Medical record search engines, using pseudonymised patient identity: An alternative to centralised medical records

Catherine Quantin\textsuperscript{a,b,*}, David-Olivier Jaquet-Chiffelle\textsuperscript{c}, Gouenou Coatrieux\textsuperscript{d}, Eric Benzenine\textsuperscript{b}, François-André Allaert\textsuperscript{b,e}

\textsuperscript{a} Inserm, U866, Dijon, F-21000, Univ de Bourgogne, Dijon F-21000, France
\textsuperscript{b} CHRU, Service de Biostatistique et d’Informatique Médicale, Dijon, France
\textsuperscript{c} Bern University of Applied Sciences et Université de Lausanne, Switzerland
\textsuperscript{d} Institut TELECOM; TELECOM Bretagne; Unité INSERM 650 LaTIM, France
\textsuperscript{e} Ceren Esc Dijon & Dpt biostat Ecole de Santé Publique, Liège, Belgium

\textbf{Article info}

Article history:
Received 2 August 2010
Received in revised form 8 September 2010
Accepted 2 October 2010

Keywords:
Medical record
Patient identifier
Direct access
Data security
Privacy
E-health

\textbf{Abstract}

Purpose: The purpose of our multidisciplinary study was to define a pragmatic and secure alternative to the creation of a national centralised medical record which could gather together the different parts of the medical record of a patient scattered in the different hospitals where he was hospitalised without any risk of breaching confidentiality.

Methods: We first analyse the reasons for the failure and the dangers of centralisation (i.e. difficulty to define a European patients’ identifier, to reach a common standard for the contents of the medical record, for data protection) and then propose an alternative that uses the existing available data on the basis that setting up a safe though imperfect system could be better than continuing a quest for a mythical perfect information system that we have still not found after a search that has lasted two decades.

Results: We describe the functioning of Medical Record Search Engines (MRSEs), using pseudonymisation of patients’ identity. The MRSE will be able to retrieve and to provide upon an MD’s request all the available information concerning a patient who has been hospitalised in different hospitals without ever having access to the patient’s identity. The drawback of this system is that the medical practitioner then has to read all of the information and to create his own synthesis and eventually to reject extra data.

Conclusions: Faced with the difficulties and the risks of setting up a centralised medical record system, a system that gathers all of the available information concerning a patient could be of great interest. This low-cost pragmatic alternative which could be developed quickly should be taken into consideration by health authorities.

© 2010 Elsevier Ireland Ltd. All rights reserved.

* Corresponding author at: Centre Hospitalier Universitaire, Service de Biostatistique et Informatique Medicale, 1 Boulevard Jeanne d’Arc, BP 77908, 21079 Dijon, France. Tel.: +33 380 29 36 29; fax: +33 380 29 39 73.
E-mail address: Catherine.quantin@chu-dijon.fr (C. Quantin).

1386-5056/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2010.10.003
1. Introduction

For more than 20 years now, many research projects have been conducted on a standardised, centralised, secure and reliable medical record (MR) system, but they have still not met with success. The French DMP project to implement personal MRs for each patient that are accessible to the patient is an illustrative example. The DMP has encountered many difficulties regarding ethical and legal aspects, the definition of a common identifier and centralised storage of all records. We are not aware of a country that has successfully implemented a standardised, centralised, secured, privacy-compliant and reliable medical record system at the national level. Thus it is time to develop a new strategy based on a pragmatic, secure, non-centralised, unstructured MR system which will be operational in the very short term. The main goal of this article is to promote this non-centralised and non-standardised MR system, which is based on an original system to search for and gain access to distributed medical data like the one that exists in Israel (Clalit HMO and government hospitals), Pittsburg (Pennsylvania – UPMC) [1] and is being implemented in Brussels (IRIS hospitals) [2] and Franche Comte, France (EMOSYST) [3]. In these examples, all focusing on the sharing of medical data, MRs are not standardised but can be structured or unstructured. However, the sharing of medical data is standardised and structured.

2. Planned standardised MR system: the reasons for the failure and the dangers of centralisation

The main reasons for the failure are related first to insufficient human and financial resources, second to the lack of or failure to properly deploy a unique patient identifier (UPI), third to the lack of standardisation or structuring of the MRs.

Many solutions concerning these aspects are being developed to provide for example, an Enterprise Master Patient Index and standards have been proposed [4]. In practice harmonisation of patient identification is very difficult to achieve in many countries, and re-indexing of previously stored old medical data has to be managed. Facing these national difficulties, the current strategy at the European level, is more to let each country define its own identification policy and to encourage the interoperability of national information systems [5]. As a consequence, a pragmatic solution, relying on existing data and not requiring a UPI, seems to be more appropriate at the European level. The approach of the new DMP project in France, using a UPI which is to be implemented immediately, is in progress in this respect.

Regarding standardisation and structuring of the MR system, a lot of time has been wasted to define a unique format for all doctors and all pathologies. The only domain where real harmonisation has been obtained is “coding”, as with the Systematised Nomenclature of Medicine clinical terms (SNOMED – this terminology is widely accepted, but is just starting to be implemented) and the International Classification of Diseases (ICD) used to assess hospital activities. However, though ICD codes are now included within MRs to record the activities of health-care establishments, they are not being used for the daily management of patients’ MRs in all European countries. The same remark can be made regarding patients’ drug treatments, which is key information, and for which there is the recognised International Common Denomination (ICoD) used by the pharmaceutical industry.

The dangers of centralising the standardised MRs can be summarised mainly as drawbacks and management of access difficulties. For many years, the public authorities have understood the danger of a centralised system, notably the considerable risk of losing all of the data if the centralised organisation is destroyed. After realising the weaknesses of a centralised system, the USA ministry of defence created in 1969 the ARPANET, a network system that would continue to function in the case of a catastrophe. The USA has accepted the idea that a network has the advantage of being able to preserve functioning structures, even after major destruction, and that new systems could be built on the remains of damaged ones. Concerning health data, it would also be safer to store it in different places to ensure protection of both privacy and information. Furthermore, hackers may see a centralised system as a challenge and try to gain access to the central patient MR system and modify patients’ medical information. It could also be a target for terrorists who wish to destabilise a country by pirating its health system and divulging health information on citizens. Secure management of access to a centralised system will be very difficult to obtain. In addition, it raises problems related to access rights, rights to read, and to write or modify data related to the patient. In a decentralised system, a doctor who has made the effort to reconstitute and synthesise his patient’s medical history will not need to connect to a centralised system for every consultation but only to update the patient’s records, if desired.

3. Methods

3.1. The alternative: decentralised management of MRs

First, in industrialised countries, each health-care structure whatever the type (public or private) has an information system that gathers structured or unstructured computerised medical records. Then, information contained in the daily routine MR is sufficient for most of the needs of health professionals. It is rarely necessary to gather distributed information on a patient for health care reasons. Thus, the additional work a doctor needs to do to reconstitute a patient’s medical history (MH) is limited unless the patient frequently seeks treatment in different places. Such cases are rare and doctors have this extra workload only occasionally. Given the two previous points, and the dangers and complexity of a centralised system, it seems reasonable to us to set up a system that allows each doctor, with the authorisation of the patient, to collect information on that patient from the different health structures. The doctor, or the medical information exchange application he uses, will then have to synthesise the patient’s MH for his specific use, save it, and update it regularly. This effort to synthesise MRs will be reduced because, one doctor can pass on information about one of his patients to other doctors when this patient moves. For example, the General
The basic organisational advantage is that it could be operational rapidly because problems of harmonisation will be reduced. The decentralised management principle supposes that the saved MR remains in its unmodified form in hospitals and clinics in terms of content and structure, and remains identifiable by certain elements that exist in all patient MRs such as first names, last names and dates of birth, and requires no additional indexing. When the patient and his/her doctor want to gain access to patients’ medical data that is distributed among the servers of various hospitals or clinics, they have to be connected to an electronic server on which they identify themselves. The identification of the doctor should be based on strong authentication credentials. Typically, the doctor might use a token activated with a PIN code (or with biometrics) to give the answer in a challenge-response protocol. The identification of the patient could be based on a smartcard (E-health card) for example.

Once the authorisation is granted, Medical Record Search Engines (MRSEs) will securely gather medical information about the patient in a privacy-compliant way and transfer it to the medical practitioner.

A practical solution describing the flow of information and how the MRSEs can achieve their task in a secure privacy-compliant way is explained below.

3.2. Description of the Medical Record Search Engine

The basic scenario we consider is as follows. A medical practitioner facing a patient wishes to gather information regarding this patient. The aim of the proposed system (cf. Fig. 1) is to request all potential providers of medical information (hospitals, health professionals) to provide information. Two main points are considered.

First, all of the retrieved information will be gathered by the system before being transferred to the medical practitioner’s office. One special feature of the proposed architecture is that the management system requests information without directly reading in the local (hospital or health professionals) information providers. More clearly, all MRs are kept and managed in a decentralised way in the “local” health structures (HS), recorded according to the system provided by the health structure software and identified with the usual identifiers (first name, last name and birth date) which are present in all MRs.

Second, the patient’s privacy is protected by using a pseudonymous code (derived from the patient’s identity). All communications are encrypted.

To access a patient’s medical records, the proposed procedure can be described in seven steps (cf. Fig. 1).

3.2.1. First step: pseudonymisation of patient’s identity

During a consultation between a Medical Practitioner (MP) and a patient, the MP enters the components of that patient’s identification. This information related to the patient’s identity will be anonymised, using a robust cryptographic hash function to provide a hashed patient identity called H(Pi). The aim of this algorithm is to obtain a pseudonymous code, but, hopefully, always the same one for a given individual in order to link all of the information concerning any given patient.

3.2.2. Second step: sending the request to the two MRSEs

When a medical practitioner wants to request a patient’s information, distributed in some other HS, he has to send a request to two MRSEs and authenticate both himself and the patient. Exchanges between the MP and the MRSEs are protected by using an asymmetric encryption algorithm (like the RSA encryption). In this communication, the public keys (PMRE1 and PMRE2) of the MRSEs are used by the MP.

The MP sends a request called “x” to the two Medical Record Search Engines MRSE1 and MRSE2. The system hinges on the prevention of communication between MRSE1 and MRSE2. As seen in Fig. 1, the information sent by the MP is split between the MRSEs. The purpose of MRSE1 and MRSE2 is to guarantee the confidentiality and the privacy of the request during its transmission.

MRSE1 receives three elements:
(a) x, the number of the request;
(b) K, a session key;
(c) e_j, the MP public key.

MRSE2 receives two elements:
(a) x, the number of the request;
(b) E_E(H(Pi)), the hashed patient identity H(Pi), previously symmetrically encrypted by the MP with the session key K.

Through this procedure, MRSE2 is unable to access the pseudonymous patient identifier as it does not know the session key K, which has been transmitted to MRSE1. This prevents MRSE2 from finding the identity of the patient through, for example, a dictionary attack. To preserve the patient’s privacy optimally, MRSE1 and MRSE2 must not be able to communicate with each other.

E2 must not be able to communicate with each other.

3.2.3. Third step: request transmitted to all health structures by the two MRSEs

To transmit the request to the hospitals, both MRSEs first have to decrypt the messages sent by the MP, using their own private keys.

Then MRSE1 and MRSE2 consult a health structure directory in order to forward the request to all health structures to which they are connected. MRSE1 and MRSE2 sign their respective part and send it to the hospitals. The requested information is encrypted with the health structure’s public keys.

3.2.4. Fourth step: search for the patient’s MR at the health structure level

Locally, each HS decrypts messages issued by MRSEs. Then, they also decrypt the pseudonymous patient identifier (H(Pi)) with the session key K.

Each hospital h_i can then search for medical records corresponding to this pseudonym (comparing it with hashed identities of the patients hospitalised in h_i). If this search is successful, i.e. if one corresponds to the previously received
H(PI), the corresponding patient’s MR will be gathered before being sent to the aggregator.

3.2.5. Fifth step: transfer of the results of the request to an aggregator
This step consists in sending to an aggregator a record containing three elements:

(a) the number of the request, \( x \);
(b) the hashed patient identity, \( H(PI) \);
(c) the patient’s MR, digitally signed by hospital \( h_i \) with an electronic signature.

This electronic signature allows non repudiation and verification of the integrity of the message. To ensure transmission security, confidential medical information such as the hashed patient identity \( H(PI) \) and the patient’s MR are asymmetrically encrypted with the MP public key \( e_j \). The MP is the only one who can decrypt this confidential medical information with his private key.

3.2.6. Sixth step: gathering all patient information at the aggregator level
The aggregator collects information received from all health structures and gathers all the results of the same request \( x \). These results are sent to the MP, after a challenge-response authentication procedure. The MP will be then able to decrypt these results with his own private key.

4. Discussion
MRSEs are platforms that coordinate a decentralised search. They never have direct access to the database of the local systems of the HS as it is the HS itself which makes the requests on it. MRSEs are platforms where encrypted infor-
mation is temporarily stored before being passed on. MRSEs do not store any MRs but may keep logs of transactions. One of the key points of the procedure we propose relies on the fact that MRSE1 and MRSE2 are not allowed to communicate. This solution guarantees confidentiality and privacy during the communications between the MP and HS. In this way, MRSE1 does not manage patient data, and MRSE2 only manages pseudoanonymous encrypted data. Hence, the MRSEs must be hosted in different locations under different responsibilities. Regarding the organisation of the MRSEs and of the aggregator, a recursive infrastructure could be envisaged at the regional, national and European level if needed. MRSEs and aggregators could be set up at a regional level, for example. In such a structure, a main central aggregator will have to manage the reception of multiple documents related to the same patient from local aggregators. Of course, the MP will have to decide and fix a maximum time for document return. The same is true for the information handled by MRSEs.

Among the different problems to consider for this approach is the question of the quality of a patient’s information transmitted to the MP and thus of the quality of data linkage. The absence of a need for a new unique identifier is a major advantage, but it may raise some discussion about doublons and collision risks. Concerning the doublon risk, i.e. the risk of losing some information (that is not being able to link information concerning the same patient) due to typing errors, for example, it is important to underline that this may also happen when typing the unique identifier. It is possible to reduce such errors by using phonetic algorithms as described in a previous paper [6] or other robust transformations like in [7] or in [8]. To improve linkage quality, it is also possible to envisage that the MP sends not only one pseudonymous identifier per patient but a list of pseudonymous partial identifiers for each patient. The creation of each pseudonymous partial identifier can be based on the different combinations of first names, last names and dates of birth of a patient as follows:

- If a patient has two last names such as in the case of a married woman or a child whose parents are divorced, one can create two pseudonymous partial identifiers (one for each last name).
- It is also possible to do the same for one patient who has two first names (like David Olivier or François André).

Concerning the risk of collision (i.e. two MRs from two different patients may be mixed), apart from the fact that it may also happen today, this risk implies that the medical practitioner has to check all information and this, eventually, with the help of the patient. A feasibility test, based on observed data could make it possible to give to each record a linkage probability level (high, medium or low). These levels could be obtained from the probabilistic modelling done on observed data and could be managed by the MREs. This will permit the centralised aggregator, when transmitting the results of the request, to give a hierarchical order with high, medium or low probability levels, so as to help the MP in the validation process.

Above all, whatever the amount of lost or false information, the situation (from the professionals’ and patients’ point of view) will be better than the great lack of information we have now.

Some medical practitioners may complain that this procedure will increase their workload because they will be obliged to read all of the different information to detect possibly false or missing information. The burden is not as heavy as they may think for many reasons. First, this enquiry on the patient’s past history is far from being mandatory for all patients and generally the contents of their local records are sufficient to treat the patient. Second, few patients (but it should be verified by a survey on a real situation) have been hospitalised in many different hospitals and therefore, the number of different medical records that MPs will have to summarise will be more frequently one or two (or eventually three) than ten or twenty. Third, it is important to take into account that this synthesis has to be made in collaboration with the patient, who could provide great help in the management of his own records. For example, concerning the doublon risk, i.e. the risk of losing information, it is easy to ask the patient if he has been hospitalised somewhere other than in the hospitals that have answered the request and provided information. Similarly, concerning the risk of collision, (the amalgamation of two or more MRs from different patients), it is usually easy except in emergency situations to ask the patient if he has in fact been hospitalised in all the hospitals that have provided parts of his medical record. The situation is less easy if the records provided came from the same hospitals, but here again the patient should be able to say if the date of the stay and the disease recorded correspond to him or not. Of course, sometimes the patient is not conscious and cannot answer the questions. This case is not frequent, and we must not build a rule to solve an exception, and secondly, it is in such a case when the patient cannot give any information that it would be useful to have information on his past history quickly even if it is incomplete or insufficiently precise. Fourth, concerning the data gathered, one may think that it could be useful to reduce the amount in order to make it easier to handle. Therefore, we may imagine adding a selection criteria tool to the basic request as in any kind of search engine. The criteria could concern the selection of a time period, a list of hospitals, a clinical department or event, the pathology about which the medical practitioner requests precise information. However, the information, even when reduced in quantity, will need to be analysed and reorganised by the practitioner.

Moreover, one last but by no means minor advantage is that these Medical Record Search Engines using pseudonymisation of patient’s identity will make it possible to retrieve the past histories of all patients without the need to reformat the data and above all without the need to build an additional specific structure. Therefore it could be implemented rapidly at a much lower cost than a central medical record. The regional level could be the right level to set up an experiment to test this proposal.

5. Conclusion

The main reasons for the failure of centralised MR management are related not only to insufficient human and
Summary points

What was already known on the topic:

- The difficulties of centralizing standardized Medical Records (MRs) at the level of a country are already known.
- They are summarized in Section 2 to enlight the interest of developing an alternative method based on the decentralized management of MRs.

What this study added to our knowledge:

- In this study, we describe a new method allowing Medical Practitioners (MPs) to reconstitute a patient’s medical history, when necessary, without needing to implement standardized and centralized medical records.
- The Medical Record Search Engines, using pseudonymisation of patient’s identity, provide upon a MP’s request all the available information concerning a patient (such as hospitalizations).
- The security protocol is very strict. The request is splitted between two MRSEs in order to protect separately medical data and patient’s pseudonymous code (derived from the patient’s identity) from any disclosure (divulgation). Exchanges between the MP and the MRSEs are protected by using an asymmetric encryption algorithm.

financial resources but also to difficulties in the implementation of a UPI and the lack of MR standardisation. In this paper, we have discussed the interest of a pragmatic solution relying on existing data. The collection of information through Medical Record Search Engines using pseudonymisation of patient’s identity could be a secure alternative solution to the Centralised Medical Record Management as it offers a real service that would be easy to set up in a short time. It would be much less expensive and avoid the need for major reorganisation of the hospital medical archives everywhere.

References


