

# Towards Implementation of Ethical Principles: Privacy and Confidentiality in Medical Data Mining—Indian Scenario

Ashwinkumar. U. M<sup>1,\*</sup>, Anandakumar. K. R<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Reva Institute of Technology and Management, 560064, Bangalore

<sup>2</sup>Department of Computer Science and engineering, S J B Institute of Technology, 560060, Banaglore

**Abstract** Datasets constitute a fundamental part of any medical research. Medical data sets (ie, medical records) are generally noted to be large, complex, heterogeneous, hierarchical, time related, and of varying quality. Such fragmented and distributed data sets challenge efficient data integration and interpretation. Application of data mining techniques for healthcare related data will be highly rewarding. Lack of domain knowledge within healthcare providers calls for active collaboration between the doctor and data miner. Ethical concerns, especially issues of consent and confidentiality have resulted in the introduction of stringent regulations in doing this form of research in India. However we found that only 16% of the reviewed publications in medical data mining within India reported of having undergone ethical review. This suggests that the ethical considerations while practicing data mining is very much undermined. It is mandatory to obtain ethical and legal clearance from regional as well as Institutional Ethics Review Board (IERB), before using data mining tools in health care research.

**Keywords** Ethics, Ethical Review Board, Data Mining, Healthcare Research

## 1. Introduction

Progress in the field of modern medical science has been attributed largely to newer methods of scientific research. Mostly doctors and researchers rely on better data collection and interpretation. Reviews of medical datasets and resulting publication thereof are always done without revealing patients identity. However debate exists about the need to obtain informed consent from patients when their identities must be revealed[1]. The use of medical datasets either takes the form of systematic record review or record linkage[2].

Systematic record review is applied when records of consecutive series of patients with same diagnosis are reviewed to identify common clinical features, response to treatment, or factors influencing prognosis. Such retrospective analysis of records forms the basis for majority of the scientific publications by medical community within India and abroad[2].

Record linkage is applied when medical information from separate sources on an individual patient needs to be collated, to identify potential association between drug and disease. Record linkage necessitates sharing of information

pertaining to personal identity such as name, date of birth and other details. This entails a greater risk of loss of confidentiality. However record linkage review is not used often in India. This probably explains the lack of concern about the use of medical records for research amongst general public in India.

## 2. Biomedical Research in India- current Scenario

The existing scene for ethical consideration in biomedical research involving patient information and the uniqueness of medical data mining are discussed in the following paragraphs.

### 2.1. Ethical, Legal and Social Issues

Numerous surveys outside India has revealed that patients are willing to support and participate in research provided they are taken through a well informed consent. Patients are mostly worried whether their data is used for marketing and insurance purposes. They are also concerned that sensitive information could be widely used and distributed without their knowledge.[3, 4, 5, 6, 7]. These concerns have led to international efforts for enhancing the protection afforded to medical records. In United States, the health insurance portability and accountability act (HIPPA)[8, 9] directs the secretary of Health and human services to establish

\* Corresponding author:

ashu12330@rediffmail.com (Ashwinkumar. U. M)

Published online at <http://journal.sapub.org/ajbe>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

safeguards for the privacy of individually identifiable health information. A variety of federal legislative proposals has also been developed to address the issue. The European commission has proposed in its draft directive that explicit patient consent should be obtained before each record can be used for research. Such stringent rules may halt record based research all together. The guidelines proposed by the United Kingdom's department of health and British Medical Association, are less stringent nonetheless restrictive[3, 4, 5].

## **2.2. Current Practice of Medical Records Review and Publication of Data in Research: Indian Scenario**

Until recently, most Indian investigators could get retrospective analysis published without an ethics review, as most international journals did not insist on such clearance previously. Currently the IERB clearance has been made mandatory by Indian Council of Medical Research (ICMR), even for a retrospective case record analysis. However ICMR allows a waiver for informed consent if the study is of minimal risk or conducted in a situation of emergency. The Medical council of India's (MCI) code of medical ethics[10] also permits such waivers if the patient identity is not revealed. Nevertheless all proposals must be cleared by ERBs in a formal meeting. Adding to the existing issues, very few IERBs exist within the country, thus slowing down the process of research.[11]

## **2.3. Practical Argument Against Regulatory Considerations Related To Obtaining Consent**

Obtaining consent from patients either through direct or indirect contact may introduce bias into research process. It may also constitute a breach of privacy. Such contact especially in a setting of retrospective study may cause psychological, social or other harm to the former patient. Undue hardship may be imposed on an organization when additional financial, material, human or other resources are required.

## **2.4. How do these Regulations Affect Indian Doctors and Researchers?**

Making it mandatory for researchers to obtain explicit consent from patients before assessing their medical records, as proposed by European commission, would prevent clinical studies that rely on personal records, with the exception of small case series. In US, HIPPA regulations appear to have reduced medical record and database based research[8, 9]. Current HIPPA implementation strategies will increase the workload for ERBs and researchers; this may result in increased dropout rate for proposed studies as investigators may not meet the requirements[8, 9]. Researchers also feel that public money from government agencies and charitable organizations is wasted by ERBs when innocuous retrospective studies are required to go through multiple ethical reviews[12]. The majority of publications from Indian institutes are related to medical records review. Only a few major institutes have ERBs and

most of this form of research is not subjected to ethics review.

## **3. Privacy and Security of Human Data**

For instance, U. S. federal rules set guidelines for concealment of individual patient identifiers. At stake is not only a potential breach of patient confidentiality, with the possibility of ensuing legal action, but also erosion of the physician-patient-relationship, in which the patient is extraordinarily candid with the physician in the expectation that such private information will never be made public. A related privacy issue may apply if, for example, crucial diagnostic information were to be discovered on patient data, and a patient could be treated if only one could go back and inform the patient about the diagnosis and possible cure. In some cases, this action may not be taken. Another issue is data security in data handling, and particularly in data transfer. Since transferring the data electronically via internet is insecure, the identifiers must be carefully concealed even for transfers within a single medical institution from one unit to another. So one uses de-identifying or concealing of medical data and guidelines suggest concealment of identifiers should be irreversible. Before the identifiers are concealed, only authorized persons should have access to the data. On the other hand, it has been noted in recent U. S. federal documents (U.S. 1999), that there are at least two legitimate research needs for re-identification of de-identified medical data: first, there is a need to prevent accidental duplicate records on the same patient from skewing research conclusions; second, there may be a compelling need to refer to original (re-identified) medical records to verify the correctness or to obtain additional information on specific patients. These special requirements could be managed by appropriate regulatory agencies, but they could not be met at all if the data are completely anonymous. There are four forms of patient data identification:

### **3.1. Anonymous Data**

Those data sets that had the patient identification detail removed right at the time of collection is called as anonymous data. For example, a block of tissue may be taken from an autopsy on a patient with a certain disease, to serve as control tissue-block in the histology laboratory. The patient's identifiers are not recorded at the time of specimen collection, and thus can never be recovered.

#### **3.1.1. Anonymized Data**

Medical data sets where the patient-identifiers that were collected initially are subsequently irrevocably removed are known as anonymized data. That is, there can never be a possibility of returning to the patient's record and obtaining additional information. This research practice has been common in the past. However, anonymized data, as described above, could be accidentally duplicated, and

cannot be verified for correction or additional data.

### 3.1.2. De-identified Data

Under de-identified data, data that are collected initially with the patient-identifiers are subsequently encoded or encrypted. The patient can be re-identified under conditions stipulated by an appropriate agency, typically an Institutional Review Board (IRB)

### 3.1.3. Identified Data

Identified data can only be collected after obtaining a fully informed written consent by the patient. Also it is mandatory that both the research protocol and consent document have been reviewed by the IERB. A variety of encryption protocols for each patient's data, suitable for specific purposes are available. It is worth to note that data hosted on a public domain is most secure only when it involves single entry of data. Multiple or serial update of data sets is less secure.

In summary, the de-identified medical data seems to carry the least risk, and is most commonly employed for medical research in US. Studies utilizing "de-identified data" employ data that are collected in the ordinary diagnosis and treatment of patients. There shall be no change in patient management as a result of the research, including no pressure on the patient to accept or refuse certain management, and no call-back for additional data that might upset the patient or next -of-kin. Nevertheless even this data set has the risk of loss of confidentiality to the patient. This is called minimal risk data.

### 3.2. Expected Benefits

Irrespective of the source and type of medical data set, even de-identified, used for data mining in medical research must justify potential benefit to the population at large. Legally and ethically one cannot perform data analysis for frivolous or nefarious purposes. Thus having a locally established IERB plays a significant role in ensuring that any medical research involving data mining is ethical and legally justifiable. And potentially contributes to the society at large.

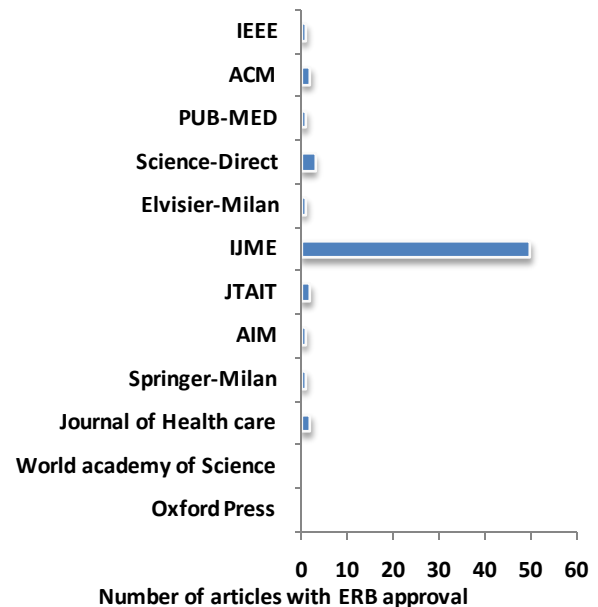
Now how do we give an evidence of this?

## 4. Methodology

We searched all major bibliographic databases and several specialist datasets during the last quarter of 2009 (see table A for search criteria). Under our primary exclusion criteria we excluded duplicate publications, editorials, letters, poster presentations, and single variable databases (such as, prescribing, disease registers). We searched for citations of papers that used a reference standard for assessment of quality. When relevance was ambiguous (for example, if we were unable to deduce whether the study involved electronic patient records or paper records) we checked the abstract and MeSH headings through PubMed. When ambiguity remained we obtained the full paper and made a collective

decision. Please refer table 1.

## 5. Results



**Figure 1.** Number of articles published on Data mining that report IERB approval

The search revealed 400 hits out of which 64 fitted our criteria based on the information needed. We found that only 16% of the reviewed publications in medical data mining reported of having undergone ethical review. This suggests that the ethical considerations while practicing data mining is very much undermined. Steps must be taken to ensure a mandatory requirement of Ethical Review Board consideration to be part of the Medical data mining research.

**Table 1.** Framework for assessing eligibility of publications for review

All three categories (A-C) needed to be satisfied for a paper to be selected	
A. Time period	Published between January 1 <sup>st</sup> 2009 to June 30 <sup>th</sup> 2009.
B. key word	Medical data mining, ethics review board, India.
C. Patient record types	Electronic patient records or Paper records.

## 6. Conclusions

Data mining in medicine is distinct from that in other fields, because the data is noted to be heterogeneous and of varying quality. In addition ethical, legal, and social constraints are applicable to medical information unlike data sets that involve physical sciences. The ethical, legal, and social limitations on medical data mining relate to privacy and security considerations, fear of lawsuits, and the need to balance the expected benefits of research against any

inconvenience or possible injury to the patient.

### 6.1. Suggestion and Recommendations

There are no reported surveys within India, to the author's knowledge that has studied the views of patients regarding use of personal data for research. However issues of confidentiality are likely to gain importance with wider insurance coverage. The investigator should anticipate this plan for the future. The ICMR guidelines allow ERBs to waive informed consent in appropriate cases where the study carries only minimal risk or in cases of emergency. However the guidelines should also provide allowances for expedited reviewer exemption from the review process. Study proposals involving medical records would be included under this category of Review. The ICMR should resist the move to universalize the new set of stringent guidelines proposed by the European commission. It would be ideal for India to adopt guidelines of the working group where the ERBs are responsible for assessing the potential importance of research proposal and deciding whether or not waive the requirement for informed consent. Circumstances under which ERBs may opt to do this include the following situations.

Access to the clinical record is essential for completion of research and consent is not applicable

The research is likely to yield information of sufficient merit

The research pertains to some future planning, preventive or therapeutic initiatives which may benefit the patients whose records are studied.

Researchers who are non-clinicians are formally instructed about their duty of confidentiality and they enlist a clinical supervisor who formally accepts professional responsibility for any breach of confidentiality, should it occur. Excessive restrictions on access to medical data for research could harm large number of people and hamper in medical care. A consensus policy respecting the rights of individuals and responsibilities of investigators are needed in India

Data mining has become an integral part of health care delivery, planning and management. There have been many studies reporting various data mining models and their effectiveness in managing huge medical database. These studies in general use existing patient information. However there seems to be a breach in that, the studies which had access to crucial patient information, have not undergone any ethical review process. There is a need for the ethical consideration with prompt review process before any study

is undertaken within the realm of data mining research.

## REFERENCES

- [1] Doyal Len. Informed consent in medical Research: Journals should not publish Research to which patients have not given Fully informed consent-with three exemptions. *BMJ* 1997; 314:1107
- [2] Wald Nicholas, Law Macolm, Meade, Tom, Miller George, Alberman Eva, Dickinson, John. Use of personal Medical Records for Research purposes. *BMJ* 1994;309:1422-1424.
- [3] Parkes S E. legal aspects of record based medical research. *Arch Dischild* 2001;89:899- 901.
- [4] Taube Daniel O, Burkhardt S. Ethical and legal risks associated with archival research. *Ethics Behav* 1997;7:59-67.
- [5] Baker R, Shiels C, Stevenson K, Fraser R, Stone M. What proportion of patients refuse consent to data collection from their records for research purposes. *Br J Gen Pract* 2000; 50(457):655-6.
- [6] Willison Donald J, Keshavjee Karim, Nair Kalpana, Goldsmith Charlie, Holbrook Anne M. Patients consent preferences for research uses information in electronic medical records: interview and survey data. *BMJ* 2003;15;326:373.
- [7] Robling M R, Hood K, Houston H, Pill R, Fay J and Evans H M. Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study. *J Med Ethics* 2004;30:104-109.
- [8] Kulynych Jennifer, Korn David. The new HIPAA (Health Insurance Portability and Accountability Act of 1996) Medical privacy rule: help or hindrance for clinical research. *Circulation* 2003;108:912-4.
- [9] O'Herrin Jacquelyn K, Fost Norman, Kudsk Kenneth A. Health Insurance Portability Accountability Act (HIPAA) Regulations; effect on medical record research. *Ann surg* 2004;239(6); discussion 776-8
- [10] Medical Council of India. Indian Medical Council (Professional Conduct, Etiquette and Ethics) Regulations, 2002. Gazette of India dated 06-04-02, part III, Section 4. {cited 2006 March 25} Available from: <http://mohfw.nic.in/code.htm>
- [11] Nundy Samiran, Gulhati Chandra M. A new colonialism: conducting clinical trials in India. *NEJM* 2005;352:1633-1636.
- [12] Wagner Richard M. Ethical review of research involving human subjects: When and why is IRB necessary. *Muscle nerve* 2003;28:27-39.