Original Article **Journal of Addiction and Health**
Vol 2, No. 3-4, Summer & Autumn 2010

Received: 15.5.2010
Accepted: 9.8.2010

Estimation of the Active Network Size of Kermanian Males

Mostafa Shokoohi MSc*, Mohammad Reza Baneshi PhD**,
Ali Akbar Haghdoost PhD***

* MSc in Epidemiology, Kerman Physiology Research Center, Kerman University of Medical Sciences, Kerman, Iran.
** Assistant Professor of Biostatistics, Regional Knowledge Hub for HIV/AIDS Surveillance, Kerman University of Medical Sciences, Kerman, Iran.
*** Associate Professor, Department of Epidemiology, Research Center for Modeling in Health, Kerman University of Medical Sciences, Kerman, Iran.

Abstract

Background: Estimation of the size of hidden and hard-to-reach sub-populations, such as drug-abusers, is a very important but difficult task. Network scale up (NSU) is one of the indirect size estimation techniques, which relies on the frequency of people belonging to a sub-population of interest among the social network of a random sample of the general population. In this study, we estimated the social network size of Kermanian males (C) as one of the main prerequisites for using NSU.

Methods: A 500 random sample of Kermanian males between 18 and 45 years old were interviewed. We asked the size of their active networks using direct questions. In addition, we received the frequency of six names from the vital registry office among Kermanian males, and we estimated C indirectly using the received frequencies and the frequency of these names among the networks of our sample.

Findings: Although different methods showed quite different Cs between 100 and 350, the best estimation for C was 303, which means that on average each Kermanian male knows around 303 males between the age range of 18 and 45 years. The estimated C did not have any strong association with the demographic variables of our subjects.

Conclusion: Using the estimated C we may use the NSU technique to assess the frequency of many important hidden sub-populations such as drug-abusers and those who have sexual contact with men and women.

Key words: **Size estimation, Social network, Networking, Addiction, Hidden population, Hard to reach population.**

Page count: 8
Tables: 3
Figures: 0
References: 18

Address of Correspondence: Ali Akbar Haghdoost PhD, Research Center for Modeling in Health, Kerman University of Medical Sciences, Kerman, Iran.
Email: ahaghdoost@kmu.ac.ir

Introduction

Population size estimation (PSE) is an essential part of health system management such as the HIV surveillance system. Especially in countries with low-level or concentrated HIV epidemics (such as Iran), estimating the size of particularly vulnerable, hard-to-reach populations such as addicts, Female Sex Workers (FSW), and Men who have Sex with Men (MSM) is very important.[1,2] Without any doubt, the number of addicts and drug abusers is a very important question which in order to address this question, direct sampling methods cannot be used.

These estimates help stakeholders in planning, resource allocation and setting up high-quality bio-behavioral surveillance studies (BSS). However, without PSE it would be hard, if not impossible, to assess the needs for sufficient services and to convince decision-makers that these needs ought to be met.[2,3]

Although there is no doubt about the importance of PES, the available statistics in this regard usually scatter with a wide range of variation. For example, we do not have any accurate estimation about the number of FSW, MSM and even addicts in Iran. Based on official figures, published in 2004, the size of injected drug users (IDU) was approximately 200,000 and during the last years it has been more or less constant while the pattern of drug use has changed and wide national methadone maintenance therapy has been implemented.[4] This controversy about the size of other hidden groups is even more profound.[5]

There are simple reasons behind these wide uncertainties; the hard-to-reach nature of these subgroups and complicated methods in estimating the size of these hidden populations. Briefly, we can classify the PSE methods into two main categories; namely, direct and indirect methods.[2] In the direct method, we either count all members of target populations (census technique) mainly in some of their venues (such as the number of IDUs in their shooting galleries), or use semi-probability methods in which we sample a defined part of the target populations (enumeration technique) and count them in their venues.[2]

Both of these two direct approaches are very hard to implement and have their own considerations, which limit their applications. Selection bias constitutes serious threat to these techniques. While social desirability affects the disclosure of membership in a stigmatized population, the inherent invisibility of hidden populations biases census and enumeration approaches toward more visible parts of the populations; indicating that these methods are not feasible in hidden populations such as drug abusers.[2,6]

In contrast to the direct approaches, indirect methods help us estimate the size of a target population without counting them directly. Capture-recapture technique is one of the first indirect methods, which estimates the size of a population by assessing the number of subjects who were captured in at least two independent samples.[7-9] Multiplier technique is an alternative method, which requires one sample from a target population with some information from a benchmark.[10,11] Although the concepts of these two techniques are easy to understand, since we sometimes cannot find appropriate samples and their assumptions are not met in some settings, their application is limited.[8] Therefore, capture-re-capture and multiplier techniques are not used in all settings.[2,11]

One of the alternative indirect approaches is the Network Scale Up (NSU) technique. Somehow, this is the only real indirect technique as we never approach our target populations in any way. The basic principle underlying NSU is that individual social networks represent the general population, and the description of these networks describes the characteristics in the general population. More precisely, the proportion of individuals belonging to a sub-population in the network of a representative sample has a direct association with the real size of that sub-population in the general population.

The NSU approach thus relies on asking a random sample of individuals from the general population whether they know any members of the population of interest. It also requires information on the average personal network size (usually unknown) in the general population and the number of the total population (usually known). The NSU method is simple and does not require contact with the hidden population directly. Relevant indicators may be added to any nationally representative survey to produce PSE for different hard-to-reach populations at the same time.[2,12-16]

In the first view, NSU is one of the easiest PSE methods. However, to be able to use the technique, we should have clear and accurate information about the social network size of the population.

Based on extended social studies in the United States (US), on average the size of the active social network of each American is around 290, which means that each American knows around 290 people (the definition of active network is given in the methods section). This very important number is a base for many NSU studies in US.[13,17] However, the size of social network in Iran similar to many other countries has not been explored deeply; therefore, we do not have this baseline information to use in NSU projects.

Based on this demand, we carried out this study as one of the first basic studies in this field in Iran not only to estimate the size of social networks in Kerman city among young males but also to standardize this technique for use in other parts of Iran.

Methods

This cross-sectional study was conducted in Kerman city (the capital of Kerman Province), located in south-east of Iran. Based on 2006 census, the total population of this city was approximately half a million inhabitants.

However, our target population were males between 18 and 45 years old who lived in Kerman city for at least over the past five years (N = 132,651). A total sample of 500 individuals of our target population was interviewed adapting a purposive sampling. Samples were selected from crowded areas; 150 persons from four main universities (Kerman Medical University (KMU), Bahonar University, Islamic Azad University, and Teacher Training University), 290 from 11 crowded areas in the city, and 60 in their work places.

To estimate C, direct and indirect methods were applied as follows. Four trained interviewers approached the samples and filled the questionnaires in face to face interviews. Having introduced themselves, the interviewers explained the main objectives of the study and convinced the samples to participate in this study. However, before asking the main questions, a verbal consent was collected. The questionnaires contained demographic questions (age, education, marriage status and job), and questions to estimate the size of their active social network (C) directly and indirectly. We also asked questions about the presence of anybody from a few hidden subgroups such as IDU in their networks. Results of the size of hard-to-reach groups will be published in a separate paper.

Definition of the active social network

We defined C as the size of the active social network which means the number of acquaintances (such as colleagues, relatives and friends) each person knows. Based on this concept, we defined 'know' as 'mutually recognizing each other by sign or name; may be contacted and has had contact in the past one year in person, face to face, phone or email[14]

Direct methods

In direct methods we broke down the networks into categories and asked the respondents the number of people they know in each category. Defined categories were work friends, casual friends, ex or current classmates, family members and neighbors (C1). We explained that each member in their network should only be counted once, belonging to one of the above categories.

We then simply asked the participants about their network size using the above 'know' definition (C2) altogether. As cross-validation, we excluded those subjects who estimated these two Cs quite differently, if C1 and C2 had more than 20% difference.

Indirect methods

In this approach, C is estimated based on the frequency of members belonging to sub-populations with defined sizes in general population. In other words, if we knew the size of some sub-populations in the general population, we would check how many of our samples know at least one people belonging to those sub-populations and would even count their frequencies. Using the following formula, we can estimate C based on this information[14,16]:

$$\frac{m}{c} = \frac{e}{t}$$

Where: m is the average number of people belonging to a sub-population who were known by our samples, c is the active network size, e is the size of the sub-population who we have to know from other sources, and t is the total population.

However, there are alternative formula which estimate C based on the frequency of people who knew at least one from our target population which are presented in details in Killworth et al.'s paper.[14]

To estimate C using the indirect method, we collected information about the frequency of six names among Kermanian males between 18 and 45 years old from the vital registry office. These names were non-common simple names to maximize the validity of responses. Subsequently, we asked our samples if they knew anybody within their active networks with these names and if they knew, we would ask the number of people they knew (C3).

To cross-validate the estimated Cs based on these six names, we calculated the number of males with each name using the estimated Cs based on other names. Then we compared the observed and expected number of males with each name using the Chi-square test.

In addition, we combined the participants' replies to all the six names. Afterwards, using the maximum likelihood method[14], we computed C that maximized the goodness-of-fit (C4) of the distribution.

For both direct and indirect methods, the 95% Confidence Interval (CI) of Cs were estimated using bootstrap technique based on 1000 iterations.

In addition, we examined whether the network size had been influenced by the demographic information (age, education, marriage status and job) using linear regression models. All analyses were performed using Excel and STATA version 10 software.

Results

Nearly two-third of our samples were aged between 18 and 25 years and were single. Furthermore, about half of our subjects were students with academic education (Table 1).

Estimation of C using direct methods (C1 and C2)

The mean (SD) of C1 (the sum of network sizes of subjects in different categories) was 125.4 (283.7). The corresponding statistics for C2 (the total active network size of subjects based on only one direct question) was 134.2 (315.6). The estimated 95% Confidence Interval (CI) for C1 and C2 based on bootstrap method were 104.6-152.9 and 109.4-163.1, respectively.

Using Pearson correlation coefficient, we examined the agreement between C1 and C2. Although the overall correlation coefficient was strong (r = 0.86), in those with large C1 (≥ 500), the association was not high enough (r = 0.26); which means that in those with a large network size the correlation between C1 and C2 was weak.

Table 1. Descriptions of all subjects (before excluding outliers) based on their demographic variables

	N	Percent
Age group (year)		
18-25	286	64.1
26-30	80	17.9
31-35	43	9.6
> 35	37	8.4
Education		
Under diploma	23	5.4
Diploma	153	35.6
Diploma-BS	218	51.1
More than BS	34	7.9
Marriage status		
Single	277	64.8
Engaged	24	5.6
Married/others	127	29.6
Job		
Jobless/soldier	25	5.9
Student	194	46.5
Retailer	130	31.3
Serviceman	18	4.4
Government worker	50	11.9

Table 2. The estimation of C3 based on each name, C1 and C2, and their goodness-of-fits in predicting the frequency of other names

Names	The frequency of each name among 18 and 45-year-old males in Kerman based on the vital registry data	C value	Chi-square statistics which shows the goodness-of-fits of C based on each name in predicting other names
Hamed	0.17%	380.4	232.1
Abolfazl	0.03%	67.5	8233.9
Afshin	0.12%	255.2	1312.9
Ghasem	0.08%	182.8	2708.4
Issa	0.03%	64.2	10676.8
Pooria	0.002%	7.6	87380.7
C1		125.4	4246.7
C2		134.2	3903.3

Table 3. Association between network size estimated using maximum likelihood method (C4) and demographic variables

Demographic variables	C4	Crude			Adjusted		
	(mean ± SD)	β	SE	P value	β	SE	P value
Age group (year)							
18-25 (n = 313)	330.7 ± 183.2	ref			ref		
26-30 (n = 93)	250 ± 178.5	-80.2	21.9	< 0.001	-40.4	26.3	0.125
31-35 (n = 47)	261.6 ± 194.9	-96.1	29.1	0.018	-26.1	36.7	0.47
> 35 (n = 42)	258.9 ± 205.4	-71.8	30.5	0.019	-31.2	39.5	0.42
Education (years)							
Under diploma (n = 30)	188.5 ± 192.2	ref			ref		
Diploma (n = 172)	320.4 ± 177.6	131.9	36.9	< 0.001	93.6	39.8	0.019
Diploma-BS (n = 254)	304.1 ± 194.5	115.5	36.1	0.001	58.6	40.7	0.15
More than BS (n = 39)	307.8 ± 171.6	119.3	45.4	0.009	52.9	49.8	0.28
Marriage status							
Single (n = 323)	323.2 ± 184.1	ref			ref		
Engaged (n = 27)	309.2 ± 210.9	-14.1	37.4	0.70	7.1	38.3	0.85
Married/others (n = 145)	258.6 ± 188.3	-64.6	18.7	0.001	-30.8	27.3	0.26
Job							
Jobless/soldier (n = 30)	302.1 ± 151.5	ref			ref		
Student (n = 223)	339.1 ± 189.1	36.9	35.9	0.30	30.7	37.2	0.41
Retailer (n = 150)	278.4 ± 174.8	-23.6	37.1	0.52	-1.3	38.6	0.97
Serviceman (n = 18)	360.7 ± 196.3	58.8	55.1	0.28	67.5	57.1	0.23
Government worker (n = 6)	241.1 ± 205.6	-60.3	41.3	0.14	-27.1	45.1	0.54

Although we interviewed 500 Kermanian males aged between 18 and 45 years, 71 cases declared a very large C (outer the range of mean plus three standard deviations); also in 85 cases C1 and C2 had 20% difference. Having excluded these subjects, we analyzed the data of 344 cases. Having excluded these subjects, the estimated C2 was 113.1 with a 95% bootstrap CI of 94.8-130.2.

Estimation of C using indirect methods (C3 and C4)

a) Names approach (C3)

The estimated C3 based on each of the six names are summarized in Table 2. However, the results of Chi-square test showed that the goodness-of-fit of each of C3 in the predictions of the frequency of other names were not acceptable.

We also estimated the frequency of names using C1 and C2. These estimations were markedly different compared with the real frequencies based on the vital registry office data.

b) Maximum likelihood approach (C4)

Applying the maximum likelihood method, the estimated C4 (SD) was 303.4 (188.9). The corresponding 95% bootstrap CI ranged from

286.1 to 320.7. Our further exploration revealed that the participants' network size (such as C4) did not have any association with the demographic variables (Table 3).

Discussion

In this study, we found that the computed C based on the direct method (C1 and C2) showed quite different results, particularly in those with extreme network sizes. The computed C based on the likelihood method was more robust and showed that young and middle aged males in Kerman know around 303 males between 18 and 45 years of age. This size did not have any association with their main demographic variables.

In the direct method, we asked our participants to use their passive memories to count everybody in their network. Therefore, we should expect missing or under reporting of some parts of their networks. In addition, extreme results in C2 also imply that the validity of direct methods in at least some of our subjects was not acceptable.

C3s were computed based on different names separately. Again, the range of C3s was very wide from 7.6 to 380.4. In addition, their goodness-of-fits in predicting the frequencies of other names was not acceptable. Therefore, again it seems that validity of C3s is not convincing.

In contrast to C3, our strategy in the estimation of C4 was based on active searching of the participants' memories. We asked the participants if they knew at least one person in their network with each name. Usually active memory searching gives more accurate responses. In addition, it seems that for participants it was much more difficult to count the number of persons with specific names in their networks (required data to estimate C3) in comparison to the reply to naming at least one person with the specific name (required data to estimate C4).[14,16]

To computer C4, using the maximum likelihood method, responses of subjects to all six questions about names were combined and a model was computed to fit the data with the maximum goodness-of-fit. Based on these results, on average each Kermanian young male knows around 303 males in this age group.

We applied four strategies which yield to different C values. However, this was not against our expectation. McCarty et al. announced that it is certainly clear for different measures and methods to produce different numbers or estimates.[16] In USA applying six different methods, the estimated numbers varied from 97 to 399.[14] However, based on the above explanation, we believe that C4 is more accurate compared with other Cs. Based on this logic, in similar studies, indirect methods using similar methodologies were used frequently.[13-16]

It should be noted that we asked our participants how many males aged 18 to 45 they knew. Therefore, for the definite active network size of males in Kerman it is greater than 303. This is because all females and males less than 18 or greater than 45 years of age were not counted. Since the social connection of males with females is less than that among males based on the Iranian and Islamic culture[18] and since people's connections are usually with people in the same age group, we do believe then the real active social network is greater than 303, but less than twofold this number.

In the univariate analysis, the network size of Kermanian young males was influenced by age and education. However, in the multivariable modeling, none of the demographic information affected the C value. More or less constant C in different subgroups is very informative. This means that we may use a valid C for the whole male population. However, this study was carried out in only a middle size city of Iran and we have to report this methodology in females and also in other parts of Iran to make sure if one C for the whole county is enough.[14,16]

This study was one of the first studies using the network scale up method in Iran, and it was carried out in only one middle size city. Therefore, these findings may not represent the network of the whole country. More wide studies with similar methods and using even more names to predict C4 are recommended.[14,16]

Conclusion

Generally, we believe that the indirect method is a more valid technique in the prediction of C4. Based on this finding, we believe on average each Kermanian young male knew around 303 males in this age group. This number is very important statistics that we can use in the network scale up method to predict the size of hidden and hard-to-reach populations such as addicts and other high risk groups (such as IDUs, FSWs and MSM).

Conflict of interest: The Authors have no conflict of interest.

References

1. Estimation of the size of high risk groups and HIV prevalence in high risk groups in concentrated epidemics. Proceedings of the UNAIDS Reference Group on Estimates; 2008 Dec 9-10; Amsterdam, Netherlands; 2008.
2. UNAIDS. Estimating the size of populations at risk for HIV. San Francisco: Family Health International; 2002.
3. UNAIDS, IMPACT, FHI. Estimating the Size of Populations at Risk for HIV [Online]. Jul 2003; Available from: URL: http://data.unaids.org/publications/External-Documents/estimatingpopsizes_en.pdf/
4. Gheiratmand R, Navipour R, Mohebbi MR, Mallik AK. Uncertainty on the number of HIV/AIDS patients: our experience in Iran. Sex Transm Infect 2005; 81(3): 279-80.
5. Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Soc Probl 1989; 36(4): 416-30.
6. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS 2005; 19 Suppl 2: S67-S72.
7. Larson A, Stevens A, Wardlaw G. Indirect estimates of 'hidden' populations: capture-recapture methods to estimate the numbers of heroin users in the Australian Capital Territory. Soc Sci Med 1994; 39(6): 823-31.
8. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epidemiol Rev 1995; 17(2): 243-64.
9. Stephen C. Capture-recapture methods in epidemiological studies. Infect Control Hosp Epidemiol 1996; 17(4): 262-6.
10. Zhang D, Wang L, Lv F, Su W, Liu Y, Shen R, et-al. Advantages and challenges of using census and multiplier methods to estimate the number of female sex workers in a Chinese city. AIDS Care 2007; 19(1): 17-9.
11. Hickman M, Hope V, Platt L, Higgins V, Bellis M, Rhodes T, et al. Estimating prevalence of injecting drug use: a comparison of multiplier and capture-recapture methods in cities in England and Russia. Drug Alcohol Rev 2006; 25(2):131-40.
12. Bernard HR, Johnsen EC, Killworth PD, Robinson S. Estimating the size of an average personal network and of an event subpopulation: Some empirical results. Social Science Research 1991; 20(2): 109-21.
13. Killworth PD, Johnsen EC, Bernard HR, Shelley GA, McCarty C. Estimating the size of personal networks. Social Networks 1990; 12(4): 289-312.
14. Killwortha PD, Johnsenb EC, McCartyc C, Shelleyd GA, Bernar HR. A social network approach to estimating seroprevalence in the United States. Social Networks 1998; 20(1): 23-50.
15. Killworth PD, McCarty C, Bernard HR, Shelley GA, Johnsen EC. Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. Evaluation Review 1998; 22(2): 289-308.
16. McCarty C, Killworth PD, Bernard HR, Johnsen EC, Shelley GA. Comparing two methods for estimating network size. Human Organization 2001; 60(1): 28-39.
17. Bernard HR, Killworth PD, Johnsen EC, Shelley GA, McCarty C. Estimating the Ripple Effect of a Disaster. Connections 2001; 24(2): 18-22.
18. Gray PB. HIV and Islam: is HIV prevalence lower among Muslims? Soc Sci Med 2004; 58(9): 1751-6.

مقاله پژوهشی

مجله اعتیاد و سلامت

سال دوم/شماره ٤-٣/تابستان و پاییز ۱۳۸۹

برآورد اندازه شبکه فعال مردان شهر کرمان

مصطفی شکوهی*، دکتر محمد رضا بانشی، دکتر علی اکبر حقدوست*****

* کارشناسی ارشد، گروه اپیدمیولوژی، مرکز تحقیقات فیزیولوژی کرمان، دانشگاه علوم پزشکی کرمان، کرمان، ایران.
** استادیار آمار زیستی، مرکز منطقه‌ای آموزش نظام مراقبت HIV / ایدز، دانشگاه علوم پزشکی کرمان، کرمان، ایران.
*** دانشیار اپیدمیولوژی، مرکز تحقیقات مدل سازی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران.

تاریخ دریافت: ۸۹/۲/۲۵
تاریخ پذیرش: ۸۹/۵/۱۸

چکیده

مقدمه: برآورد اندازه جمعیت زیرگروه‌های پنهان همچون برآورد تعداد افراد مصرف کننده مواد مخدر، بسیار سخت اما امری مهم است. روش NSU یا Network scale up یک روش غیر مستقیم برای برآورد اندازه زیر گروه‌های جمعیتی است. این روش بر اساس فراوانی تعداد افراد موجود در جمعیت مورد نظر در شبکه اجتماعی فعال افراد جامعه بنا نهاده شده است. هدف از این مطالعه، برآورد اندازه شبکه اجتماعی فعال (C) مردان شهر کرمان (به عنوان یکی از پیش نیازهای اصلی روش NSU)، برای برآورد اندازه جمعیت این زیر گروه‌ها بود.

روش‌ها: در این مطالعه ۵۰۰ مرد کرمانی ۱۸ تا ۴۵ ساله، به طور تصادفی انتخاب و مورد مصاحبه قرار گرفتند. از این افراد با استفاده از سؤالات مستقیم، اندازه شبکه فعال آن‌ها پرسیده شد. در ادامه، از افراد مورد مطالعه در مورد فراوانی ۶ اسم در داخل شبکه آن‌ها سؤال شد. بر اساس اطلاعات اداره ثبت و احوال شهر کرمان، فراوانی این ۶ اسم در شهر کرمان دریافت شد و بر اساس آن مقدار عدد C به صورت غیر مستقیم نیز برآورد گردید.

یافته‌ها: با استفاده از روش‌های متفاوت، مقادیر مختلفی برای C به دست آمد. این اعداد از مقدار ۱۰۰ تا ۳۵۰ متفاوت بود، اما بهترین برآورد عدد ۳۰۳ بود. این عدد به این معنی است که به طور متوسط هر مرد کرمانی تقریباً ۳۰۳ نفر مرد ۱۸ تا ۴۵ ساله را در شبکه خود می شناسد. ارتباط معنی‌داری بین مقدار به دست آمده برای عدد C با متغیرهای دموگرافیک افراد مورد مطالعه مشاهده نشد.

نتیجه‌گیری: با استفاده از عدد C به دست آمده، ما می‌توانیم از روش NSU برای برآورد زیر گروه‌های جمعیتی پنهان و گروه‌هایی که دسترسی به آن‌ها سخت می‌باشد (همچون فراوانی تعداد افراد مصرف کننده موارد مخدر، زنان تن فروش و مردان همجنس باز)، استفاده کنیم.

واژگان کلیدی: **برآورد اندازه جمعیت، شبکه اجتماعی، شبکه بندی، اعتیاد، جمعیت‌های پنهان، جمعیت‌های غیر قابل دسترس.**

تعداد صفحات: ۸
تعداد جدول‌ها: ۳
تعداد نمودارها: -
تعداد منابع: ۱۸

آدرس نویسنده مسؤول: دکتر علی اکبر حقدوست، مرکز تحقیقات مدل سازی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران.
Email: ahaghdoost@kmu.ac.ir