Estimating the Size and Age-gender Distribution of Women’s Active Social Networks

Maryam Zamanian MSc1, Mohammad Reza Baneshi PhD2, Ali Akbar Haghoost PhD3, Tayebeh Mokhtari-Sorkhani4, Fatemeh Amiri MD5, Farzaneh Zolala PhD6

Abstract

Background: Network Scale-up (NSU) method is an indirect method for the estimation of hidden behaviors. In NSU, respondents are asked about the number of members they know from a subpopulation of interest (e.g., injecting drug user) and assume that the prevalence of risky behavior in the networks of a random sample of respondents is similar to that of the population. However, first, we need to identify the total number of people each respondent knows [the social network size (C)]; Moreover, certain risky behaviors happen in particular age and gender groups. Our aim was to determine the size and age-gender distribution of female networks.

Methods: This cross-sectional study was conducted in the city of Kerman, Iran. A total sample of 1275 women was recruited using multistage sampling. In this study, 25 first names were selected as reference groups. Participants were asked how many people they know with the selected names. The respondent's answers were categorized into eight separate age-gender subgroups and C was estimated for each subgroup.

Findings: The results of this study showed that, on average, each Kermanian woman knows about 234 people and about two-thirds of them are female (82 males and 152 females); moreover, participants were more likely to communicate with their peers. The majority of males (88%) known by Kermanian women were in young and middle age groups; in contrast the female young and middle age groups, who are at reproductive age, form only 45% of the female part of their networks.

Conclusion: We have seen that the age-gender distribution of the networks is not the same as that of the general population. Our figures can be applied in NSU studies focusing on risky behaviors of particular age and gender groups.

Keywords: Network Scale-up; Network size; Age-gender distribution; Women; Iran


Received: 04.02.2016 Accepted: 25.04.2016
**Introduction**

Estimating the size of hidden/hard to reach subpopulations (such as injecting drug users, female sex workers) is a highly challenging issue in the public health context. These groups are hidden in the society because of fear of legal punishment and the stigma surrounding these behaviors. The use of standard methods of sampling and direct estimations requires huge sample sizes due to the relatively low prevalence of such behaviors in the general population. Another problem of direct methods for size estimation of these groups is that respondents may refuse to answer or may provide wrong answers because of the stigma surrounding these behaviors and the tendency to maintain privacy. This leads to underestimation of the results. To convince policymakers and investors in the health-care sector of the existence and importance of any public health problem, you need to have a valid estimate of the size of the population of interest. The use of indirect methods that overcome the limitations of direct methods can be helpful. One of the indirect methods for size estimation is the Network Scale-up (NSU) method, which is relatively new in the field of epidemiology and public health and does not require direct contact with the population of interest.

In the NSU method, respondents are asked about the number of members they know from the subpopulation of interest (e.g., injecting drug user). The NSU method is based on the concept that the proportion of individuals known by participants is linearly proportional to their real size in society at large.

In addition to the indirect and anonymous nature of questioning which desensitizes the topic and increases response rates and accuracy, this new method is efficient with respect to time and expense considerations. It could, therefore, be appealing to employ this method in middle- and low-income countries. Moreover, it is possible to estimate the size of several target subpopulations in one NSU study, reducing the need of conducting distinct studies for each subpopulation.

To employ this method, we need to first identify the total number of people each respondent is acquainted with the social network size (C). To estimate C, the respondents are asked about the number of people they know in certain groups whose real sizes are known (e.g., “How many people do you know named Afshin?”).

C is dependent on the social and cultural structure of each community. Quite varying estimated values for C in different parts of the world (from 55 to 611) are evidence for this claim. Hence, when investigating the hidden subpopulations in a region, a study needs to estimate the specific local C.

In Iran, a national survey estimated an average size of about 300 people for C for the whole population. This means that every Iranian knows around 300 people, on average. This estimated size of C is an average for the entire country and could be very different for each city. It is also not clear how many of them are women, young people, etc. In other words, calculated C in the previous studies represents the average total number of people that everyone knows. However, many high-risk behaviors occur in a particular gender or age group. For example, abortion only can happen among females of reproductive age. It is not enough to know that each Iranian knows about 300 individuals; we should know the gender and age structure of networks to identify the average number of women of reproductive age that everyone knows. The easiest way is to assume that the age-gender distribution of the network is the same as the age-gender distribution of the society. However, we believe that this is not always the case, and individuals could have more contact with their peers.

This study aims to estimate the size and the age-gender distribution of women’s social networks. Our results provide basic information, which is essential for studies focusing on risky behaviors of particular age and gender groups.

**Methods**

This cross-sectional study was conducted in the city of Kerman (the capital of the largest province of Iran) and aimed at establishing the size and age-gender distribution of the social network size of Kermanian women. According to the latest Iranian Census (2011), the city of Kerman has a population of 534,441 and females make up 50% of this value. In this regard, 1275 women, proportionate to the age distribution of the latest census, were interviewed. Multi-stage sampling
was used in this study. In the first step, we used the governor’s office experts’ view to classify the city of Kerman with respect to the socioeconomic classes. Accordingly, the city was classified into three categories: high, medium, and low socioeconomic status. Then, five regions were randomly selected from each category as clusters. Finally, in crowded streets of each region, 85 pedestrians were recruited in accordance with the age distribution of women using the convenience sampling approach. Interviews were carried out both in the morning and evening with eligible women aged over 18 residing in the city of Kerman for the past 5 years. A trained same-sex interviewer explained the purpose of the study and collected the required data with a structured, face-to-face interview, once the interviewees verbally consented to participate in the study. The study protocol was approved by the Ethical Committee of Kerman University of Medical Sciences (ir.kmu.rec.1394.223).

The known population method is typically used to estimate social network size and it is recommended to include 20-30 reference groups with known population sizes. In this study, 25 first names, including 13 female and 12 male first names were selected as reference groups (due to the ease with which participants can identify these groups, compared to other reference groups). The selection of names was done in consultation with the civil registration offices and based on the guidelines provided in the literature as follows:

- Proportion in general population being from 0.1% to 4%,
- Popularity not changing in recent decades,
- No two-part or gender-neutral/ambiguous names.

In the first section of the interview form, participants were asked about the number of people they know with the selected names. We also asked them to classify their replies based on age group (< 18, 18-29, 30-49 and > 50). The participants were explained the standard definition of “know” as “people whom you know and who know you by name, with whom you can interact, if needed, and with whom you have had contact within the last 2 years personally, or by telephone or e-mail.” The other section of interview form was related to the demographic and socioeconomic status of participants.

The basic equation in the NSU method is \( e/t = m/C \). In this equation, \( e \) is the size of a certain reference group and \( t \) is the size of the total population, both of which should be obtained from other sources. \( m \) is the number of individuals that each respondent knows from the certain reference group, and \( C \) is the respondent’s network size. For example, assume we have two thousand females named Sara in a city with population of 500000. If one participant knows two females named Sara, then this participant’s \( C \) will be 500. This equation is extended to estimate \( C \) with more precision by averaging multiple reference groups and many participants’ responses. The \( C \) and its standard error (SE) were calculated with the formula 1 and 2, respectively:

\[
c_i = \frac{(t^i \sum m_i)}{\sum e_i} \quad (1)
\]

\[
SE = \sqrt{\frac{(t^i C_i)}{\sum e_i}} \quad (2)
\]

The indices \( j \) and \( i \) stand for the reference group and respondent, respectively. Ineligible reference groups were eliminated in an iterative back calculation process. For this purpose, using the \( C \) value derived, we back calculated the size of reference groups:

\[
\hat{c} = \frac{(t^i \sum m_i)}{\sum \hat{c}_i} \quad (3)
\]

The ratio of back-calculated to real-size was then calculated for all reference groups, and the reference group with the worst ratio (farthest away from one) was considered ineligible and eliminated. \( C \) was re-estimated using the remaining reference groups, and this process was repeated until all the ratios were between 0.5 and 2. We also used the approach recommended by Habecker in elimination of ineligible reference groups. Habecker et al. used log base 2 of the ratios and continued the process until all absolute values were below one. This rule is the same as having ratios between 0.5 and 2 but guarantees symmetric values.

Each respondent’s answers were categorized into eight separate age-gender subgroups, and the \( C \) calculations were performed separately in each of these age-gender subgroups.

To assess the internal validity, intraclass correlation (ICC) between the back-calculated and real sizes of the eligible reference groups and root mean square of error (RMSE) were calculated. RMSE was the root of the mean of square differences between the real and back-calculated...
sizes of the eligible reference groups. We also fitted a regression line so that the dependent and independent variables were back-calculated and real sizes of the reference groups, respectively. Intercept of zero and slope of 1 was considered evidence of agreement between back-calculated and real sizes. Microsoft Office Excel and SPSS (version 20, SPSS Inc., Chicago, IL, USA) were used to analyze the data.

## Results

Out of 1946 female pedestrians, 1275 agreed to participate in our study. The youngest and oldest participants were 18 and 83-year-old, respectively, and the mean ± standard deviation (SD) age of participants was 36.46 ± 13.96 years old. Approximately, 28% of participants were single. Among women participating in the study, 43.2% had higher education (more than 12 years) and 27.7% were employed. Among the husbands of married women, 32.7% had higher education and among their different job categories, self-employed was the most frequent (51.3%) (Table 1).

In the back-calculation process, the results of log base 2 of the ratios were the same as the results of counting only ratios between 0.5 and 2.

The results of this study showed that on average, each Kermanian woman aged over 18 knows around 234 persons. Two-thirds of the social network of women was formed by women. In other words, Kermanian women know 82 males and 152 females. The majority of males known by Kermanian women were in the young and middle age groups (72 out of 82, corresponding to 88% including 40 in 18-29 and 32 in 30-49 age groups, respectively). In contrast, Kermanian women knew females in all age groups. Those in the young and middle age groups, who are of reproductive age, form only 45% of the female part of their network (105 out of 152) (Table 2).

### Table 1. Demographic characteristics of participants

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td></td>
<td>231 (23.6)</td>
</tr>
<tr>
<td>25-34</td>
<td></td>
<td>391 (30.7)</td>
</tr>
<tr>
<td>35-49</td>
<td></td>
<td>328 (25.7)</td>
</tr>
<tr>
<td>≥ 50</td>
<td></td>
<td>255 (20.0)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td></td>
<td>358 (28.1)</td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td>829 (65.0)</td>
</tr>
<tr>
<td>Divorced/widowed</td>
<td></td>
<td>88 (6.9)</td>
</tr>
<tr>
<td>Job</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housewife</td>
<td></td>
<td>632 (49.6)</td>
</tr>
<tr>
<td>Employee</td>
<td></td>
<td>181 (14.1)</td>
</tr>
<tr>
<td>Student</td>
<td></td>
<td>195 (15.3)</td>
</tr>
<tr>
<td>Self-employed</td>
<td></td>
<td>172 (13.5)</td>
</tr>
<tr>
<td>Retired</td>
<td></td>
<td>59 (4.6)</td>
</tr>
<tr>
<td>Unemployed</td>
<td></td>
<td>36 (2.8)</td>
</tr>
<tr>
<td>Education (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 9</td>
<td></td>
<td>267 (20.9)</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>457 (35.8)</td>
</tr>
<tr>
<td>12-16</td>
<td></td>
<td>455 (35.7)</td>
</tr>
<tr>
<td>≥ 18</td>
<td></td>
<td>96 (7.5)</td>
</tr>
<tr>
<td>Husband’s job of married</td>
<td></td>
<td></td>
</tr>
<tr>
<td>women</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employee</td>
<td></td>
<td>215 (25.9)</td>
</tr>
<tr>
<td>Worker</td>
<td></td>
<td>61 (7.4)</td>
</tr>
<tr>
<td>Self-employed</td>
<td></td>
<td>425 (51.2)</td>
</tr>
<tr>
<td>Retired</td>
<td></td>
<td>114 (13.7)</td>
</tr>
<tr>
<td>Un-employed</td>
<td></td>
<td>14 (1.7)</td>
</tr>
<tr>
<td>Husband’s education of 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>years</td>
<td></td>
<td>238 (28.7)</td>
</tr>
<tr>
<td>12 years</td>
<td></td>
<td>320 (38.6)</td>
</tr>
<tr>
<td>12-16 years</td>
<td></td>
<td>202 (24.4)</td>
</tr>
<tr>
<td>≥ 18 years</td>
<td></td>
<td>69 (8.3)</td>
</tr>
</tbody>
</table>

### Table 2. The size and performance of C values, divided by different age-gender subgroups

<table>
<thead>
<tr>
<th>Age group (year)</th>
<th>The number of eligible reference groups</th>
<th>C</th>
<th>RMSE†</th>
<th>ICC‡</th>
<th>Slope¶ (P†)</th>
<th>Intercept¶¶ (P§)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male part of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>women’s network</td>
<td>&lt; 18</td>
<td>8</td>
<td>6</td>
<td>241</td>
<td>0.77</td>
<td>0.62 (0.309)</td>
</tr>
<tr>
<td></td>
<td>18-29</td>
<td>12</td>
<td>40</td>
<td>127</td>
<td>0.95</td>
<td>0.74 (0.035)</td>
</tr>
<tr>
<td></td>
<td>30-49</td>
<td>8</td>
<td>32</td>
<td>122</td>
<td>0.94</td>
<td>0.93 (0.764)</td>
</tr>
<tr>
<td></td>
<td>≥ 50</td>
<td>4</td>
<td>4</td>
<td>70</td>
<td>0.97</td>
<td>1.00 (0.997)</td>
</tr>
<tr>
<td>Female part of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>women’s network</td>
<td>&lt; 18</td>
<td>6</td>
<td>15</td>
<td>188</td>
<td>0.83</td>
<td>0.88 (0.797)</td>
</tr>
<tr>
<td></td>
<td>18-29</td>
<td>8</td>
<td>59</td>
<td>142</td>
<td>0.86</td>
<td>0.48 (0.008)</td>
</tr>
<tr>
<td></td>
<td>30-49</td>
<td>9</td>
<td>46</td>
<td>124</td>
<td>0.93</td>
<td>0.85 (0.489)</td>
</tr>
<tr>
<td></td>
<td>≥ 50</td>
<td>5</td>
<td>32</td>
<td>162</td>
<td>0.78</td>
<td>0.76 (0.702)</td>
</tr>
</tbody>
</table>

†RMSE (Root Mean Square of Error): The root of the mean of square differences between real and back-calculated size of the eligible reference group. ‡ICC (Intraclass correlation coefficient) between back-calculated and real size of reference groups. ¶Slope of regression line of back-calculated versus real sizes. ¶¶Intercept of regression line of back-calculated versus real sizes. †Two-tailed P-value produced from t test compared to one. ††Two-tailed P-value produced from t test compared to zero.
Low RMSE and high ICC between the back-calculated and real sizes of reference groups indicate a high agreement between the real and back-calculated size. Most of the slopes of regression lines were close to one and their intercepts were reasonably close to zero; almost all corresponding P values were above 0.05, which indicates an appropriate agreement between the real and back-calculated sizes. The poorest C was calculated for men aged < 18. Here, the ICC was as low as 0.77; the slope of the regression line was 0.62, which still was not significantly different from 1 (Table 2).

In a complementary analysis, we calculated the network size of Kermanian women stratified by their age. We observed that young women were less likely to communicate with old women and vice versa (Figure 1).

**Discussion**

In this study, the average social network size for Kermanian women was calculated to be around 234 persons and the share of females (152) was much more than that of males (82), which is indicative of much more same-gender than opposite-gender communication; this issue is also clearly seen in all the age groups (Figure 1). This network size was also calculated divided by eight different age-gender subgroups. Internal validity of results was checked by calculating different criteria such as RMSE, ICC between the back-calculated and real size of reference groups and regression after back calculation of the size of reference groups.

Various other studies have been conducted in other countries in order to estimate C, and quite different values have been obtained. Various sizes have been estimated for C in the USA, including 55, 108 and 291.\(^1\)\(^8\),\(^2\)\(^0\) C has been estimated at 175 in Ukraine\(^3\) and 364 and 310 in Japan and China, respectively.\(^2\)\(^2\) The differences observed in estimated C in different countries can have two possible causes: methodological and real.

The methodological causes may be due to the lack of uniformity in the method of estimating C, or the number and nature of reference groups in different studies. For example, the Kadushin et al. study\(^2\)\(^1\) in the USA used only six reference groups, which possibly caused the large SD of C. Furthermore, their reference groups were crime-related, which could cause underestimation of C due to the potential stigmatization. In two other studies in the USA, McCarty et al.\(^2\)\(^0\) used two methods to estimate C; the NSU method with 29 subpopulations and the summation method with 16 relation types; while reference groups in
another study\textsuperscript{18} included 26 subpopulations plus 14 first names. In Japan, Ezoe et al.\textsuperscript{3} considered 7 reference groups among 10 primary reference groups as inappropriate in a primary pilot study, and the main study was conducted with only three remaining reference groups including firefighters, police officers, and military personnel; and Guo et al.\textsuperscript{22} in China, used 8 reference groups for estimating C.

As other methodological aspects, for example in Kadushin et al. study\textsuperscript{23} very far from one (from 0.1 to 3) values can be seen among the ratios between the predicted and the real rates for different cities and different reference groups, that could affect the size and reliability of network size estimation; Or in Guo et al. study\textsuperscript{22} the network size has been estimated after three steps of adjustment for demographic distribution, back estimation, and log transformation.

In the Iranian national study mentioned above,\textsuperscript{14} the average size of the social network for all Iranian persons (regardless of gender) was estimated between 308 and 380 and the corresponding figure for women was estimated between 303 and 341; this is more than the estimated size in the present study. While the national size is an overall average for the whole country, it is not suitable for use in specific locations, due to regional cultural and social differences, which result in different network sizes for various locations. The need to perform independent studies with adequate sample sizes in different provinces is also emphasized in the national study, due to lack of generalizability of the national study results.\textsuperscript{14} Moreover, age-gender distribution has not been determined in either the national study or in other studies, while it is required in NSU studies. Whenever in NSU studies a corresponding distribution has been required, it has been assumed that the age-gender distribution of the network is the same as the age-gender distribution of the society, which could contribute to considerable errors. For example, in the national study, in order to estimate the abortion rate using the NSU method,\textsuperscript{21} the share of women of reproductive age from total C was needed; by assuming the society distribution, they assumed that females composed half of women’s network size, whereas the present study showed that females composed about two-thirds of women’s networks. Generally more contact with the same gender is not unexpected, moreover in Iran it could also be affected by Islam. In order to determine the share of the intended age group, studies were performed using the proportion of the general society age groups; whereas the present study showed that these sizes are not proportionate to the age-gender distribution of society and with increasing age, the social network gets older. For example, younger women have much younger networks than older women (Figure 1), which could be due to more contact with peers. This figure shows that the study participants are far more communicative with same-age cohorts.

Another study conducted in Iran has estimated the total network size of 259 for Tehranian people and 230 for Tehranian women,\textsuperscript{16} which is nearly consistent with the results of the present study. The slightly lower estimated C was expected for Tehran, Iran, due to the special lifestyle in capital cities, less communication and cognition in big cities than in smaller cities, and/or that many people in Tehran are immigrants from other cities who have transplanted to Tehran for employment or other reasons and so could have smaller networks in Tehran.

Shokoohi et al.\textsuperscript{24} in another study in Iran estimated the Kermanian men’s network size of 303, which is greater than our estimation for Kermanian women. Other studies have also estimated larger networks for men than women.\textsuperscript{14,16,22} This could be because of more business contacts and social activity among men, which leads to a larger social network.
We could not find similar studies in other places or involving the opposite gender in order to compare our results about network distribution. While determining age-gender distribution of social networks is essential for assessing hidden behaviors through the useful method of NSU, we recommend performing similar research in other regions and on males, which could show new configurations of the size and distribution of C, and will enable comparison of findings.

Limitations
An inherent limitation of this method is its dependence on the short amount of time given the participants to remember information, which could affect the accuracy of answers.

Another limiting factor is cognition error, which could occur in this method. Despite all the interviewers’ efforts in explaining the concept of recognition to participants, participants may interpret it differently.

Despite the existing limitations, this is the first study to specifically assess the size of women’s social networks and establish the networks’ age-gender distribution.

Conclusion
Social network size in different parts of the world, even within a single country, can be quite different due to different levels of communication that is related to the cultural and social structures of each region. In addition, the age-gender distribution of the network is not necessarily the same as the age-gender distribution of the general population. This study showed much more same age-gender than opposite age-gender communication as well. Our figures can be applied in NSU studies focusing on risky behaviors of particular age and gender groups.

Conflict of Interests
The Authors have no conflict of interest.

Acknowledgements
We would like to thank the Kerman University of Medical Sciences that financially supported the project. We would also like to thank all those who participated in the study.

References
Active Social Network Size & Age-gender Distribution


http://ahj.kmu.ac.ir,  5 July
برآورد اندازه و توزیع سنی- جنسی شبکه اجتماعی فعال زنان

مريم زمانیان، 1 دکتر محمدرضا بانی‌شي 2، دکتر علی اکبر حقوقدست 3، طبه مختاری سرخانی 4،
دکتر فاطمه امیری 5، دکتر فرزانه دولالی 6

مقاله پژوهشی

چکیده
مقدمه: روش تعمیم شیکه‌ای، روش غیر مستقیمی به منظور برآورد رفتارهای پنهان می‌باشد. در این روش، شرکت‌کنندگان افرادی از جمعیت هدف (برای مثال توریست) که نگهدارنده بررسی می‌شود را به عنوان شرکت‌کننده انتخاب می‌کنند. شرکت‌کنندگان در مرحله مطالعه بانک اطلاعاتی ملی به کار می‌روند.

روش‌ها: این مطالعه به صورت مقیاسی در شهر کرمان انجام شد. ۱۲۵ زن با روش نمونه‌گیری چند مرحله‌ای مورد بررسی قرار گرفتند. در این مطالعه، ۲۵ نام کوچک به عنوان گروه‌های انتخاب انتخاب شد. شرکت‌کنندگان در مرحله انتخاب فردی که با این اساسی می‌شناختند مورد سوال قرار گرفتند. پس از آن، شرکت‌کنندگان را به کار برگردانیدند. 

یافته‌ها: بافت‌های شبکه‌ای همراه با افزایش سن و جنسیت می‌تواند در بالای ۸ سال، ۳۳ و در بالای ۱۲ سال، ۳۴ و در بالای ۱۵ سال، ۳۵ تفاوت معنی‌داری داشته باشد. بافت‌های شبکه‌ای معلول می‌تواند تأثیر بگذارد.

نتیجه‌گیری: نتایج حاکی از آن بود که توزیع سنی- جنسی شبکه، به طور قطعی شبیه توزیع سنی- جنسی جمعیت عمومی نخواهد بود. نتایج مطالعه حاضر می‌تواند در مطالعات تعمیم شیکه‌ای با توجه به فناوری برقراری رفتارهای پرخاطر گروه‌های سنی- جنسی خالص به کار گرفته شود.

واژگان کلیدی: روش تعمیم شیکه‌ای، اندازه شبکه، توزیع سنی- جنسی، زنان، ایران

ارجاع:
زمانیان مريم، بانی‌شي محمدرضا، حقوقدست علی اکبر، مختاری سرخانی طبه، دکتر علی اکبر، فرزانه دولالی، براورد اندازه و توزیع سنی- جنسی شبکه اجتماعی فعال زنان. مجله اجتناب و سلامت 1395 (3)؛ 167-178.

تأريخ دریافت: 94/11/15

Email: zolalafarzaneh@gmail.com

Addict Health, Summer 2016; Vol 8, No 3

http://ahj.kmu.ac.ir, 5 July