D30120C09551 made to Duke University from the Centers for Disease Control and Prevention (CDC), US Department of Health and Human Services (HHS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC or US HHS. Dana K Pasquale (PI) was also supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), NIH, Grant Award 1R25HD079352-01A1.