ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 36975

Understanding How Personal Networks Change: Wave 1

Claude Fischer
University of California, Berkeley

Data Documentation

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

The University of California Social Networks Study

Funded by NIA R01AG041955, "Understanding How Personal Networks Change"

Data Documentation – UCNets – Wave 1

December 5, 2017

This document is subject to revisions. Please check back by visiting the UCNets website, http://ucnets.berkeley.edu.

Principal Investigator:
Professor Claude S. Fischer
Department of Sociology
University of California
Berkeley, CA 94720-1980

Director:
Dr. Leora Lawton
Berkeley Population Center
University of California
Berkeley, CA 94720-2120

Table of Contents

Preface

This document details the research design, sampling frame, weighting and data usage. It is designed to be used with the survey instrument and the codebook, both available online at ucnets.berkeley.edu. Because this study is still ongoing this document is subject to change. We released this guide to facilitate use of the Wave 1 data, which are now available.

I. SAMPLE DESIGN

1.1 Introduction

The data from the UC Berkeley Social Networks Study (UCNets) are part of a five-year NIA-funded study, "Understanding how Personal Networks Change" (NIA R01AG 041955), focusing on two age groups likely to experience life transitions: 21-30 year olds and 50-70 year olds. The sample was drawn from households in the six-county San Francisco Bay Area. The first interview, Wave 1, was fall 2015 – winter 2016. The re-interview for Wave 2 was conducted in Feb-May 2017, and Wave 3 is scheduled for Feb-May 2018. The goal of this study is to understand how network composition and character change over time following life transitions, such as college graduation, marriage, retirement or widowhood. The survey contains items regarding households, personal networks, family milestones, employment, health status and behavior, personality and demographic characteristics.

1.2 Definition of the Population

The population of interest is the adult population of the United States. However, the actual survey population will be limited to adults aged 21-30 and 50-70 who reside in one of six counties in the San Francisco Bay Area (Alameda, Contra Costa, Marin, San Francisco, San Mateo, and Santa Clara counties). We limited the geographic range because the name-eliciting portion that forms the foundation of the network study is not feasible for telephone surveys and wanted to conduct face-to-face interviews because they are deemed to provide much better data quality. Eligibility for the survey was also limited to those who speak English or Spanish. In the end, only one interview was conducted in Spanish. Census tracts with a high proportion of residents who do not speak either of those languages were excluded from the sample.

Although the survey population restricts the study sample to a subset of the population of interest, we do not expect the results to be seriously biased. The types of analysis to be carried out are quite robust, and we expect the results to approximate what we would have learned from a more extensive (and expensive) sample.

1.3 Sampling Frame

The households for the study were selected from address lists maintained by the U.S. Postal Service, and acquired from Marketing Systems Group, an official vendor of the address-based sample. Addresses were sampled from those lists, within certain census tracts selected as primary sampling units. The sampling frame for selecting census tracts was the list of census tracts for the six-county area of the study taken from the 2010 U.S. Census. Each tract on the list has geographic codes and the number of households in the tract. In order to enable analyses that looked at neighborhood effects, the sample was drawn from 30 specific census tracts, the census tracts sorted, ten into each three strata. The strata and pool of possible tracts were defined (by the P.I.). These strata are city; inner suburban, and outer suburban: City centers of San Francisco, Oakland and San Jose, inner suburbs as those within 25 miles of the three centers and outer suburbs, those tracts further than 25 miles. The outer suburbs tracts typically are "over the hills," i.e., over the East Bay Hills, through the Caldecott tunnel; on or over the hills on the Peninsula (e.g., Woodside; Pacifica); in or over the Marin hills.[1] A cluster sample design within strata was by census tract initially selecting 10% of the households but was increased to 42% because of recruitment issues, discussed below. This stratification was based on a number of social indicators. Separate samples of tracts were drawn from each of the three major strata. Within the major strata the census tracts were sorted by geographic codes, to enable a systematic selection of tracts distributed across each of the counties in the frame.

In the first stage, 120 census tracts were allocated to the three major strata, selected in each stratum with probability proportional to the number of housing units in each tract, with a random start, from the list of tracts sorted by geographic codes, to ensure geographic dispersion throughout the counties in the study. Within each selected tract, we selected approximately 100 addresses with probability inversely proportional to the number of addresses in the tract, with a random start, from to ensure geographic dispersion throughout the tract. A random selection of 30 addresses in each tract was designated as the initial sample, and the remaining addresses put into random order, to be used in the fieldwork as needed. The center city tracts represented 34% of the six-county households; the inner suburban ones, 45%; and the outer suburban ones, 22%. To establish equally large strata in the respondent pool, we sampled the center tracts, proportional to population, at 0.93, inner suburban at 0.79, and the outer suburban at 1.49. *Note that, in end,* we rely on post-stratification weighting (described below) to estimate data for the entire region's (English- and Spanish-speaking) 21-to-30 and 50-to-70-year olds pooled. The location-based strata designations and weighting are moot for the Facebook recruits we used to fill out the 21-to-30 year-old sample. Weights by strata are available for each ABS case from the project in addition to the general post-stratification weights described below.

All geographic data – tracts, cities, counties, zip codes – have been excluded from the public-use data sets.

II. RECRUITMENT INTO THE STUDY AND DATA COLLECTION

2.1 *Procedures for Screening, Enrolling, and Assigning Respondents*

The first wave of UCNets was fielded from May to December, 2015, with additional cases completed in January, 2016. Using an address-based sampling (ABS), potential respondents were sent a letter inviting them to participate in the survey. Where feasible, names of residents were appended to the letters and addresses. We had phone numbers for about 40% of the contacts and attempted follow-ups with them by phone to encourage participation. The letter directed them to call a toll-free number or visit a website to take a screener survey to confirm age group, as well as randomize the selection by age group

[1] Complications: To take into account actual travel access of neighborhoods to the center city, tracts that would otherwise be outer suburban were classified as inner suburban if the drive time (as determined by Google Maps) was <= 25 minutes from either downtown San Francisco or downtown Oakland. That left assorted individual cases of tracts that had be assigned in an ad hoc manner. Details are available from the PI.

and within age group. A somewhat higher proportion of young persons was initially programmed (2 of 3 potential respondents), but as the study came closer to reaching the older cohort target the ratio was changed more in favor of younger recruits. Then qualifying respondents were randomly selected such that 3 out of 4 were directed to a face-to-face (FTF) survey and the others to a web survey.

The screening procedure informed the interested party about the study–including the financial incentive[2]--checked to assess his or her eligibility, randomly selected a respondent from a household when more than one qualified,[3] presented “consent” information and received agreement, assigned the would-be respondent to mode, made any necessary adjustment to the assignment (this was rare), began the process of scheduling an interview for those assigned to the in-person condition or provided a link to the web survey.

The face-to-face survey and web survey each took approximately 1 hour[4] to complete. Respondents were given $25 for the first survey, $35 for the wave 2 survey, and $50 for the wave 3 survey. Each FTF survey was also recorded with digital audio if the respondent approved (414 did in the first wave out of 647 FTF interviews). The two interview instruments were as identical as possible, given the method of presentation. Both versions are available to users.[5]

The FTF data collection was done by a San Francisco firm, Nexant, Inc., with trained interviewers. Nexant also staffed the toll-free number and the study email account (ucnets@berkeley.edu). For both web and FTF surveys, the software was provided by the Center for Economic and Social Research at the University of Southern California. Bart Orriens of CESR programmed the instruments. Wave 3 FTF data will be collected by The Henne Group, also in San Francisco, as Nexant management decided to terminate the data collection unit in their business.

2.2 The New Hard-to-Reach: Young Adults

Addressed-based sampling sufficed to yield the anticipated 50-to-70 year-olds, but not the 21-to-30 year-olds, who are notoriously hard to reach. (See below.)We may have been at about a 1.5 percent completion rate with that group, when we turned to other methods. We tried, in Santa Clara County, to use colorful postcards to get more young people, but the postcards had the effect of encouraging a few older respondents to respond, and we saw no lift in response from younger respondents. We then expanded the outreach in two ways: We added a few dozen more respondents by getting referrals from existing panelists and added 290 more by soliciting 21-to-30-year-olds in the metropolitan area through Facebook advertisements. (See below.) These recruitment differences are another reason for analyzing the age cohorts separately. Note that we generally control–and recommend that analysts control--for sample recruitment procedure in multivariate analyses of the younger sample.

However, recruiting the younger generation to the study was difficult. They did not respond well to the initial mailing effort. An experiment using colorful postcards was then used and this served as a follow-up mailing to encourage more participation, but it was almost all from the older generation.

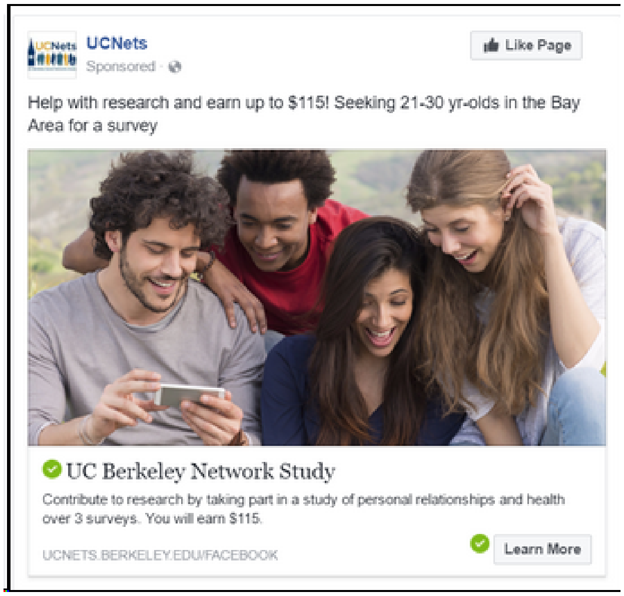
Another step we took to increase young cohort response was using referrals. Those who had completed the study – initially just 21-30 year olds, and then expanded to the 50-70 year olds – were asked to refer

---

[3] Initially the selection ratio for households of mixed ages was 2:1 for the young over the old, but as the study came closer to reaching the older cohort target, the ratio was changed more in favor of younger recruits.
[4] Means were 60 and 62 minutes respectively; medians were 58 for both. The web time does not include breaks respondents took.
[5] Discussion of variations appear in Fischer and Bayham, in process.

21-30 year olds they knew who were not immediate family members or members of their household to take part. We do not know how many of the initial respondents forwarded the referral email, but we do know that 78 people started the process, and 37 of those finished the main survey.

Facebook Recruitment. A colleague, Prof. Danny Schneider, told us that he had successfully used Facebook ads to recruit participants in a study on low-wage earners. One of the (then) graduate student project members, Eric Giannella, became our Facebook expert, and created several targeted 'ads' presented here, following this procedure: We bought advertisements from Facebook targeting 21-to-30 year-olds who spoke English or Spanish in the San Francisco-Oakland-San Jose region. We ended up advertising on Facebook for desktop exclusively as these were the people actually took the survey. We set the ad to run from 11 am to 2 pm and 8 pm to 11 pm, hoping that this would catch people both at work and after work. The screener was adjusted slightly to accommodate this new influx path.



Facebook reported that about 420,000 people were exposed to the ad and 2,120 clicked on it. As shown in the tables above, 786 began the screener and 290 completed the survey.

Of the 3,207 Facebook recruits, 878 began the enrollment process, and 433 completed the main survey. While we had to discard some of the Facebook respondents because of incomplete surveys, we felt successful in attracting a sample size sufficient for complex analyses. In our initial analyses, we have not seen that the Facebook sample skews substantially from the mail-invited respondents of the same age group who completed the survey. However, as seen in Table 2, below, Facebook respondents were much more likely than others to drop out before completion of the survey.

Facebook Ads Results:

- 414,059 saw the ad (impressions).
- 2,102 clicked the ad (taken to ucnets landing page)

Through the iterative adjustments that are part of the process, Giannella found the desktop news feed to have better take-up.

Facebook desktop news feed results:

- 127,307 saw the ad
- 1,105 clicked (taken to our landing page)

III. ENROLLMENT TAKE-UP RATE

3. 1 Rates of Recruitment and Progression through the Process

Our procedure entailed inviting a hopefully representative sample of two cohorts to join our panel. For the ABS sample, the path from receiving the letter to being in the final data set entailed numerous steps: Opening the letter; determining qualification and interest; contacting the fieldwork center; being positively screened as qualified for the study; being selected as *the* qualified member of the household for the study or being asked to pass the invitation on to another age-qualified member of the household; being randomly assigned to the in-person or to the online condition (and accepting that assignment); scheduling and making an in-person interview *or* opening the link to the web version; starting the survey; and completing the survey. Each step entailed drop-outs, surely non-random ones. Estimating and evaluating our final take-up or enrollment rate requires understanding the losses at each stage. (The referral and Facebook cases followed the same path once they contacted the fieldwork sample.)

3.2 Rates of Contacting the Field Office and Starting the Survey

Of the 2,447 individuals who contacted the UCNets field office by phone or through a website, 1,558 had responded to the letter (103 had been referred by a previous respondent and 786 had been solicited by Facebook or referral–all of these should have been 21 to 30). The steps that followed included confirming that someone in the household qualified; selecting a person in the household for the sample (in about 290 cases, it was someone other than the initial contact); reading the consent form; and randomly assigning the respondent to either the FTF or the web condition.

3.3 Assignment to mode

As part of a methodological experiment, the screening procedure randomly assigned qualifying respondents to either a face-to-face interview (75 percent of cases) or a web survey (25 percent)--except for the Facebook recruits, who all did the survey online. The in-person and online instruments are substantively identical and we control for mode of administration in all the multivariate analyses. After excluding incomplete and invalid cases, we attained completed surveys from 508 50-to-70 year-olds interviewed in-person and 166 on the web and from 139 21-to-30 year-olds interviewed in-person and 346 from the web.

In the ABS sample 789 were assigned to FTF and 265 assigned to web. (Some respondents balked at their assignments and a few percent were re-assigned.[6]) At the end of the screening process, interviews were set up or links to the online survey distributed.

There are two major points here about sample loss. About 30 percent of those who called or logged in from the ABS invitations dropped out either because of loss of interest or ineligibility. Then, there was a further loss of 12 percent from the eligible who never started the survey.

Table 1, below, shows the drop-offs at various stages in the process.[7]

---

[6] In the final ABS sample that completed the survey, 13 of those assigned to FTF–two percent-- ended up doing web and 14 of those assigned to web–six percent–ended up doing FTF, largely because of internet unavailability or discomfort.

[7] The numbers have some noise; note that more ABS respondents were assigned to a condition than were coded as eligible. But the general picture is clear.

Table 1: Progress of Recruited Respondents through the Screening Process			
	ABS	Ref	Facebook
Entered Screener (n=)	1558	103	786
Asked Consent	71%	77%	79%
Gave Consent	68%	77%	79%
Eligible	67%	76%	65%
Given Intended Mode	68%	76%	65%
Started Survey as % of Entered	57%	40%	55%
Started Survey as % of Eligible	85%	53%	85%
Started Survey (n=)	885	41	433

What sort of selection biases occurred along these steps?[8] We know little about the respondents who dropped out early in the process, but we can say this: The more persons age-qualified in the inquiring respondent's household, the less likely he or she would end up as eligible (either because someone else, by random selection among the age-qualified, ended up as the respondent or because the selected other person did not follow up). Persons in households with at least one 50-to-70-year-old were much likelier to end up in the eligible sample than those not. Similarly, screened 50-to-70-year-olds were likelier to end up as eligible–81%--than were 21-to-30-year-olds–69%. We also have the zip codes of the invited and eligible.[9] Respondents from San Francisco and Marin counties were more likely (75%) and residents of San Jose and of rural San Mateo county were less likely (61%, 62%) to end up as eligible.

We performed a similar analysis of the Facebook recruits. As the table above shows, 786 people responded to the ads. Of those, 100 were not in the correct age group (45 were coded as missing on age). Unlike the ABS sample, those from households with other 21-to-30-year-olds were neither more nor less likely to end up in the sample. In the end, the ratio of the Facebook recruits who *started* the survey to those who contacted the field office was comparable to the ABS sample. (We did not do a similar analysis of the referred recruits because of small numbers.)

3.4 Rates of Completing the Survey

Of the 1,359 respondents who began the survey; 1,159 completed, by the criterion of having answered questions in the final section. Their progress through the survey, by recruitment method, is show in the next table. Recall that all Facebook and referred recruits were 21 to 30 and all Facebook recruits did the survey on the web. There is thus a confound between recruitment and both age and mode. (An analysis of mode effects on survey completion appears in Fischer and Bayham, in process) The table below shows the rates of progression through the survey by recruitment, adding in a column for ABS recruits 21 to 30 to allow age-constant comparison.[10]

---

[8] This paragraph draws on tests of the differences among respondents to the screener survey between those ended up coded as "eligible" and those not.
[9] This analysis use the 3-digit zip codes to aggregate cases.
[10] We did not do that for the previous table, because we do not learn the age of the targeted respondent until after many have dropped away.

Table 2: Progress of Respondents who started the Survey				
	ABS(all)	ABS(21-30)	Ref	Facebook
Started Survey (n=)	885	170	41	433
Completed First Section	99%	98%	98%	94%
Completed Name-Eliciting	99%	98%	95%	90%
Completed All Network Sect's	97%	95%	88%	69%
Completed Survey	94%	85%	85%	67%
Completed Survey (n =)	834	160	35	290

The major takeaway is that, while younger respondents were likelier than older ones to not complete (compare first two columns), Facebook respondents were especially likely to drop out, notably during the name-descriptor part of the network section.

3.5 Estimating Rates of Enrollment

Because the referral and Facebook samples were non-probability, the following estimation applies only to the ABS sample. Assessing the yield is difficult, because—in addition to the facing generally high resistance to polling in the current era (National Research Council 2013)—we required recipients of the invitation to qualify by age, reach out to our fieldwork center, be randomly chosen from eligible members of the household, enroll in a panel for three waves, and in most cases arrange an in-person interview. We estimate that someone from about five percent of the eligible older households contacted our field office and that in the end about three percent of them fully completed the survey.[11] This panel uptake is comparable to contemporary, high-quality commercial surveys (e.g., Pew[12]).

Addressed-based sampling, as described above, sufficed to yield the anticipated 50-to-70 year-olds, but not the 21-to-30 year-olds, who are notoriously hard to reach. We may have been at about a 1.5 percent completion rate with that group, when we turned to the other methods. We tried, in Santa Clara County, to use colorful postcards to get more young people, but the postcards had the effect of encouraging a few older respondents to respond, and we saw no lift in response from younger respondents. We then expanded the outreach in two ways: We added a few dozen more respondents by getting referrals from existing panelists and added 290 more by soliciting 21-to-30-year-olds in the metropolitan area through Facebook advertisements. (See above.) These recruitment differences are another reason for analyzing the age cohorts separately. Note that we generally control–and

---

[11] We estimate that about 50,000 letters went out to valid addresses in the metropolitan area. Of those, census data suggest that about 22,500 would have gone to households with someone between the ages of 50 and 70. (Thanks to Daniel Schneider for the calculations.) Our field office received replies from about 1,000 of such households and in the end 674 older qualifying respondents completed the survey to the end either in-person or on the web, for a cumulative response rate of about three percent among the older population.

[12] On the Pew comparison: In 2014, Pew abandoned their RDD practice and built a panel for continuing surveys. Pew asked respondents to a 2014 telephone survey-itself with a response rate of about 10 percent-to join an ongoing web- or mail-based panel. Of those asked, 54 percent agreed, but only 43 percent participated in at least one of their subsequent surveys. Pew (2015) estimates their cumulative response rate as 3.5 percent.

recommend that analysts control--for sample recruitment procedure in multivariate analyses of the younger sample.

ABS Recruitment: Because only about 25% of the region's population falls into the age groups, and because of the effort required to participate, the enrollment rate was approximately 7.72%[13], which is low, yet note that the Pew American Trends Panel study, with a far less demanding study, garnered a 3.5% enrollment rate.[14] The enrollment rate for the mailing is therefore about 1.5% but taking into consideration incidence, the mailing generated a 7.72% enrollment rate.

Facebook and Referrals: It is not possible to calculate a response rate for the referral or Facebook samples as they are non-probability samples with unknown denominators.

3.6 Sample Composition by Mode and Age Group

Table 3 shows the resulting sample:

Table 3: Sample Composition by Age and Mode

		21-30 year olds			50-70 year olds		
	TOTAL	**All**	**Web**	**Face-to-Face**	**All**	**Web**	**Face-to-Face**
Male	468	212	163	49	256	52	204
Female	891	432	339	93	459	133	326
TOTAL	1,359	644	502	142	715	185	530

After deleting 200 respondents for incomplete data, the net sample is 1,159. However, the incomplete responses have useful data, and so it is left to the researcher to decide whether to use the complete sample. The public use files have only the 1,159.

Of these final 1,159, 346 were 21-30 year olds who completed the survey via the Web mode; 203 were female Facebook recruits, and 87 of the 21-30 year olds were male Facebook recruitments, for a total of 290 Facebook recruits.

Table 4: Final Panel Composition by Age and Mode

		21-30 year olds			50-70 year olds		
	TOTAL	**All**	**Web**	**Face-to-Face**	**All**	**Web**	**Face-to-Face**
Male	395	152	103	49	243	47	196
Female	764	333	243	90	431	119	312
TOTAL	1,159	485	346	139	674	166	508

[13] Internal note: If you use 25% of 60,000 mailings, which is about 15,000, and subtract the returned mail, about 5% of the total, we get 7.72%. (15,000-3,000 = 12,000. 1359-433=926. 926/12000 = 7.72.

[14] http://www.pewresearch.org/methodology/u-s-survey-research/american-trends-panel/.

IV. WEIGHTS

4.1 Post-Stratification Weighting

The weighting was done by Marco Angrisani, of USC's Center for Economic and Social Research (CESR).[15] He calculated two weights for each case, the first simply to match the conjoint sociodemographic distributions of six attributes in our sample to that of the six Bay Area counties (using the 2014 American Community Survey) for each age group, the second weight adding to that calculation adjustments for the sampling strata described earlier–missing for the 325 Facebook and referral cases in the younger sample. The first weight is based on the distributions –*within age groups*–of: gender (male/female); specific age (21-25 /26-30 for the younger; 50-60/61-70 for the older); race (white/Asian/other); ethnicity (Hispanic/non-Hispanic); marital status (married/not married); and education (less than B.A./exactly B.A./more than B.A.). This weight is labeled *weight_demo.* The weights sum to 1159.

These weights were highly skewed and could lead to a one case being weighted as much as 130 times that of another. Consulting with Dr. Angrisani, we explored "trimming" the weights by one percent or 5 percent. Both are available. In the end, we recommend using the 95 percent trims.[16] Though the resulting weighted sample is not as exactly representative, the trimmed weights better assure that outliers do not determine the results. The most extreme contrast in weights is a ratio of 19.4.

V. DATA STRUCTURE

There are two data sets for each wave of UCNets. In one file, the respondent is the unit of analysis, such that every record (line) is one case. As indicated in the codebook, some questions are not available for unrestricted public use as they contain personal identifiers. The public use data set contains therefore a selection of screener items and the majority of main survey items. For the face-to-face interviews, the file includes a number of items about the interviewer's impressions of the respondent, specifics for the face-to-face interview itself, and details about the interview setting. In Wave 1, there are 1,359 records, however, many of these are not of high quality, and after tagging those who did not complete the survey, remaining are 1,159 respondents.

The 'Names' file contains the names and attributes of named alters tabulated from the name-eliciting questions. Each record is one alter. Each alter has a unique identifier, which is prefixed by the ego's identifier (PRIM_KEY). For example, if the ego's PRIM_KEY is 300000000036, then the first name in the names file is 3000000003601.

The core emphasis of the interview is the name-eliciting portion, that is, respondents (egos) were asked to name people in their lives (alters), such as members of their household, people they do activities with, people who they help, and who people who help them. Each name also has data about them, such as the relationship to the respondent, how they know them, whether the named alter is of the same religion, sex, age group, etc. The names data is contained in the second data file. In this file there are over 14,000 records, one for each alter named by the egos. On average, there were x names per ego. To each alter is the corresponding data, and also the ego's data. Thus, one record is 'Nancy' who is the first

---

[15] Project document: "Berkeley Study: Weighting Procedure" (available upon request). One technical note: The document reports weight 1,157 cases. Three late ones were added by the PI with appropriate weights constructed by using a regression model to "predict" predict their weights.

[16] Original values of *weight_demo* above the 95th percentile are recoded to the 95th percentile value and those values originally below the 5th percentile are recoded to the 5th percentile value.

person named by the ego (identified by the respondent's unique identifier, prim_key), and 'Nancy' also is assigned the ego's identifier, with a suffix of _1 to denote that she is the first name.

5.1 Curating the Data

As mentioned above, the initial 1,359 respondents were reduced to 1,159 in the final set for analysis by requiring that all respondents in that set have completed the survey. (Users who wish to use the larger set should inquire to the project. Those cases lack sociodemographic information such as education, income, and ethnicity, as well as some of the network data and the health data.)

In addition, we carefully combed through the lists of alters whom the respondents provided and are collected in what is described below as the "Names" file. Entries that were determined be inappropriate – e.g., "my friends," "the homeless," "golfing" – were dropped. And entries that were determined to be duplicates – e.g., a "Bill" and a "Billy," or a "Robert" and a "Dad" both coded as parents – were *merged*. We estimate that fewer than three percent of all the alters originally listed in the complete cases were in this sense errors and required correction. The correction process is inherently imperfect and always in process. We plan to provide revisions of the files as we find errors.

5.2 Ego-Level versus Alter-Level Analysis

The data were delivered to us in a wide format, with all the variables describing the named alters in a row with the respondent data. We converted that into two "long" files. One file is the respondent data (with a small handful of variables that summarize network data[17]), n = 1,159. The second file is called the "Names file." The unit of analysis is the alter, n = apx. 12,000, which codes which questions the particular alter was named on and all the descriptive variables for that alter. The files are connected by the variable prim-key which identifies the respondent on both files. Users can attach respondent data (e.g., gender, marital status) to the Names file and, conversely, aggregate Names data (e.g., number of kin named, percentage of social partners who are same gender) to the respondent level.

One can use the data at the ego level or the alter level, using each each data set in a stand-alone analysis. If one wishes to utilize the names for a ego-centric analysis, then one needs to create new variables that provide counts, or specify the nature of the relationship, between the ego and the alter. One may wish to analyze the data by focusing on the egos, and how their personal networks manifest. For example, how many people – or what kinds of people – are there in the ego's network? To do that, one first creates a new variable that aggregates all the responses for each name in the desired category for the ego into a single variable from the names file that is specific to the ego, and then outputting the new variable to a separate file, and then merging this new variable to the main survey data.

For example, you wish to see how many people in the ego's network that were met at work.

Using SPSS syntax, first compute the co-workers variable in the Names file.

```
Compute coworkers = 0.

IF c1a_12 EQ 7 COWORKERS = 1.
```

Then aggregate them into a new variable:

```
COMMENT aggnames is the new data set with the new variable.

DATASET ACTIVATE DataSet1.

AGGREGATE
```

[17] Specifically, it includes the number of names provided in answer to each name-eliciting question.

```
 /OUTFILE='C:\data\aggnames.sav'
 /BREAK=prim_key
 /name_coworkers_sum=SUM(coworkers).
```

Then merge this new file with the main file (first, sort both files by prim_key).

In SPSS for a full description of AGGREGATE Sub-commands, see

https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/base/syn_aggregate_functions.html.

STATA

Similar to the SPSS code above, below is a short bit of Stata code that creates a "number of coworkers" variable from the names file and merges it to the respondent file. The file 'names_file.dta' is the file name we use for this example.

```
cd <<working directory>>
use "names_file.dta", clear
gen coworker = c1a_12
recode coworker 9=.
collapse (sum) coworker, by(prim_key)
save <<file name>>, replace

use "names_file.dta", clear
merge 1:1 prim_key using <<file name>>
tab coworker
```