



## Social Network Size Estimation and Determinants in Tehran Province Residents

*Mohsen SHATI<sup>1</sup>, \*AliAkbar HAGHDOOST<sup>2</sup>, Reza MAJZADEH<sup>1</sup>, Kazem MOHAMMAD<sup>1</sup>, SeyedSalehe MORTAZAVI<sup>3</sup>*

1. Dept. of Epidemiology and Biostatistics, School of Public Health, Tebran University of Medical Sciences, Tebran, Iran
2. Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
3. Dept. of Health Promotion and Education, School of Public Health, Tebran University of Medical Sciences, Tebran, Iran

\*Corresponding Author: Tel: +98 341-3205055; Email: Ahaghdoost@kmu.ac.ir

(Received 19 Apr 2014; accepted 13 Jul 2014)

### Abstract

**Background:** Network scale-up is an indirect method for estimating the size of hidden, hard-to-count or high risk populations. Social network size estimation is the first step in this method. The present study was conducted with the purpose of estimating the social network size of the Tehran Province residents and its determinants.

**Methods:** Maximum Likelihood Estimation was applied to estimate people's network sizes by using populations of known sizes and the scale-up method. Respondents were selected from Tehran province through convenience sampling in 2012. Out of thirteen selected subpopulations with known size, ten had minimum accuracy which used in our analysis.

**Results:** Of the 1029 respondents in this study, 46.7% were male. The social network size of Tehran Province residents was estimated to be 259.1 (CI<sub>95%</sub>: 242.2, 276) based on the ten known populations remained in this study. This size was 291.8 in men and 230.4 in women. Younger people (18-25 years old) had larger network sizes compared to the other age groups ( $P < 0.001$ ).

**Conclusion:** Our estimation for social network size of Tehran inhabitants was smaller than that previously estimated size for the whole country ( $c=380$ ). In addition, we found that the social network of subpopulations was different. This difference means that we need local estimations for sub-populations to improve the accuracy of population size estimation using network scale up method.

**Keywords:** Network scale up, Population size, Acquaintance, Social network size, Hidden population

### Introduction

Planning, evaluating, and budget allocation for prevention programs are contingent upon estimating the size of their target groups. Population size estimation of the target groups in some context, such as HIV/AIDS, always entail numerous challenges in countries where such diseases are concentrated in specific sub-populations with stigmatized behaviors & characteristics (e.g.: injecting drug users) (1-5). Traditional sampling methods that are used to estimate the size of these high risk

populations, including multiplier, capture-recapture, etc., are very complicated, if not impossible, since locating people with high risk behaviors is difficult, and approaching them directly can greatly reduce response reliability (5-10).

Estimating social network size of a representative sample, through a general population survey, and asking participants questions about specific high-risk behaviors among their acquaintances (11-13) is one of the best means of gathering information

on the sizes of such populations, also called hidden or hard-to-count populations due to the fact that it is not possible to calculate their sizes through regular, direct methods (14). One such method that has held the interest of researchers for the past decade is network scale-up (NSU) (6). In this method, the average social network size of the respondents (shown by  $c$ ) is used as a prerequisite for estimating the size of a high risk population in the society (shown by  $e$ ) based on the mean number of persons with high risk behavior known by respondents (shown by  $m$ ) (15).

In the past two decades, several studies in different countries including the United States, Ukraine, Brazil, Moldova, Rwanda, Japan, China, and Thailand have used this method to determine social network sizes as a pre-requisite for estimating high risk population sizes. Results have a wide range from 55 to 399 people. This difference points to the need for local studies to determine this value and its determinants (14-21).

In Iran, the estimated social network size on a national scale is between 308 and 380. However, Tehran, as a megacity and the capital of Iran, has unique demographic, cultural, and social features, special inter-individual relationship patterns, and different subpopulation proportions. In addition, due to the migration of different ethnic groups to this city, various ethnic networks have been formed (22).

Therefore to observe these different characteristics and the technical considerations which are recommended in the NSU method, the present study was designed to examine the social network size of the Tehran Province residents and the demographic factors affecting it in order to prepare the grounds for estimating the size of hidden populations.

## Materials and Methods

The social network size was estimated using an indirect approach, that is, respondents were asked how many persons they knew in a certain subpopulation within the society, and estimations were made based on the respondents' answers.

The term "knowing a person" has a specific definition in this study, and involves a certain time span and space (23, 24); therefore by saying A knows B, it is understood that A knows B by name and face; A can visit, call or email B whenever he/she wants, and vice versa. Moreover, A has contacted B by phone, email or in person at least once within the past 2 years, and B is a resident of Tehran province (25-27).

For the total sample size 1029, we selected 829 persons from Tehran and 200 from Robat Karim, the capital of Robat Karim County in Tehran Province in 2012. The city of Tehran was divided into 5 geographical zones: north, south, east, west, and center and the sample size were divided proportional to the population size of these five zones. Sampling was done in crowded areas, such as streets and parks; individuals were selected through convenience sampling, and questions were asked in the form of street interviews. All participants were 18 or older, and had been residents of Tehran Province for the past five years.

Upon entering the survey, participants were briefed on the study objectives, and their informed consent was obtained. In order to determine the social network size, participants were asked questions regarding the number of people in their personal networks using the indirect method, that is, they were asked how many people they each knew in specific subpopulations.

After coordination with the Ministry of Health and the Statistical Center of Iran, twenty-three known populations with clear and available provincial level statistics were selected to be used in this study. In order to improve estimation accuracy, and in view of the findings of similar studies (20, 28), several considerations were taken into account in preparing the final list of the known populations to be used in the present study; for instance, the size of all selected populations was between 0.1% and 4% of the total population of Tehran Province. Moreover, we excluded populations that could potentially induce a transmission error in the estimations for which there was no practical correction method. Popularity of names in the past three decades was observed in assigning names, and more popular names were used at

an almost even distribution. Out of the twenty-three initial subpopulations, thirteen were eventually used in the study: first graders, high school graduates, university applicants, university students, married people, divorced people, those who had had normal vaginal deliveries (NVD), those who had had Cesarean sections (C/S), primary school staff, people with the first names Hamed, Abulfazl, Sara and Marjan.

In this study, social network size was estimated using the maximum likelihood estimation (MLE) method and the equation below:

$$\hat{c}_i = \frac{\sum_j m_{ij}}{\sum_j e_j} \cdot T$$

Where  $\hat{c}_i$  is the social network size of the respondent  $i$ ,  $m_{ij}$  is the number of people that  $i$  knows in the known population  $j$ ,  $e_j$  is the real size of the known population  $j$ , the statistical information for which is available through previous censuses or surveys, and  $T$  is the total size of the general population in the survey area (11, 13).

To compute the 95% confidence intervals of the social network size, standard error  $C$  was calculated through the formula below: (29)

$$SE c_i = \sqrt{\frac{tc_i}{\sum_j e_j}}$$

This simple model will only work if the following strong assumptions exist (11, 24):

- A) All respondents have equal opportunity to know each member of the subpopulations under study.
- B) All respondents are fully informed about their own social network.
- C) All respondents can recall the number of their acquaintances in the subpopulations under study quickly and clearly.

Violation of each of these assumptions brings about errors in the results achieved through the network scale-up method. One thing that can violate assumption A is the barrier effect, which pertains to barriers causing some respondents to know the subpopulations in the study with stronger or weaker possibility.

Transmission error, on the other hand, can violate assumption B, and that is when all people in a given person's network are similarly likely not to be

aware of him/her belonging in a subpopulation.

Estimation effect is another factor that can violate assumption C, and it is due to a lack of precise knowledge of certain details about the people in a high risk social network.

In order to control various effects and errors that can bias estimations, different correction methods are used in different studies. In the present study, the following methods were employed to investigate the sources for these potential errors: 1) back estimation of each of the known populations to determine the best groups for estimating the ultimate social network size, and 2) variable stratification of the study samples, carrying out all analyses independently in each subgroup, and comparing results with that for the total sample. Moreover, acquaintances who were among the respondents' social network but were not residents of Tehran Province were eliminated from all estimations.

In order to perform the first correction, a preliminary calculation of the social network size was done, the size of each known population was assumed unknown ( $e$ ), and then the size of each subpopulation was estimated through the formula below:

$$\hat{e}_0 = t \frac{\sum_i m_{i0}}{\sum_i c_i}$$

At this point, the estimate/real ratio ( $e/r$  ratio) for each population was calculated by dividing the estimated population sizes by the real sizes throughout the province. The next step was to eliminate the first known population in which this ratio was not between 0.5 and 2. After removing the population in which this ratio was the farthest away from the above-mentioned range of 0.5 and 2, the whole process was repeated one more time. The procedure was repeated for each of the populations, starting with the farthest outlier, until none of the calculated ratios was outside of this range. All remaining populations were used to calculate  $e$  in the first formula above. The final social network size was computed from the average social network size calculated for people in this stage ( $C_i$ ) (14, 24, 30).

In the end, the effect of various demographic and

background factors on the social network size of the residents of Tehran Province was analyzed using a linear regression model. Data analysis was performed using Excel 2010 and SPSS version 17.

## Results

A total of 1029 people were interviewed in this study; 46.7% participants were male. The majority (42.3%) was in the 18 - 25 year age group; the last educational degree of most participants was high school graduate or bachelor's degree, and they were mostly university students (Table 1).

Using the thirteen known populations, the social network size was estimated at 200.4 (CI<sub>95%</sub>: 188.3, 212.5), which was used for the back estimation of the known populations in this study. The Pearson correlation coefficient between the real and estimated sizes of these known populations was 0.44 (*P* value = 0.12).

The e/r ratio was approximately 1 in all but three of the subpopulations; exceptions were first graders, and NVD, C/S where e/r ratios were 0.40, 0.43, and 0.3 respectively, and therefore the social network size of Tehran province was estimated by eliminating the subpopulations C/S, first graders and NVDs in that order.

**Table 1:** Demographics of study participants

	n	%
<b>Sex</b>		
Male	481	46.7
Female	548	53.3
<b>Age (yr)</b>		
18-25	435	42.3
25-40	423	41.2
> 40	170	16.5
<b>Education</b>		
Illiterate	72	7.0
Elementary School	194	18.9
High School	375	36.5
Associate / Bachelor	345	33.5
Master / Doctorate	42	4.1
<b>Marital Status</b>		
Single	555	54.0
Married	426	41.5
Divorced / Widowed	46	4.5
<b>Occupation</b>		
Unemployed	80	7.8
University Student	268	26.2
Service Provider	129	12.6
Businessman / woman	169	16.5
Government Employee	148	14.5
Retired	28	2.7
Housewife	202	19.7

**Table 2:** Estimate/Real (E/R) ratio & Estimate-Real correlation changes after stepwise excluding the known populations with out of range E/R ratio

Estimation Steps	Social Network	E/R ratio					E-R correlation	
		C/S	PS	NVD	Marjan	Total	Coefficient	<i>P</i> value
Based on data of 13 known groups	200.4	0.34	0.4	0.43	2.2	1.17	0.32	0.28
All groups except C/S*	220.8	-	0.36	0.39	2	1.13	0.51	0.09
All groups except C/S, PS**	243.5	-	-	0.35	1.81	1.09	0.72	0.01
All groups except C/S, PS, NVD***	259.1	-	-	-	1.7	1.09	0.82	0.004

C/S\*: Had "Cesarean section" last year/ PS\*\*: Started "primary school"last year/ NVD\*\*\*:Had "Normal Vaginal Delivery" last year

Table 2 represents the estimated social network size after excluding each of the above subpopulations, as well as the e/r ratio for populations whose sizes have been calculated based on the

estimated network sizes that continue to be out of the desired range.Having eliminated C/S, primary school and NVD groups, the estimated Cs were 220.8, 243.4 and 259.1 respectively. In addition,

the correlation between the real and estimated sizes of remaining group was improved significantly in every steps ( $r= 0.32$  based on groups, 0.51, 0.72 and 0.82 by dropping C/S, primary school and NVD groups respectively).

Table 3 demonstrates results of applying the back

estimation method to subpopulations with estimated sizes within the acceptable range that were used in the final network size estimations. Back estimations and  $e/r$  ratios of all populations are presented in Table 4.

**Table 3:** Known populations, source of data, back estimation of their population sizes, and the E/R ratio

Reference group	Organization providing data	Real size (proportion)	E/R Ratio*	
			Estimate	Ratio
Started primary school last year	Ministry of Education	95544 (0.64%)	-	-
Graduated from high school last year	Ministry of Education	55645 (0.37%)	73289	1.3
Took part in university entrance exam last year	Higher Education Organization	96275 (0.65%)	88547	0.9
Started university last year	Higher Education Organization	48600 (0.32%)	38725	0.8
Got married last year	Civil Status Registration Organization	101488 (0.68%)	73789	0.7
Got divorced last year	Civil Status Registration Organization	30403 (0.20%)	19917	0.7
Had normal vaginal delivery last year	Ministry of Health	53112 (0.35%)	-	-
Had Cesarean section last year	Ministry of Health	106275 (0.71%)	-	-
Have an office job in an elementary school	Ministry of Education	37373 (0.25%)	43829	1.2
First name "Hamed"	Civil Status Registration Organization	34059 (0.23%)	55868	1.6
First name "Abolfazl"	Civil Status Registration Organization	63465 (0.42%)	53705	0.8
First name "Sara"	Civil Status Registration Organization	47624 (0.32%)	53871	1.1
First name "Marjan"	Civil Status Registration Organization	18958 (0.12%)	32345	1.7
Total	-	-	-	1.2

**Table 4:** E/R ratio in separate estimations of social network size for men and women (before and after excluding the three out of range E/R ratio known groups)

Reference group	Initial estimation in men		Final estimation in males		Ratio in Women Before*		Ratio in Women after*	
	Estimate	Ratio	Estimate	Ratio	Estimate	Ratio	Estimate	Ratio
Started primary school last year	29545	0.3	-	-	48191	0.5	49995	0.5
Graduated from high school last year	107014	1.9	79778	1.4	82114	1.5	85188	1.5
Took part in entrance University Exam last year	135994	1.4	101382	1.1	92306	1.0	95761	1.0
Started University last year	51316	1.1	38255	0.8	48774	1.0	50599	1.0
Got married officially last year	83264	0.8	62073	0.6	107884	1.1	111922	1.1
Got divorced last year	20639	0.7	15386	0.5	31011	1.0	32172	1.1
Had normal vaginal delivery last year	19084	0.4	-	-	26935	0.5	27943	0.5
Had Cesarean section last year	24032	0.2	-	-	50229	0.5	52110	0.5
Having an office job in an elementary school	41420	1.1	30878	0.8	72360	1.9	75068	2.0
First name "Hamed"	95563	2.8	71242	2.0	48191	1.4	49995	1.5
First name "Abolfazl"	79589	1.3	59333	0.9	58965	0.9	61172	1.0
First name "Sara"	64321	1.4	47951	1.0	75126	1.6	77938	1.6
First name "Marjan"	37038	2.0	27611	1.5	46735	2.5	-	-

Figure 1 represents the scatter plot of the real against estimated sizes of the thirteen known populations, and the same measures after excluding the three abovementioned subpopulations. The final network size estimated by using the remaining ten subpopulations was 259.1 (CI<sub>95%</sub>: 242.2, 276).

In the univariate analysis, social network size was significantly associated with gender ( $P=0.01$ ), age ( $P<0.001$ ), and occupation ( $P<0.001$ ). Men had larger networks comparing to women (291.8 versus 230.4), and C in younger people was larger than that in older ones (328.5 versus 201.1).

Moreover, retirees had smaller networks comparing to the others.

Multivariate analysis showed that variables of gender and age impacted social network size (Table 5). Age was also proved to have a significant impact on the respondents' knowing the studied subpopulations ( $P < 0.001$ ).

The possibility of knowing a person who had graduated from high school, participated in the university entrance exam, or been admitted into university in the past year was higher in the younger age group (56.6%, 63.4% and 39.3% respectively) (Table 6).

**Table 5:** Comparison of social network size in different sub groups' based on different background variables

	C mean (SE)		Crude P value	Adjusted P value
<b>Sex</b>			.010	.01
Male	291.8	(15.5)		
Female	230.4	(8.6)		
<b>Age (years)</b>				
18-25	328.5	(16.9)	< .001	<.001
25-40	211.6	(9.6)		
> 40	201.1	(12.7)		
<b>Education</b>				
Illiterate	276.3	(44.5)	.96	-
Elementary School	236.2	(19.1)		
High School	265.3	(13.7)		
Associate / Bachelor	270.4	(15.1)		
Master / Doctorate	184.1	(23.6)		
<b>Marital Status</b>				
Single	270.1	(11.8)	.13	0.22
Married	248.8	(13.6)		
Divorced / Widowed	222.9	(27.7)		
<b>Occupation</b>				
Unemployed	266.3	(23.7)	< .001	.055
University Student	362.4	(25.6)		
Service Provider	229.4	(16.5)		
Businessman / woman	211.0	(13.5)		
Government Employee	215.5	(17.5)		
Retired	198.9	(37.4)		
Housewife	220.3	(12.9)		

Table 6: Probability of knowing subgroups by age

Known group	Knowing probability	Age Groups		
		18-25	25-40	>40
High school graduate	Count	246	134	60
	% within High school graduate	55.90%	30.50%	13.60%
	% within Age groups	56.60%	31.70%	35.30%
University applicant	Count	276	199	83
	% within University applicant	49.50%	35.70%	14.90%
	% within Age groups	63.40%	47.00%	48.80%
Started University last year	Count	171	111	38
	% within Started University	53.40%	34.70%	11.90%
	% within Age groups	39.30%	26.20%	22.40%

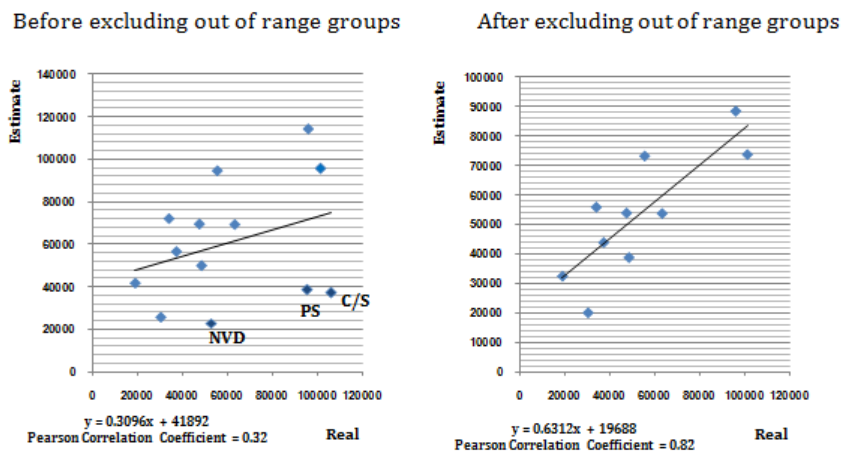


Fig.1: Real – Estimate scatter plot of known population sizes (before and after excluding the three out of range E/R ratio known groups)

## Discussion

Based on the findings of this study and by using the MLE method, the social network size of the residents of Tehran Province was estimated at 259.1 which were calculated using 10 known populations with an e/r ratio between 0.5 and 2. Age and gender were determinants of peoples' network sizes. Estimated network sizes were different in male and female respondents, and appropriate known populations used for C estimation in these two groups were not equal.

Our experience showed that all subpopulations with known size were not appropriate to be used in the estimation of C. We dropped three subpopulations to improve the internal validity of our size estimations, which also used in other comparable studies (11, 14, 20, 22).

By eliminating these 3 subpopulations, the social network size increased by 29% (from 200.4 to 259.1). The national study was primarily biased in a similar manner as well (22). Therefore, estimation of C based on the size of some known subpopulations is a stepwise technique and has to be applied with some considerations; otherwise the estimated

C might have big bias.

The estimated sizes for known subpopulations C/S, NVD, and first graders were less than half the real sizes, and for the subpopulation of people with the first name Marjan the estimated size was more than twice the actual size. After all three subpopulations were eventually eliminated, the correlation coefficient was more than double (0.82 as opposed to 0.32), and became statistically significant. Moreover, the estimations for the subpopulation Marjan—a Persian female first name—fit within the acceptable range, and was closer to 1. This indicated that the social network size estimated by using the ten remaining known populations would yield a more reliable estimate of hidden subpopulations too.

Similar studies confirm likewise that elimination of populations with estimations outside of the range 0.5 – 1.5 increases the correlation coefficient between back estimations and real sizes. Paniotto et al. conducted a research in Ukraine to estimate the number of IDU's FSW's and MSM. Of the 22 known populations used in this study, 13 fell within the acceptable  $e/r$  ratio range, and were used in the final network scale-up calculations. The estimated sizes of known populations had a high correlation coefficient with the real sizes using the back estimation technique ( $r = 0.912$ ), and this number reached 0.94 after nine of the known populations were eliminated (14).

In another study conducted on 1554 individuals in the U.S., the correlation coefficient was 0.79 using 29 known populations. After elimination of two populations with a discrepancy between their real and estimated sizes, the correlation rose to 0.94 (11).

In the present study, the average  $e/r$  ratio using thirteen known populations was calculated to be 1.17. Removing each of the outliers brought this ratio closer to 1, and after eliminating the 3 subpopulations mentioned earlier, the  $e/r$  ratio reached 1.09. This figure indicated an average overestimation of about 9% in the calculations. In the Ukrainian study, the  $e/r$  ratio was 1.65, which denotes an overestimation of 65% in population size estimations (14).

Overestimation seems to be one of the common

limitations of the network scale-up method with small subpopulations (11, 30). In the present study, this problem occurred with the subpopulation of people with the first name Marjan, which was the smallest subpopulation with only 0.12% of the total population of Tehran Province. Since nicknames are common in Iran, this overestimation may be because the respondents' network includes people called Marjan who are registered in the census with a different name; this can be true with several names. After eliminating the C/S subpopulation, the  $e/r$  ratio for the subpopulation Marjan fell within the acceptable range ( $e/r = 2$ ), and after the two subpopulations NVD and first graders were removed, this ratio decreased even further ( $e/r = 1.7$ ).

Another common calculation error in the network scale-up method is underestimation of the size of large subpopulations (11, 30). The largest subpopulations used in this study were C/S, first graders, married people, and university applicants. The first two were underestimated as expected, while the other two had more accurate estimations, probably on account of a more satisfactory transmission and the respondents being more informed on these properties due to the significance that Iranians attribute to marriage and university education. Based on the Killworth et al. study on seroprevalence in the United States, it is difficult for respondents to estimate the number of acquaintances they have in large subpopulations (11).

In addition to the estimation error in the C/S subpopulation, another explanation for this group having the lowest  $e/r$  ratio (0.34) and an  $e/r$  ratio lower than 0.5 for the NVD group is transmission error. The similar natures of these two subpopulations as well as their connection to gender suggested that respondents' gender impacts their awareness of NVD or C/S occurring within their social networks. In order to assess the validity of this theory, network size estimations and back estimations for populations were performed based on the respondents' gender.

Based on our findings, the  $e/r$  ratios for known populations were different in male and female respondents (Table 4). The  $e/r$  ratios were out of range for the subpopulations C/S, NVD, and first



graders when estimated through the networks of male respondents, but this limitation did not exist in case of female respondents. The difference seems to be understandable considering the nature of these subpopulations and the Iranian culture: an Iranian male respondent would be less likely to be aware of childbirth conditions or first graders in his social network compared to a female respondent in a similar situation. In the Ukrainian study, similar circumstances were encountered when estimating the number of militiamen, since there was a higher likelihood for male respondents to know someone in this subpopulation on account of their cultural status and social relations compared to women (77% vs. 68%). Therefore, using the militiamen subpopulation to estimate the social network size leads to different results in male and female respondents (14).

Barrier effect is another significant factor that can compromise the estimation of population sizes in the network scale-up method. One common barrier, which could be a concern in the present study, is the respondents' age. Those in the 18 – 25 year group were more likely than others to know someone in the subpopulations high school graduates, university applicants, and university students, considering how these categories are age-related. Based on the results shown in Table 6 however, barrier effect did not affect the results of this study. This is probably because the 18-25 year old group is relatively larger in the general population, and our 18-25 year old group was proportionately larger too. In the Ukrainian study, the two subpopulations Polish and Moldavian were removed due to the risk of barrier effect and consequent estimation errors.

In the final stage of the present study, the remaining ten subpopulations were used to estimate the social network size at 259.1, and this was used as the base for estimating the known population sizes with the most suitable  $e/r$  ratio and the highest correlation with their true sizes.

This number was smaller compared to similar study in the country, carried out by Shokoohi et al. in Kerman Province in 2010 ( $c = 303$ ). Their study was performed on 500 men between 18 and 45 years of age, and the data were collected

through interviews. The difference between the social network sizes of the two provinces of Tehran and Kerman seems to be due to the lifestyle specific to metropolitan areas and capital cities that somehow limits social relationships; one other reason may be that the Kerman study was restricted to male samples, who are expected to have more social contacts and consequently larger networks than females in a city like Kerman (27).

Our estimate was also smaller than the national average. In the national study, two estimates were computed by entering 23 known groups in the study using regression-based ( $c=308$ ) and ratio-based ( $c=380$ ) methods. According to their results, the ratio-based method, which is also used in the present study, is the recommended approach because of higher internal validity and prediction validity (22). The ratio of the population size to the total Tehran population was beyond the range of 0.1% to 4.0% for 10 of the 23 known groups in the national study; differences in population compositions and their effect on results can partly explain the different estimates in these two studies. Moreover, unique social, cultural, political, and economic processes in Tehran have created a different pattern of social interactions that restricts inter-individual relationships and limits the social network size.

Social science studies also confirm that in recent years, the nature of urban development in megacities such as Tehran has entailed loosening in social ties. This is due to concentration of population in metropolitan areas and absence of alternative social practices (31).

In the Ukrainian study, the social network size was estimated at 202 using a network scale-up approach and the MLE method, which was also employed by the present study (27).

In a 2012 study conducted in Japan by Ezoe et al. with the purpose of estimating the MSM population, the network size was estimated to be 363.5 regardless of gender, with 174 being male. Out of the initial ten known populations used in this study, only three were eventually used to estimate the social network size in the pilot stage: male fire fighter, policemen, and military personnel. Of the seven eliminated populations, five were removed

on account of estimation error, that is, their estimated sizes were not consistent with their real sizes, and two were removed due to transmission error (20). The social network size was estimated at approximately 291 in the United States, as shown in the 2001 study by McCarty et al., using four separate national sets of samples. The study used the network scale-up method on 29 known populations, three of which were similar to the subpopulations in the present study and those were: first names, people who gave birth within the last year, and victims of car accidents (24).

The social network size estimated by Killworth et al. was 286, which is larger than the present study, and this may be due to the numerous cultural and social differences between Iran and the United States (11).

Although in the Kerman study, there was no significant relationship between the social network size and any of the demographic factors (age, education, marital status, and occupation) (27); the results of the present study showed that men comparing to women and younger people comparing older ones had a significantly larger social network, which seems understandable on account of cultural and social considerations.

In the national study and the Chinese study, the social network sizes of men and younger people was larger compared to those of women and other age groups, and this is due to men and young people having wider social circles (21, 22). Married people had smaller social networks compared to single people in the two studies mentioned above; this difference was found to exist in the present study as well, although the difference between these two groups was statistically not significant in Tehran Province (248.8 versus 270.1).

In this study, there were limitations associated with the network scale-up method.

Firstly, it is unrealistic to expect accurate and flawless estimations, because although there is a precise and specific definition for the term “knowing a person”, data collected on social networks is self-reported. In minimize this limitation, the present study attempted to use known populations that were easily recognizable by all respondents, and had clear and uniform definitions for all

members of the society. Furthermore, the time span for the questions was the past year so that respondents could remember the details with fewer errors. Another issue, which is quite unavoidable, is lack of definite boundaries separating many subpopulations in the respondents' points of view. Researchers are limited to subpopulations for which official statistics exist, and these subpopulations have been assigned specific definitions that are not necessarily consistent with those of various members of the society.

The last issue was our sampling method. We intended to select a representative sample of general population in order to minimize different types of selection and information biases. Although it was not a fully random sample of whole community and there are some methodological considerations, based on the existing experiences in Iran and the result of a methodological study that was performed for comparing three popular sampling method in this type of studies (street-based, telephone-based and home-based interviews) in this regard, it seems street-based sampling from different geographical zones would be a feasible sampling scheme which was used in the national study as well.

## Conclusion

Based on the above explanation, it seems that the social network size of Tehran's residents is different not only with similar studies in other countries but also with the results of national survey in Iran, which might be mainly because of the special social and cultural pattern of communications among different communities. We also showed that the  $C$  varies considerably in males and females, in young and old people. Therefore, local estimations for  $C$  in different sub-populations are needed to improve the accuracy of NSU estimations.

## Ethical considerations

Ethical issues (Including plagiarism, Informed Consent, misconduct, data fabrication and/or fal-

sification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

## Acknowledgment

This work was supported by Tehran University of Medical Science and Kerman University of Medical Science. The authors declare that there is no conflict of interest.

## References

1. UNAIDS/IMPACT/FHI workshop (2003) *Estimating the size of populations at risk for HIV, Issues and methods*. In. United States Agency for International Development.
2. Potts M, Halperin DT, Kirby D, Swidler A, Marseille E, Klausner JD, Hearst N, Wamai RG, Kahn JG, Walsh J (2008). Reassessing HIV Prevention. *Science*, 320: 749-50.
3. Mills S, Sidel T, Magnani R, Brown T (2004). Surveillance and modelling of HIV, STI, and risk behaviours in concentrated HIV epidemics. *Sex Transm Infect*, 80(Suppl II): ii57-ii62.
4. Magnania R, Sabin K, Sidel T, Heckathorn D (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS*, 19 (suppl 2): S67-S72.
5. Bernard HR, Hallett T, Iovita A, Johnsen EC, Lyerla R, McCarty C, Mahy M, Salganik MJ, Saliuk T, Scutelnicu O, Shelley GA, Sirinirund P, Weir S, Stroup DF (2010). Counting hard-to-count populations: the network scale-up method for public health. *Sex Trans Infect*, 86: ii11-ii15.
6. UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2010) *Guidelines on estimating the size of populations most at risk to HIV*. In. Geneva
7. Laska EM, Meisner M (1993). A Plant-Capture Method for Estimating the Size of a Population from a Single Sample. *Biometrics*, 49 (1): 209-20.
8. Wittes JT, Colton T, Sidel VW (1974). Capture-recapture models for assessing the completeness of case ascertainment using multiple information sources. *J Chronic Dis*, 27: 25-36.
9. Zhang D, Wang L, Lv F, Su W, Liu Y, Shen R, Bi P (2007). Advantages and challenges of using census and multiplier methods to estimate the number of female sex workers in a Chinese city. *AIDS Care*, 19: 17-19.
10. Hook EB, Regal RR (1995). Capture-recapture methods in epidemiology: methods and limitations. *Epidemiology Review*, 17: 243-64. (Corrigenda in *American Journal of Epidemiology*, 1998, 148:12-19.).
11. Salganik MJ, Mello MB, Abdo AH, Bertoni N, Fazito D, Bastos FI (2011). The Game of Contacts: Estimating the Social Visibility of Groups. *Soc Networks*, 33 (1): 70-78.
12. Killworth PD, Johnsen EC, McCarty C, Shelley GA, Bernard HR (1998). A social network approach to estimating seroprevalence in the United States. *Social Networks*, 20 (1): 23-50.
13. Paniotto V, Petrenko T, Kupriyanov O, Pakhok O (2009) *Estimating the Size of Populations with High Risk for HIV Using the Network Scale-Up Method*. In. Kiev International Institute of Sociology, Ukraine.
14. Killworth P D, McCarty C, Bernard H R, Shelley G A, Johnsen E C (1998). Estimation of Seroprevalence, Rape, and Homelessness in the United States Using a Social Network Approach. *Evaluation Review*, 22 (2): 289-308.
15. Kadushin C, Killworth PD, Bernard HR, Beveridge AA (2006). Scale-up methods as applied to estimates of heroin use. *J Drug Issues*, (6): 417-40.
16. Salganik M J, Fazito D, Bertoni N, Abdo A H, Mello M B, Bastos F I (2011). Assessing network scale-up estimates for groups most at risk of HIV/AIDS: evidence from a multiple-method study of heavy drug users in Curitiba, Brazil. *Am J Epidemiol*, 174 (10): 1190-96.
17. Johnsen E C, Bernard H R, Killworth P D, Shelley G A, McCarty C (1995). A social network approach to corroborating the number of AIDS/HIV+ victims in the US. *Social Networks*, 17 (3): 167-87.
18. UNAIDS & the US Office of the Global AIDS Coordinator (March 28-30th, 2012) *Consultation on Network scale-up and other size estimation methods from general population surveys*. In. consultation on network scale-up & other size estimation

- methods from general population surveys., New York City, .
19. Rwanda Biomedical Center/ Institute of HIV/AIDS DPaCDRI (2012). *Estimating the Size of Populations through a Household Survey (ESPHS), Rwanda 2011*. ICF International Calverton, Maryland, USA.
  20. Satoshi Ezoe, Takeo Morooka, Tatsuya Noda, Miriam Lewis Sabin, Soichi Koike (2012). Population Size Estimation of Men Who Have Sex with Men through the Network Scale-Up Method in Japan. *Plos ONE*, 7 (1): e31184.
  21. Bao S L, Wu G H, Zhang W, Zhou C, Guo W, Zhou X L., Wang L (2012). Application on size estimation through the network scale-up method on men who have sex with men in Chongqing municipality. *Zhonghua Liu Xing Bing Xue Za Zhi*, 33 (10): 1036-9.
  22. Rastegari A, Haji-Maghsoudi S, Haghdoost A, Shatti M, Tarjoman T, Baneshi M R (2013). The Estimation of Active Social Network Size of the Iranian Population. *Glob J Health Sci*, 5 (4): 217-27.
  23. Midanika LT, Greenfield TK (2003). Telephone versus in-person interviews for alcohol use: results of the 2000 National Alcohol Survey. *Drug and Alcohol Dependence*, 72 (3): 209-14.
  24. McCarty C, Killworth PD, Bernard HR, Johnsen EC, Shelley GA (2001). Comparing two methods for estimating network size. *Human Organization*, 60 (1): 28-39.
  25. Russell HB, Killworth PD, Johnsen EC, Shelley GA, McCarty C (2001). Estimating the Ripple Effect of a Disaster. *Connections*, 24 (2): 18-22.
  26. Pridemore W, Damphousse KR, Moore RK (2005). Obtaining sensitive information from a wary population: A comparison of telephone and face-to-face surveys of welfare recipients in the United States. *Soc Sci Med*, 61 976-84.
  27. Shokoohi M, Baneshi MR, Haghdoost AA (2010). Estimation of the active network size of Kermanian Males. *Addict Health*, 12 (3-4): 81-88.
  28. Bernard HR, Johnsen EC, Killworth PD, Robinson S, Barbara S (1991). Estimating the size of an average personal network and of an event subpopulation: some empirical results. *Soc Sci Res*, 20 (2): 109-21.
  29. Snidero S, Zobec F, Berchialla P, Corradetti R, Gregori D (2009). Question order and interviewer effects in CATI scale-up surveys. *Sociological Methods & Research*, 38 (2): 287-305.
  30. Johnsen EC, Bernard HR, Killworth PD, Shelley GA, McCarty C (1995). A social network approach to corroborating the number of AIDS/HIV+ victims in the US. *Social Networks*, 17 (3-4): 167-87.
  31. Mosavi Y (2002). Weakening of social relationships of inhabitants of metropolitan cities. *Modares*, 6 (4): 113-32 [In persian].