Arch Iran Med 2009; 12 (6): 559 – 565

Original Article

Comparison of Databases for Iranian Articles; Access to Evidence on Substance Abuse and Addiction

Masoumeh Amin-Esmaili MD MPH*, Saharnaz Nedjat MD PhD**, Abbas Motevalian MD PhD***, Afarin Rahimi-Movaghar MD MPH*, Reza Majdzadeh DVM PhD•**

Background: With the expansion of scientific documents worldwide and the growth of Iranian articles published in recent years, highly sensitive electronic resources containing scientific studies can be of great help to researchers at different stages of research. This study has been conducted to examine the sensitivity of domestic and international databases to locate Iranian articles related to substance abuse disorders or addiction.

Methods: Iranian articles in three domestic databases (Iranpsych, Iranmedex, and SID) and three international databases (PsycInfo, Embase, and Medline) were compared, separately. All articles in English or Persian related to the subject in one year period (2002) were selected by two reviewers. The log linear model was applied to estimate the total number of articles and sensitivity of each database.

Results: As a whole, 129 Persian articles and 29 English articles related to addiction were found in six domestic and international databases. Among domestic databases the sensitivity of each database (Iranpsych, Iranmedex, and SID) was estimated to be 66.5, 55.7 and 40.5%, respectively. Among the international databases (PsycInfo, Embase and Medline) were estimated to have 22.1, 19.1 and 17.6% coverage, respectively. A crossover search of articles retrieved from domestic databases showed that 11.6% of the Iranian articles had been published in English in international journals.

Conclusion: Our findings suggest that after searching at least one general and one specialized Iranian database, 80% of studies relevant to addiction in Iran can be accessed. Introduction of developing countries' domestic databases can be beneficial in a more comprehensive access to scientific documentations.

Archives of Iranian Medicine, Volume 12, Number 6, 2009: 559 – 565.

Keywords: Information storage and retrieval • Iran • methods • periodicals • substance-related disorders

Introduction

Due to "information explosion", busy clinicians and health care providers find it difficult to keep themselves up-to-date. Studies show that as of 2001, over 3 billion scientific documentations have been registered on the internet, of which 2% are health-related. Data have shown that the aforementioned number doubles every 173 days.[1] The expansion of scientific evidence and documents in various medical fields in a country is a sign of the promotion of research. And fortunately, Iran has also developed by expanding its scientific researches; and between the years 1992 and 2002 Iranian articles have shown a 733% growth.[2]

The evidence based medicine movement and ever growing success of systematic reviews in introducing interventions in health necessitate the existence of highly sensitive databases that would allow access to studies and systematic reviews with the least selection bias. If we are to look at the

Authors' affiliations: *Iran Research Center of HIV/AIDS (IRCHA), Tehran University of Medical Sciences, **School of Public Health, Knowledge Utilization Research Center (KURC), Tehran University of Medical Sciences, ***School of Public Health, Iran University of Medical Sciences, Tehran, Iran.
•**Corresponding author and reprints:** Reza Majdzadeh DVM PhD, Knowledge Utilization Research Center (KURC), Tehran University of Medical Sciences, Tehran, Iran.
Telefax: +98-216-649-5859, E-mail: rezamajd@tums.ac.ir
Accepted for publication: 5 August 2009

subject from a researcher's point of view who is going to do a primary research; it is obvious that he would have to review the literature for many reasons: to avoid useless and parallel work which waste human and financial resources as well as time, to use others' experiences, to learn other studies' strengths and weaknesses, and in many cases to determine the sample size and perform the study correctly. Therefore access to relevant studies and data on a subject, especially if it is comprehensive enough, is essential in the primary stages of designing a research study.

According to a study conducted by the Ministry of Health and Medical Education's Applied Research Secretariat substance abuse disorders came in third in its burden of disease in men, preceded by traffic accidents and coronary diseases. Substance abuse disorders are among the first top ten of disease in both genders.[3,4] Access to valid data on nationwide addiction is hence required for planning and correct allocation of resources. Specific cultural and social characteristics and native variables make it a necessity to use national research. Fortunately, nowadays there are a good number of mental health publications in the country. Their data could easily be accessed if databases existed in the country.[5]

One of the ways of accessing scientific evidence throughout the world is by using electronic databases that register scientific evidence. However, this question has always kept researchers thinking: to what extent do these databases cover the concerned researches. In recent years Iran a few databases has been created. Three of them have been selected for evaluation in this study: 'Iranian Database of Medical Articles' (Iranmedex), 'Scientific Information Database' (SID), and Iran's Database of Mental Health Studies (Iranpsych).[6] A brief introduction on these databases and their characteristics at the time of this study (June to August 2007) is presented in Annex 1.

Determining the facilities of databases is important to researchers, because there are now a large number of international databases and domestic ones are developing fast. The sensitivity of a database in its access to data and access to full texts help researchers set their expectations while designing their studies, especially secondary research in order to have an estimate of the validity of their search.

In this study we have used the capture-recapture method for estimating addiction-related studies in this country. This method has been used to estimate hidden populations such as substance abusers, illegal drug abusers, prostitutes and the homeless.[7–9] The simplest type of capture-recapture model is the 'two sample model'. Here a sample is captured from a specific population, tagged, and returned to the same population, recaptured, and a second sampling is done. Using the number of cases captured in both samples an estimate is made of the total population size. Few studies have used this technique for evaluating the sensitivity of databases and examining the completeness of a systematic search.[10–12] However, none of the studies have examined Iranian databases.

In this study we have attempted to examine the sensitivity and number of addiction-related articles in domestic and international databases by the capture-recapture method. Also, three domestic and international databases have been compared separately for Iranian articles.

Materials and Methods

All articles related to the subject in English or Persian published in 2002 were searched with a certain search strategy in the three domestic databases of Iranpsych, Iranmedex and SID as well as the three international databases of PsycInfo, Embase, and Medline. The search strategy for Iranian databases included all English and Persian keywords related to substance abuse, addiction and names of drugs. For international databases, all geographical locations including name of the country, universities and major cities were searched and combined with the names of drugs, and keywords related to substance abuse. Finally, the results were limited to the period of the study. There was no limitation regarding experimental studies carried out on animals if they were related to substances. Abstracts were obtained through the internet and were sufficient in determining the subject relevance.

The articles were evaluated both for their subject relevance and the abovementioned criteria by two independent reviewers and their particulars were noted in the evidence table. The total number of relevant articles and the number of articles separately retrieved from each database were determined at this stage.

One of the important assumptions of the capture-recapture method is the independence of databases. When there are two sources then there is

no choice but to have two independent databases. In a similar study, three databases were included the log linear model can be applied to enter more than two sources as well as the dependency of the data sources. Log-linear is a model that adjusts more than two cases taking into consideration their dependency by including interaction terms.[13] In this way no bias will be created in the estimates if the prerequisite of source independence does not exist.

As a result, this model can be applied to estimate articles not displayed in any of the databases (X in Tables 1 – 3 of the results) and also to the estimate of the total number of eligible studies (N in Tables 1 – 3). Akaike's information criterion (AIC) is the main criteria for model selection in log-linear. The best model is indicated by the smallest AIC.[14]

Results

Upon examining abstracts and deleting repetitive articles in each database, a total of 129 and 29 addiction-related articles in Persian and English respectively, were found in all six domestic and international databases. In the domestic databases, 129 Persian and 11 English articles were found. The total number of Iranian articles published in English was 29; 27 of which were published in non-Iranian journals and 2 published in English Iranian journals, respectively.

It is worth mentioning here that Iranpsych and SID respectively displayed 35 and 14 Iranian articles published in non-Iranian journals; but Iranmedex only displayed articles published in Iranian journals that were both in Persian and English. Since our objective was to compare databases that displayed Iranian journals' articles; the articles published in international journals were not considered in the evaluation of Iranian databases sensitivity.

The numbers of addiction-related articles recaptured separately for each database was as follows: 105 from Iranpsych, 88 from Iranmedex, 64 from SID, 15 from Medline, 13 from PsychInfo and 12 from Embase.

A comparison of unique articles in the databases showed that among domestic and international databases, Iranpysch and Medline contained more unique articles as compared to the other databases (Table 1). Unique articles are those articles which are exclusively recaptured from the same database and not appeared in other databases. Diagrams 1 and 2 demonstrate the coverage and overlap of Iranian and international databases with respect to addiction related articles. The log linear model results have been shown in Tables 2 and 3. Among domestic databases the model with the lowest AIC and the highest fit, was a model that had an interaction with Iranmedex-SID; the number of addiction-related articles that haven't been displayed in any of the three domestic databases would therefore be 18. The total number of studies related to addiction in a year was therefore estimated to be 158 (CI95%: 152 – 190). It should be kept in mind that 158 is not the number of articles observed, but the total estimate of possible eligible studies related to the subject including those which were found by three databases as well as those that were hidden and not captured by them. On these grounds, the sensitivity of the three databases of Iranpsych, Iranmedex and SID were estimated to be 66.5, 55.7 and 40.5% respectively. Based on the findings of the current study, searching Iranpsych alone will yield 66.5% of the country's Persian addiction-related articles. If both Iranpsych and Iranmedex are searched then the sensitivity will reach 85% (95%CI: 78 – 90). Adding SID to the search will only raise the sensitivity of the search by 4%, i.e. to 89% (95%CI: 83 – 93). Here, we must keep in mind that the articles related to addiction in Iran and published in non-Iranian journals were only displayed in Iranpsych and SID and therefore

Table 1. Frequency of total and unique retrieved articles relevant to addiction in Iran in six domestic and international databases*

Database	Number of Persian articles	Number of English articles	Total number of articles	Number of unique articles
Iranpsych	97	8	105	35
Iranmedex	82	6	88	17
SID	58	6	64	8
Medline	—	15	15	9
PsycInfo	—	13	13	4
Embase	—	12	12	2

*Since searches within domestic and international databases were not done for a similar period of time (1381 Solar Hegira and 2002 Christian calendar), comparison between domestic and international databases based on data of this table is not applicable

Table 2. Characteristics of log linear fitness models in three domestic databases in 1381(Solar Hegira calendar)

Log linear model	df*	G2*	AIC*	X★	N✦	CI95% for N
IM*, IP*, SID*	3	14.69	8.69	11	151	147–158
IM×IP, SID	2	13.67	9.67	8	148	143–159
IM×SID, IP	**2**	**2.55**	**-1.45**	**18**	**158**	**150 – 170**
SID×IP, IM	2	14.32	10.32	12	152	147 – 163
IM×SID, SID×IP	1	0.57	-1.43	25	165	152 – 190
IM×SID, IM×IP	1	2.52	0.52	19	159	147 – 194
IM×IP, SID×IP	1	13.61	11.61	9	149	143 – 167
IM×SID, IM×IP, SID×IP	0	0	0	39	179	150 – 297

*df=degree of freedom; AIC=akaike information criterion; BIC=Bayesian Information Criterion; IM=Iranmedex; IP=Iranpsych; SID=Scientific Information database; G2=AIC = G2 – 2 (d.f.)

★ An estimate of articles not displayed in any of the domestic databases

✦ An estimate of the total number of addiction-related articles

omitted from the calculations.

Among international databases the model with the lowest AIC and the highest fitness was a model that had an interaction with Embase-PsychInfo; the number of addiction-related articles that were not displayed in any of the three databases would therefore be 39. The total number of studies related to addiction in 2002 was therefore estimated to be 68 (CI95%: 40 – 165). On these grounds, the coverage of each of the databases of Medline, PsychInfo and Embase were estimated to be 22.1, 19.1, and 17.6%, respectively. A simultaneous search in Medline and PsychInfo had a 37% (95%CI: 25 – 49) sensitivity. Adding Embase to this search improved the sensitivity by only 6%, i.e. to 43% (95%CI: 31 – 55).

Upon comparison of two domestic and international databases i.e. Iranpsych and Medline respectively), we discovered that out of 129 Persian addiction-related articles published in 1381 in Persian-Iranian journals; 15 (11.6%) English versions of the articles had been indexed in Medline. For this purpose we expanded the search year and did not limit the search to 2002.

Discussion

The current study is the first of its kind to compare Iranian databases and estimate the total number of substance-abuse related articles by the capture-recapture method. Though similar studies elsewhere have examined other subjects with different methods in databases and compared their coverage on that specific subject.[15] Among these are a simple descriptive comparison of retrieved and unique articles, and determination of the recall index of seven databases (Medline, CINAHL, Caredata, PsycLIT, Cochrane Library, Sociofile and Social Science Citation Index) on the subject of rehabilitation of individuals with severe mental disorders in Brettle et al.'s study,[14] or comparison of various methods of evaluating publication bias namely Funnel plot and capture-recapture in three electronic search sources, manual/hand searching and search by experts in Bennett et al.'s study.[16]

Based on the findings of this study, a total of 158 articles related to addiction in 2002 were displayed in six domestic and international databases. The Iranian articles in Persian and English were examined separately in the three domestic databases of Iranpsych, Iranmedex, and SID. Likewise articles in non-Iranian journals were separately examined in the three international databases of Medline, PsychInfo and Embase. It was estimated that 18 Iranian articles in Persian

Table 3. Characteristics of log linear fitness models in three international databases in 2002

Log linear model	df*	G2*	AIC*	X★	N✦	CI95% for N
ML*, EB*, PI*	3	8.47	2.47	14	43	35 – 67
ML×EB, PI	2	7.48	3.48	10	39	32 – 63
ML×PI, EB	2	6.78	2.78	9	38	31 – 60
PI×EB, ML	**2**	**0.28**	**-3.72**	**39**	**68**	**40 – 165**
ML×PI, PI×EB	1	0.26	-1.74	44	73	34 – 455
ML×PI, ML×EB	1	4.31	2.31	4	33	30 – 48
ML×EB, PI×EB	1	0.12	-1.88	55	84	35 – 539
ML×PI, ML×EB, PI×EB	0	0	0	88	117	32 – 3027

*df=degree of freedom; AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion; G2= AIC = G2 – 2 (d.f.); ML=Medline; EB= Embase; PI=PsycInfo

★An estimate of articles not displayed in any of the international databases.

✦An estimate of the total number of addiction-related articles.

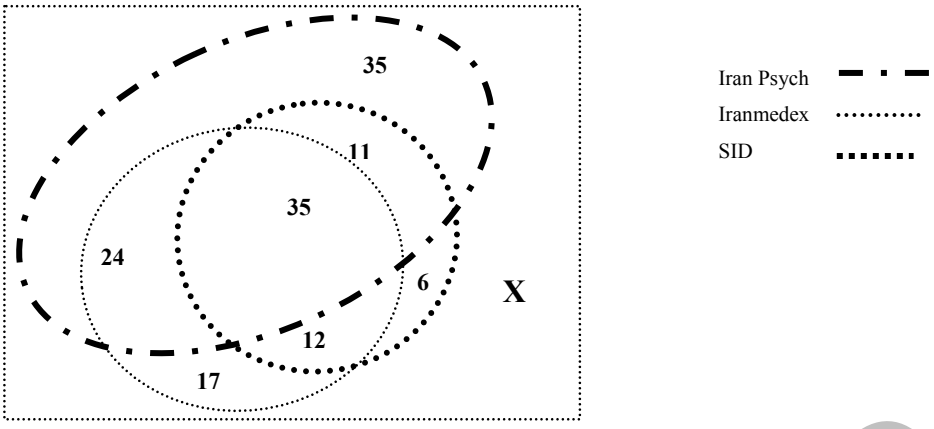

Diagram 1. The Venn diagram shows the number of addiction-related articles in Iranian journals published in 1381 (Solar Hegira calendar) and displayed in three domestic databases

and English were not seen in any of the domestic databases. Similarly, the number of Iranian addiction related articles in 2002 that were not displayed in any of the international databases were 39.

Access to Iranian articles published in non-Iranian journals indexed in Iranpsych and Iranmedex confirmed our estimations of the international databases. The present study revealed that a total of 37 Iranian articles published in non-Iranian journals were indexed in Iranmedex and Iranpsych that had not been displayed in any of the concerned international databases. The other reason behind this rationale is McDonald et al.'s study[16] in which the number of psychiatry journals displayed in Medline, PsycLIT, Embase and Biosis had been compared. There the results showed 52% of psychiatry journals had been displayed in these four databases. Researchers were advised to search more than two databases to ensure a complete coverage in search because of the overlap in databases.

A cross-sectional search of recaptured articles from domestic databases revealed that 11.6 % of Iranian studies had been published in English in non-Iranian journals. Researchers and especially those doing systematic reviews must take this matter into account.

The four assumptions of capture-recapture will help us more in ascertaining the authenticity of the results. The first assumption of capture-recapture is that the population is closed. The sources currently used in this study are databases containing articles. No doubt databases are not confined systems and the interval between updates, especially in international databases is short (e.g. in Medline updates are weekly), though this duration is longer in domestic databases. However, in order to observe this principle in our study, the publication date was limited in the search. That is why the year

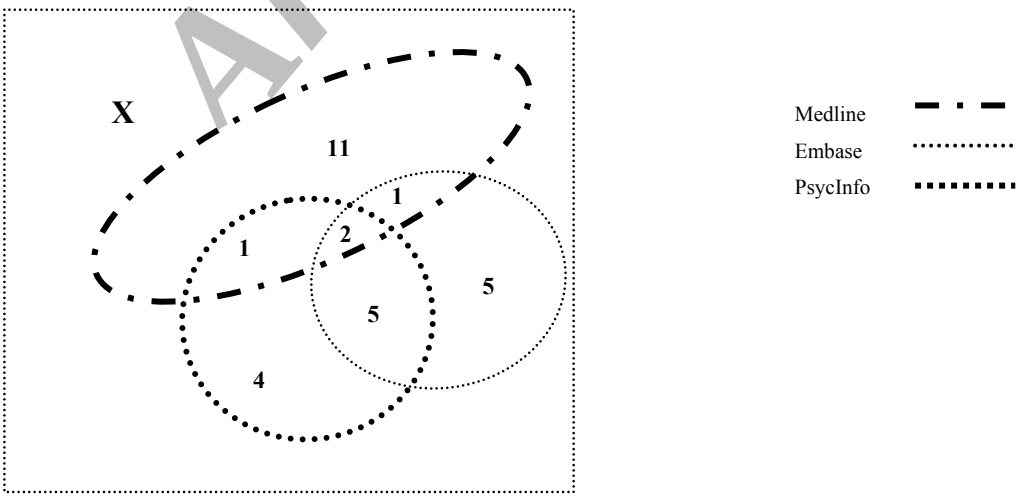

Diagram 2. The Venn diagram shows the number of addiction-related articles in international journals published in 2002 and displayed in three international databases

1381 (Solar Hegira calendar for Iranian databases) and 2002 (Christian calendar for international databases) was chosen, so that by the time of the study, i.e. 2008, the process of entering articles into the bank had been completed.

The second assumption is the ability to identify common cases among the lists. Since all the particulars used to identify common cases (including authors' names, journal title, article title, abstract and publication date) were present in all these lists, this principle has been observed, too. The possibility of erring at the time of registering the study's particulars has also been overcome by examining the abstracts for possessing the inclusion criteria.

In the third assumption we should have independent lists or the capture in each occasion should be independent. It was possible to find the same study in more than one database, [it seems that at times some Iranian databases find and index articles from other resources]. Although the third assumption of independent lists was not followed but the results were not affected because three sources were present for each domestic and international database, and the log linear model was used.

The fourth assumption is the dependency of capture on individual characteristics. The characteristics of each study such as title, abstract and keywords were the same in every database they were displayed in. The only thing that sets the databases apart from each other are their search engines, though this problem is seen less in international databases. Advanced search, search of keywords and MeSH terms are possible with little differences in the three databases of Medline, PsycInfo and Embase. On the other hand, there were multiple problems in searching Iranian databases. Problems such as 'absence of an advanced search at the time of this study', 'lack of simultaneous acceptance of Boolean operators of "and", "or", "without", and use of two-word keywords in parentheses', and 'not being able to use multiple keywords simultaneously' were seen in all three databases. Since the possibility of recapturing articles in different databases was heterogeneous, we even had to enter the keywords one by one and study the articles visually in order to put the repetitive cases aside. This was done to ensure the sensitivity of our search.

Apart from the fact that addiction is a mental disorder and is closely related to other clinical diseases, it also has social and humanistic dimensions. If all the dimensions of addiction-related articles are to be registered, then all journals related to social and human sciences, sociology and education need to be included, too. Iranspych gives the most coverage for clinical, para-clinical, and social and human science journals; the total number of articles and the number of unique articles registered in this bank also support this observation.

Our findings suggest that by searching at least one general and one specialized Iranian database, over 80% of addiction-related studies of the country can be accessed. Researchers doing a systematic review however, definitely need comprehensive sources of high sensitivity to find documents related to their subject. Our findings suggest that researchers should consider the limitations of domestic and international databases and search at least two databases in each category to increase the sensitivity of their search strategy.

This study is representative of the fact that many systematic reviews conducted worldwide that confine themselves to searching international databases miss many of the developing countries' documents such as Iran.

Domestic data banks' authorities should introduce themselves at an international level to give other researchers the chance to more exclusively use domestic study findings.

Acknowledgment

This study is a brief presentation of the first author's MPH thesis, at School of Public Health, Tehran University of Medical Sciences.

References

1 Powell J, Clarke A. The WWW of the World Wide Web: Who, What, and Why? *J Med Internet Res*. 2002**; 4:** e4.

2 Rezaei-Ghaleh N, Azizi F. The impact factor-based quality assessment of biomedical research institutes in Iran: Effect of impact factor normalization by subject. *Arch Iran Med*. 2007; **10:** 182 – 189.

3 Ministry of Health and Medical Education. *National Study of Burden of Disease and Pathologies; Burden of Health Risk Factors and Health Associated Life Expectancy in Islamic Republic of Iran in 2003 at National and Provincial Level (Six Provinces)*. Tehran: MOHME Deputy of Health; 2007.

4 Mahdavi N. *Determining the Burden of Mental Disorders in Three Provinces of Eastern Azerbaijan, Bushehr and Chaharmahal-o-Bakhtiari in 2002. Seventh National Conference on Psychiatric and Psychological Research in Iran; 2005*. Tehran: Shaheed Beheshti University of Medical Sciences; 2005: 99 – 101.

5 Rahimi-Movaghar A, Sharifi V, Mohammadi MR, Farhoudian A, Sahimi Izadian E, RadGoodarzi R, et al. Researches on substance use in Iran; 3 decades evaluation [in Persian]. *Hakim Res J*. 2006; **8:** 37 – 44.

6 Rahimi-Movaghar A, Mohammadi MR, Sahimi Izadian E, Mansouri N, Sharifi V. Iranian Database of research articles on mental health: IranPsych [in Persian]. *Social Welfare Quarterly*. 2004; **4:** 248 – 253.

7 Bloor M, Leyland A, Barnard M, McKeganey N. Estimating hidden populations: a new method of calculating the prevalence of drug-injecting and non-injecting female street prostitution. *Br J Addict*. 1991; **86:** 1477 – 1483.

8 Berry B. A repeated observation approach for estimating the street homeless population. *Eval Rev*. 2007; **31:** 166 – 199.

9 Gurgel RQ, da Fonseca JD, Neyra-Castaneda D, Gill GV, Cuevas LE. Capture-recapture to estimate the number of street children in a city in Brazil. *Arch Dis Child*. 2004; **89:** 222 – 224.

10 Spoor P, Airey M, Bennett C, Greensill J, Williams R. Use of the capture-recapture technique to evaluate the completeness of systematic literature searches. *BMJ*. 1996; **313:** 342 – 343.

11 Motevalian SA, Holakouie Naieni K, Mahmoodi M, Majdzadeh R, Akbari ME. Estimation of death due to road traffic injuries in Kerman district: application of capture-recapture method [in Persian]. *J School Pub Health & Inst Pub Health Research*. 2007; **2:** 61 – 72.

12 Brettle AJ, Long AF. Comparison of bibliographic databases for information on the rehabilitation of people with severe mental illness. *Bull Med Libr Assoc*. 2001; **89:** 353 – 362.

13 Chao A, Tsay PK, Lin SH, Shau WY, Chao DY. The applications of capture-recapture models to epidemiological data. *Stat Med*. 2001; **20:** 3123 – 3157.

14 Hook EB, Regal RR. Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. *Am J Epidemiol*. 1997; **145:** 1138 – 1144.

15 Bennett DA, Latham NK, Stretton C, Anderson CS. Capture-recapture is a potentially useful method for assessing publication bias. *J Clin Epidemiol*. 2004; **57:** 349 – 357.

16 McDonald S, Taylor L, Adams C. Searching the right database. A comparison of four databases for psychiatry journals. *Health Libr Rev*. 1999; **16:** 151 – 156.

Annex 1. Brief explanation of Iranian databases:

- 'Iranpsych' was run in March 2004 by the 'National Medical Research Center'. This database is a site specifically devoted to studies published on mental health and related to Iranian populations and/or research conducted in Iran in domestic and international journals and conference proceedings. At the time of this study there were almost 5000 and 1400 articles indexed in the Persian and English languages respectively. (http://iranpsych.tumc.ac.ir)
- 'Scientific Information Database' (SID) has been established in August 2004 by 'Jahad-e-Daneshgahi' which is a non-profit and non-governmental organization. It has a wide range of subject coverage including health, humanities, social sciences, engineering, agriculture and basic sciences. SID serves both Persian and English articles and more than 170,000 records were available in this database at the time of searching for this study. (www.sid.ir)
- 'Iranian Database of Medical Articles' (Iranmedex) was run in June 2004 by a private sector. More than 39,500 health related articles from Iranian journals in both Persian and English were covered by this database. (www.iranmedex.com)