Application of Capture-Recapture for Fine-tuning Uncertainties About National Maternal Mortality Estimates

Bahareh Yazdizadeh, Kazem Mohammad[1], Saharnaz Nedjat[1], Nasrin Changizi[2], Arash Azemikhah[2], Nahid Jafari[3], Laleh Radpoyan[2], Reza Majdzadeh[4]

Department of Biostatistics and Epidemiology, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran, [1]Department of Biostatistics and Epidemiology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, [2]Department of Family Health and Population, Ministry of Health and Medical Education, Tehran, Iran, [3]Center for Network Development, Ministry of Health and Medical Education, Teran, Iran, [4]Department of Biostatistics and Epidemiology, School of Public Health, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran

Correspondence to:
Prof. Reza Majdzadeh,
Knowledge Utilization Research Center,
No 1547, Floor 7, Unit 7, North Karegar
St, Enghelab Ave, Tehran, Iran.
E-mail: rezamajd@tums.ac.ir

Original Article

ABSTRACT

Background: Maternal mortality ratio (MMR) is one of the main indicators of the millennium development goals and its accurate estimation is very important for the countries concerned. The objective of this study is to evaluate the applicability of capture-recapture (CRC) as an analytical method to estimate MMR in countries.

Methods: We used the CRC method to estimate MMR in Iran for 2004 and 2005, using two data sources: The maternal mortality surveillance system and the National Death Registry (NDR). Because the data registry contains errors, we defined three levels of matching criteria to enable matching of cases between the two systems. Increasing the matching level makes the matching criteria less conservative. Because NDR data were missing or incomplete for some provinces, we calculated estimates for two conditions: With and without missing/incomplete data.

Results: According to the CRC method, MMR in 2004 and 2005 were 33 and 25 in the best-case scenarios respectively and 86 and 59 in the worst-case scenarios respectively. These estimates are closer to the ones reported by United Nations Agencies published in 2010, 38 and Hogan's study, 30 in 100,000 live births in 2005.

Conclusions: The MMR estimation by CRC method is slightly different from the international studies. CRC can be considered as a cost-effective method, in comparison with cross-sectional studies or improvement of vital registration systems, which are both costly and difficult. However, to achieve accurate estimates of MMR with CRC method and decrease the uncertainty we need to have valid databases and the absence of such capacities will limit the applicability of this method in developing countries with poor quality health databases.

Keywords: Capture-recapture, epidemiologic methods, Iran, maternal mortality

INTRODUCTION

Maternal death is defined as death during pregnancy or up to 42 days after termination of pregnancy, irrespective of the

duration and site of pregnancy, resulting from any relevant cause or as a result of care provided during pregnancy, excluding accidental death as a cause. The maternal mortality ratio (MMR) is defined as the ratio of maternal deaths to live births within a given time period. Various methods are used to determine MMR, which may be classified as "empirical" or "analytical" according to the goals and sources of information available.[1] Individual countries generally use empirical methods, while global estimates use modeling methods. In developing countries, the most commonly used method is based on surveys that directly and indirectly measure MMR.[2-4] Various studies have explored the best method for determining MMR.[2,5-7] The maternal mortality surveillance system (MMSS) was established in countries such as Iran to determine the exact MMR and consequently to control and prevent maternal mortality resulting from complications arising during pregnancy.

Because MMR is one of the main indicators of development and one of the millennium development goals (MDGs), its estimation indicates a country's progress through the MDGs.

The capture-recapture (CRC) method is an analytical method,[1] in which different data sources are used that separately register a single outcome. Taking into account the number of cases registered by each source, the number of cases common among the sources and the number of cases unregistered, eventually the overall numbers of incident cases of an outcome are estimated in a population in this method. This method is used in estimating health outcomes in situations as diverse as AIDS,[8] tuberculosis,[9] diabetes,[10] meningitis,[11] occupational disabilities,[12] genetic diseases such as Down's syndrome,[13] traffic accidents[14,15] and is even used in showing the sensitivity of databases for finding articles.[16]

To examine the validity of CRC, the results of this study is compared with MMR estimates obtained in other studies through other methods. In estimating MMR, the Inter-Agency Group (Inter-Agency Maternal Mortality Estimate Group; World health Organization, United Nations Children's Fund, United Nations Population Fund and the World Bank), classifies the country according to the available data and uses different methods for estimating MMR in each class. In the 2007

estimation, Iran was in Group F (using census data) and MMR was estimated using the reported proportion of maternal deaths among all female deaths and the estimated deaths among those of fertile age.[17] In the 2010 estimations, Iran was assigned to Group B and MMR estimates were modeled using available country data from the MMSS;[18] Hogan *et al*. estimated MMR using data available from various countries regarding maternal deaths, overall deaths and the adult death rate.[19] The objective of this study is to evaluate the applicability, firstly by means of feasibility in conducting the calculation with data sources from a developing country and secondly agreement with other estimates, of CRC as an analytical method to estimate MMR in countries.

METHODS

The MMSS and National Death Registry (NDR) databases were used to estimate maternal mortality.

Sources of data
Iranian MMSS

All maternal deaths in Iran are registered in the MMSS, using the definitions of the Ninth International Classification of Diseases (ICD-9). The sources of information for the MMSS are as follows: A list of the deaths of married women aged 10-49 years who died of any cause except accidents and provided by the district death registry officer; sudden maternal death that occurred in hospital, reported by the hospital supervisor to the district health network manager; and telephoned or personal reports to the nearest health-care unit for women not covered by the healthcare system. If a death occurs at home or on the way to the health-care unit, a team of investigators is sent to visit the deceased's house and the centers where she has received her maternal care and the appropriate questionnaires are completed. All the collected data are examined by the District Health Committee and the causes of death and their relevancy to pregnancy are assessed.[20]
Iranian NDR

The NDR registers and collects information regarding all deaths in Iran, along with their causes. The sources of information for the NDR are as follows: Hospitals, official cemeteries, medico-legal organizations and all basic health units and rural health centers. The 2004 data

from three provinces and the 2005 data from one province are not complete in the NDR. Therefore, all estimates were calculated twice: Including and excluding these incomplete provinces. We estimated the missing data using the number of maternal deaths registered in the MMSS with the assumption that the sensitivity of the registration system in the MMSS was similar throughout the country.
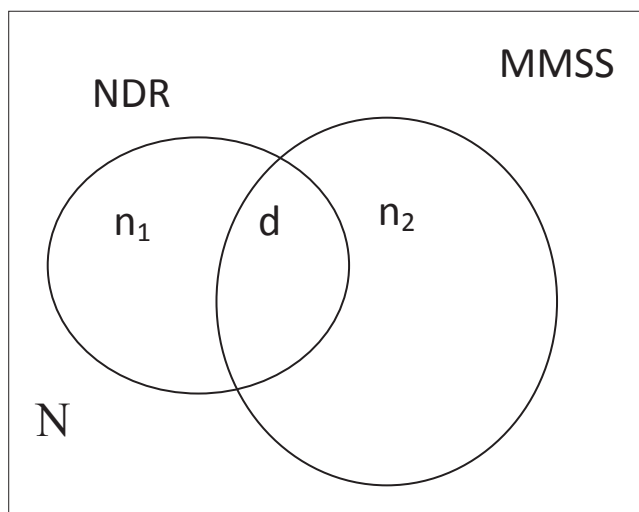
Capture re capture method

When it is not possible to identify and obtain a correct consensus of the individuals of a community with common characteristics (e.g. a specific disease), the CRC method is used to estimate the population. In this situation different sources that have registered the concerned population are used. In fact, in this method, the overall population is estimated by the number of individuals that have not been identified by different sources [Figure 1]. The figure and the relevant text have been added to the article. Based on the number of maternal deaths in each data source, the CRC method estimates the overall number of maternal deaths (including all cases not registered by the data sources). If n_1 is the number of maternal deaths in the NDR, n_2 is the number of cases registered in the MMSS, and d is the number of cases common to the two data sources, then the total number of maternal deaths is calculated as follows:

$$N = [(n_1 + 1)(n_2 + 1)/(d + 1)] - 1,$$

and its variance is calculated using the equation[21]

$$\text{Var}(N) = (n_1 + 1)(n_2 + 1)(n_1 - d)(n_2 - d)/(d + 1)^2 (d + 2).$$[21]

Figure 1: Capture-recapture method with two sources

In applying the CRC method, we must assume that the population under study was closed, the individuals were homogenous, complete matching between the data sources is possible and that the data sources were independent in data-collection. In the present study, the first two assumptions are acceptable; the other assumptions were examined as follows.

Matching between two databases (determining the number of cases common to the two banks)

Prior to performing CRC, it is necessary to determine the number of deceased that are common to the two databases and the requisites of matching must be clearly defined. The similarity of individuals was defined at three levels, taking into account problems in the documented registration of data, differences in computer data entry between the two databases and differences in the database officers' opinions:

Perfect matching

When the following criteria are all the same: name and surname (without taking into account the prefix and suffix), age (±5 years), address and time of death (±1 month).

Semi-perfect matching

Despite small differences in name and surname (by altering or omitting a few letters), two individuals are considered the same when all other criteria are similar.

Non-perfect matching

When the name and surname are worded differently, the address is the same, but the ages are different (±5 years) or months of death are different (difference of 1-2 months), the two individuals are still considered the same. The degree of rigidity in determining the similarity of the two databases is reduced from level one to level three.

Matching process

Because the NDR includes all the deaths, we separated maternal deaths (according to the ICD coding) in the beginning and then compared them with the MMSS data. Two databases were imported to one data file and because the number of cases was few, we handled the matching process manually. We used deterministic approach in each matching level for estimating the MMR.[22]

Examining the independence of the databases

The assumption of independence of databases is most important in estimating the desired outcome

in the CRC method. The term "independence of databases" means that the presence of an individual in one database will not influence the presence or absence of the same individual in another database. No analytical method can test for independence if only two sources of data are used for estimation. One of the sources of data used to identify cases of death in the MMSS is the NDR list of deaths in the past 3 months; this step in the MMSS can lead to dependence of the MMSS on the NDR. Therefore, two procedures were performed to examine the independence of the data-collection procedure: We examined the source of information regarding the occurrence of maternal death for cases registered in the MMSS and performed a sensitivity analysis on the dependency of the two databases.

In the first procedure, all provinces were requested to report to the Family Health Unit of the "Ministry of Health and Medical Education" the source of information for each case of maternal death recorded in the MMSS. The response rate was 84%. These data showed that 46% of maternal mortalities were reported from hospitals by supervisors, while the rest were reported through other routes to the district Family Health Unit. The data also revealed that among cases common to the two databases in 2005, the source of information was the hospital in 60% of cases. Information regarding maternal mortalities is actively sent from the hospital to the district health network by the hospital supervisor, while information regarding hospital deaths is sent passively to the coordination affairs office of the district health network once every 3 months in NDR. Therefore, at least 60% of the data in each database is collected independently. Under the most pessimistic conditions, the degree of dependency between the two databases is 40%.

In sensitivity analysis, if dependency is assumed to be 40% between the two databases for MMR, then the MMR would increase (the number of cases estimated to be registered by only one database is equal to the product of "number of cases currently registered by that database" and "the ratio of cases common in the other database").

$N_2 = (n_2 - d) \times (1/([d/n_1] - dependency))$.

Hence, the numbers of cases in the MMSS databank have been calculated after a certain rate of dependency; they have then been added to the other databank's cases. The total number has then been calculated.

We assumed that the two databases were independent; however, if they are in fact dependent, then the present estimates are lower than the real ratios and the ratios obtained in the present study are the lowest possible.

Examining the validity of cause of death in the NDR

Some of the cases registered in the NDR as maternal deaths are not recorded in the MMSS. To improve the accuracy of the MMR calculation, we inspected all such cases for accuracy by extracting information from the NDR database and by visiting the recorded street addresses.

RESULTS

An analysis of cases common to the two databases, based on the three matching levels revealed that the number of common cases increased as the matching level increased [Table 1]. As seen in Table 1, 100 more cases are reported each year by the NDR than by the MMSS. Table 2 shows the results of MMR, calculated using the CRC method for each level of matching after excluding provinces with incomplete documentation.

A review of the cases of maternal death present in the NDR but absent in the MMSS (thus causing major changes in the MMR) revealed 40 such cases in 2004, with the real cause of death assessable in 27 cases: Of these, the cause of maternal death was pregnancy-related in only seven. In 2005, 30 such cases were present and the real cause of death was assessable in 18 cases; of these, the cause of maternal death was pregnancy-related in only three.

By "best-case" scenario, we are referring to the MMR calculated for 100% independence of two

Table 1: Cases of registered maternal death and common cases in the two databases after matching (excluding provinces with incomplete data)

Year	Observed cases in each database		Number of common cases according to matching level		
	MMSS	NDR	Perfect matching	Semi-perfect matching	Non-perfect matching
2004	259	153	104	110	115
2005	278	151	114	122	126

MMSS = Maternal mortality surveillance system, NDR = National death registry

Table 2: MMR calculated for various matching levels before and after correction of NDR data and assuming complete independence or 40% dependency, for 2004 and 2005

Various matching levels before and after correction of NDR data	2004		2005	
	100% independence of databases	40% dependency	100% independence of databases	40% dependency
Before correction of NDR data				
Perfect matching				
Excluding provinces with incomplete data	46 (43-50)	86	35 (33-38)	59
Including provinces with incomplete data	40	72	32	53
Semi-perfect matching				
Excluding provinces with incomplete data	44 (41-47)	76	33 (31-35)	52
Including provinces with incomplete data	38	63	30	46
Non-perfect matching				
Excluding provinces with incomplete data	42 (39-45)	69	32 (30-34)	48
Including provinces with incomplete data	36	58	29	43
After correction of NDR data				
Non-perfect matching				
Excluding provinces with incomplete data	38 (36-40)	58	27 (26-29)	43
Including provinces with incomplete data	33	49	25	38.56

NDR=National death registry. Numbers in parentheses are the confidence interval, MMR= Maternal mortality ratio

databanks and under circumstances where the calculations have been done in spite of the provinces' incomplete data. By "worst-case" scenario, we are referring to a state where there is 40% dependency between the databanks and the provinces' data have not been entered in completely. According to the CRC method, MMR in 2004 and 2005 were 33 and 25 in the best-case scenarios respectively and 86 and 59 in the worst-case scenarios respectively.

DISCUSSION

We used the CRC method to estimate the Iranian MMR for 2004 and 2005. According to our study, MMR could range between 25 (under the conditions of 100% independence of databases, including provinces with incomplete data, with matching level three and after correcting the NDR) and 59 (under the conditions of 40% dependency of databases, perfect matching level and excluding provinces with incomplete data, perfect matching).

Three main infrastructures are required to do CRC and obtain accurate estimates: Having valid databases, having common variables for homogeneity and the pre-assumptions of CRC.

The number and quality of variables available for comparison is a strong predictor of estimate accuracy in record linkage methods.[23] Correct registration of cause of death in databases is the

first stand to achieve accurate estimates by CRC. In this study the correction of causes of death in one database, i.e. NDR has significantly changed the MMR estimation.

All methods for determining, the number of maternal deaths are influenced by two major errors: Identification of deaths and determining the relevance of the cause of death to pregnancy.[1] In the present study, dependency between the databases existed only in identification of deaths. Because all of the cases reported in the MMSS are examined for the true cause of death, there was no dependence between the two databases in determining the cause of death.
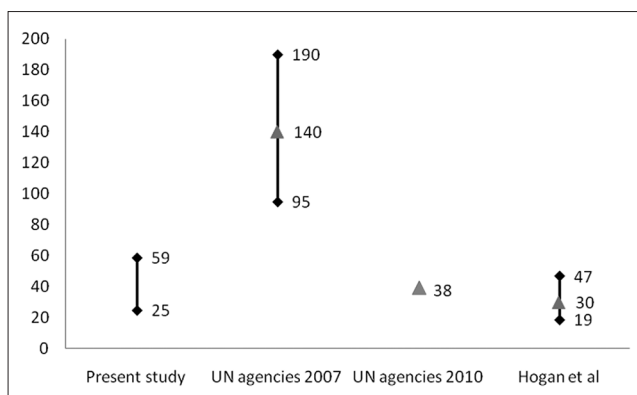
There are various ways of managing the disadvantage of being unable to analytically test the independence of the two databases: The source of information for each individual can be accurately registered in both databases and three databases could be used to estimate MMR in the future. Currently, a third database does not exist in Iran, but data available from public and private insurance firms could possibly be modified for this purpose. Another solution is directly using the "data sources" of NDR rather than using NDR data directly (e.g. using data from cemeteries and the National Organization for Civil Registration along with the MMSS). However, how close is the MMR rate calculated by CRC to the true estimate?

To investigate this issue, we can use the agreement between estimates made at international level by Hogan and WHO [Table 3]. As illustrated, the CRC estimates are close to the estimates made by Hogan and WHO in 2010, but widely differ from the estimates made by international organizations in 2007.

To examine the validity of CRC, the results of this study were compared with MMR estimates obtained in other studies through other methods [Table 3]. As illustrated in Figure 2, the CRC estimates for 2005 are close to the estimates for the same year made by Hogan and WHO (published in 2010), but widely differ from the estimates made by international organizations in published in 2007. Even though, the uncertainty of CRC is high, it implies that the international estimations in 2007 are far from the truth. Therefore, at this point, CRC is a cost-effective method for estimating health related indicators. Since every database has been designed according to its own specific goals and this method is an optimal method of using available data, upgrading the validity of data in databases should be kept in mind. Therefore, quality control should be executed regularly.

Limitation

One of the greatest challenges in the present study was the matching procedure. As we can see, MMR estimation dependent on matching levels are different. Inaccurate registration of case's attributes will damage the estimates. Numerical variables (such as National ID) could be used to identify individuals in all data registration systems,

to increase the flow and validity of the matching process.

As mentioned previously, the most important assumption in using the CRC method is that the data-collection process of the databases is independent. Because only two sources of data were used in the present study, it was not possible to test for independence using analytical tests. Previous similar studies were either certain of the independence of the two databases and provided no proof for this assumption,[24,25] or examined dependency through a questionnaire to the data-collection sites of both sources.[26] Another limitation of the study is that databases were not accessible at the appropriate time; the NDR data was not accessible for the years following 2005. There are two expansions in this regard; the first is conducting new calculation with most recent data to get more reliable estimates of MMR and the second is conducting other methods such as "sisterhood within a 5-year census" along the country to validate the MMR from another perspective.

CONCLUSIONS

In developing countries using available data is a good way of estimating health rates. Therefore, it is vital for part of health systems' activities to focus on promoting the quality of registering the diseases. In order to improve the MMR estimation, it's suggested to add some questions in the survey, which is carried out every 5 years in Iran. In order to increase the feasibility of the CRC, doing studies for improving the quality of NDR and use of three sources for CRC is suggested.

Figure 2: Comparison of maternal mortality ratio estimates for 2005 in Iran. † = In 100,000 live births; ▲: Point estimate; ♦: Lower and upper of the estimates

Table 3: Comparison of various estimates of the Iranian MMR

Source of estimate	2004	2005	2008
Capture-recapture method (present study)	33-89	25-59	-
UN agencies (Iran in Group F for 2007)	-	140 (95-190)	-
UN agencies (Iran in Group B for 2010)	-	38	30 (18-50)
Study by Hoganetal	32 (19-49)	30 (19-47)	28 (17-43)

MMR=Maternal mortality ratio, UN=United nations

REFERENCES

1. Graham WJ, Ahmed S, Stanton C, Abou-Zahr C, Campbell OM. Measuring maternal mortality: An overview of opportunities and options for developing countries. BMC Med 2008;6:12.
2. Chandy H, Heng YV, Samol H, Husum H. Comparing two survey methods for estimating maternal and perinatal mortality in rural Cambodia. Women Birth 2008;21:9-12.
3. Danel I, Graham W, Stupp P, Castillo P. Applying the sisterhood method for estimating maternal mortality to a health facility-based sample: A comparison with results from a household-based sample. Int J Epidemiol 1996;25:1017-22.
4. Abouzahr C, Wardlaw T. Maternal Mortality in 2000. Estimates Developed by WHO, UNICEF and UNFPA. Geneva: World Health Organization; 2004. Available from: http://www.who.int/maternal_child_adolescent/documents/9241562706/en/. [Last accessed on 2014 Apr 08].
5. Berhane Y, Andersson T, Wall S, Byass P, Högberg U. Aims, options and outcomes in measuring maternal mortality in developing societies. Acta Obstet Gynecol Scand 2000;79:968-72.
6. Hill K, El Arifeen S, Koenig M, Al-Sabir A, Jamil K, Raggers H. How should we measure maternal mortality in the developing world? A comparison of household deaths and sibling history approaches. Bull World Health Organ 2006;84:173-80.
7. Hakkert R. Country estimates of maternal mortality: An alternative model. Stat Med 2001;20:3505-24.
8. Kruglov YV, Kobyshcha YV, Salyuk T, Varetska O, Shakarishvili A, Saldanha VP. The most severe HIV epidemic in Europe: Ukraine's national HIV prevalence estimates for 2007. Sex Transm Infect 2008;84 Suppl 1:i37-41.
9. VAN Hest NA, Story A, Grant AD, Antoine D, Crofts JP, Watson JM. Record-linkage and capture-recapture analysis to estimate the incidence and completeness of reporting of tuberculosis in England 1999-2002. Epidemiol Infect 2008;136:1606-16.
10. Verlato G, Muggeo M. Capture-recapture method in the epidemiology of type 2 diabetes: A contribution from the Verona diabetes study. Diabetes Care 2000;23:759-64.
11. Gjini A, Stuart JM, George RC, Nichols T, Heyderman RS. Capture-recapture analysis and pneumococcal meningitis estimates in England. Emerg Infect Dis 2004;10:87-93.
12. Morse T, Dillon C, Warren N, Hall C, Hovey D. Capture-recapture estimation of unreported work-related musculoskeletal disorders in Connecticut. Am J Ind Med 2001;39:636-42.
13. Savva GM, Morris JK. Ascertainment and accuracy of down syndrome cases reported in congenital anomaly registers in England and Wales. Arch Dis Child Fetal Neonatal Ed 2009;94:F23-7.
14. Amoros E, Martin JL, Lafont S, Laumon B. Actual incidences of road casualties, and their injury severity, modelled from police and hospital data, France. Eur J Public Health 2008;18:360-5.
15. Khorasani Zavareh D, Mohammadi R, Laflamme L, Naghavi M, Zarei A, Haglund BJ. Estimating road traffic mortality more accurately: Use of the capture-recapture method in the West Azarbaijan province of Iran. Int J Inj Contr Saf Promot 2008;15:9-17.
16. Amin-Esmaili M, Nedjat S, Motevalian A, Rahimi-Movaghar A, Majdzadeh R. Comparison of databases for Iranian articles; access to evidence on substance abuse and addiction. Arch Iran Med 2009;12:559-65.
17. WHO, UNICEF, UNFPA, Word Bank. Maternal Mortality in 2005: Estimates developed by WHO, UNICEF, UNFPA, and the World Bank. Geneva: World Health Organization; 2007.
18. WHO, UNICEF, UNFPA, Word Bank. Trends in Maternal Mortality: 1990-2008. Geneva, Switzerland: World Health Organization; 2010.
19. Hogan MC, Foreman KJ, Naghavi M, Ahn SY, Wang M, Makela SM, *et al*. Maternal mortality for 181 countries, 1980-2008: A systematic analysis of progress towards millennium development goal 5. Lancet 2010;375:1609-23.
20. National maternal mortality surveillance system. Tehran: Ministry of Health and Medical Education; 2006.
21. Chapman DG. The estimation of biological populations. Ann Math Stat 1954;25:1-15.
22. Clark DE, Hahn DR. Comparison of probabilistic and deterministic record linkage in the development of a statewide trauma registry. In Proceedings of the annual symposium on computer application in medical care. Maryland USA: American Medical Informatics Association 1995. p. 397.
23. Silveira DP, Artmann E. Accuracy of probabilistic record linkage applied to health databases: Systematic review. Rev Saude Publica 2009;43:875-82.
24. Grijalva CG, Craig AS, Dupont WD, Bridges CB, Schrag SJ, Iwane MK, *et al*. Estimating influenza hospitalizations among children. Emerg Infect Dis 2006;12:103-9.

25. Tagliabue G, Tessandori R, Caramaschi F, Fabiano S, Maghini A, Tittarelli A, *et al*. Descriptive epidemiology of selected birth defects, areas of Lombardy, Italy, 1999. Popul Health Metr 2007;5:4.
26. Bernillon P, Lievre L, Pillonel J, Laporte A, Costagliola D. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990-1993. The Clinical Epidemiology Group from Centres d'Information et de Soins de l'Immunodéficience Humaine. Int J Epidemiol 2000;29:168-74.

Staying in touch with the journal

1) **Table of Contents (TOC) email alert**
Receive an email alert containing the TOC when a new complete issue of the journal is made available online. To register for TOC alerts go to **www.ijpm.ir/signup.asp.**

2) **RSS feeds**
Really Simple Syndication (RSS) helps you to get alerts on new publication right on your desktop without going to the journal's website. You need a software (e.g. RSSReader, Feed Demon, FeedReader, My Yahoo!, NewsGator and NewzCrawler) to get advantage of this tool. RSS feeds can also be read through FireFox or Microsoft Outlook 2007. Once any of these small (and mostly free) software is installed, add **www.ijpm.ir/rssfeed.asp** as one of the feeds.