Proposal of a French Health Identification Number Interoperable at the European Level

Catherine Quantin\textsuperscript{a}, François-André Allaert\textsuperscript{b}, Paul Avillach\textsuperscript{c,d}, Benoît Riaudey\textsuperscript{e}
Marius Fieschi\textsuperscript{c}, Maniane Fassa\textsuperscript{a}, Olivier Cohen\textsuperscript{f}

\textsuperscript{a}. Service de Biostatistique et Informatique Médicale, CHU de Dijon, BP 77908, 21079 Dijon Cedex, INSERM EMI 0106, France
\textsuperscript{b}. Department of Epidemiology & Biostatistics, McGill University, Montreal, Canada
\textsuperscript{c}. LERTIM Faculté de Médecine Université de la Méditerranée 27 Bd Jean Moulin 13385 Marseille cedex 05, France
\textsuperscript{d}. ISPED Université Victor Segalen Bordeaux II, 146 rue Léo Saignat, 33076 Bordeaux cedex, France
\textsuperscript{e}. Institut National d’Etudes Démographiques (INED), 133 Bd Davout, F-75980 Paris Cedex 20, France
\textsuperscript{f}. Laboratory TIMC-IMAG UMR 5525, CNRS University of Joseph Fourier Medical School of Grenoble, France

Abstract

The French ministry of Health is setting up the Personal Medical Record (PMR). This innovative tool has long been expected by French Health Authorities, Associations of Patients, other Health’s associations, those defending Individual Liberties and the French National Data Protection Authority. The PMR will lead to improvements in many areas such as Diagnosis (Research and monitoring) Healthcare (Management of emergencies, urgent situations, Temporal health monitoring and evaluation), Therapy (Cohorts of patients for Clinical trials and epidemiological studies). The PMR will foster safe healthcare management, clinical research and epidemiological studies. Nevertheless, it raises many important questions regarding duplicates and the quality, precision and coherence of the linkage with other health data coming from different sources. The currently planned identifying process raises many questions with regard to its ability to deal with potential duplicates and to perform data linkage with other health data sources. Through this article, using the electronic health records, we develop and propose an identification process to improve the French PMR. Our proposed unique patient identifier will guarantee the security, confidentiality and privacy of the personal data, and will prove to be particularly useful for health planning, health policies and research as well as clinical and epidemiological studies. Finally, it will certainly be interoperable with other European health information systems. We propose here an alternative identification procedure that would allow France to broaden the scope of its PMR project by making it possible to contribute to public health research and policy while increasing interoperability with European health information systems and preserving the confidentiality of the data.

Keywords: electronic health records, unique patient identifier, security

Introduction

In the majority of industrialized countries, at the heart of many of the concerns relating to electronic processing of health information lies the problem of patient identification. In August 2004, the French government decided by law to initiate a national project concerning Electronic Health Records called the “Dossier Médical Personnel”, the Personal Medical Record (PMR) [1]. It intends to promote health care coordination, enhance the communication of health information and reduce iatrogenic accidents. The most important aim of this article is to show and present the principles of the French unique patient identifier created relative to the PMR. However, it will also explain and show its major disadvantage, which is its interoperability characteristic. For example, this current patient identifier seems to be incompatible with the identifier of the European health card [2]. This paper will thus demonstrate, in its first section that the French health identification number does not ensure interoperability at the European level. The second section proposes an alternative identification process that would allow France to improve the quality, security, precision and coherence of its PMR project. The developed and proposed health identifier will be extremely useful for health planners, those dealing with health policies, public health research and clinical and epidemiological studies, and this, at national, regional and international level.

Materials and methods

The issues and criteria of a unique patient identifier

The identification of the patient in a health-care structure and particularly in the framework of the electronic health-care record of a care network is a major issue [3]:

- Patient care continuity requires secure, precise, coherent and reliable patient identification through all of the health information systems.
The reduction and management of identification-related error is one of the major constraints necessary to maintain and improve the quality of care;

A reliable unique format for patient identification is required in order to design and implement an interoperable health information system focused on the patient.

In the context of the “Principle and Process of Patient Identification” project, the French Group for the Modernization of Hospital Information Systems (GMSIH) [4] inventoried the principles and the architecture of identification systems by specifying two aspects of the problem:

1. “Identification” will specify which information will be used to uniquely identify a patient. Multiple identifiers can be proposed for the same patient in different applications.

2. “Merging” identification areas, requiring the implementation of specific methods (intra-health structures and among health structures).

This group also inventoried the international patient identification experience of ten countries (Germany, Australia, Canada, Denmark, Finland, Luxembourg, the United States, New Zealand, The Netherlands and the United Kingdom) considered by the GMSIH as representative of best practices in the areas of socialized health care, electronic health care information systems or patient identification.

The American Society for Testing and Materials (ASTM), a standards development organization accredited by the American National Standards Institute, identified 30 criteria that were published in the Standard Guide for Properties of a Universal Healthcare Identifier [5]. The most recent standard on Health Care Client Identification we know is from Australian Standard® AS 5017-2006 published in June 2006 [6]. This standard includes data elements that jointly comprise a unique identifier for health care clients. It is explained that the combination of the client identifier and the health care establishment that assigned the identifier is one way to indicate unique identification. The logical structure for these data elements has shown that, there may be multiple identifiers collected for any one individual. Together, the data elements that makeup one complete Health Care Client Identifier are a) Health Care Client Identifier Designation b) Health Care Client Identifier Geographic Area c) Health Care Client Identifier Issuer d) Health Care Client Identifier Type.

Most health care clients have more than one identifier. The most commonly used identifier should be collected as the Primary Client Identifier (the first listed client identifier). This identifier is generally that assigned by the organization as the means of uniquely identifying the client. Often, there are other identifiers also related to the client, and these should be collected and recorded as Other Client Identifiers (listed second or subsequently in the identifier list). There can only be one Primary Client Identifier, but there may be many Other Client Identifiers collected for any one application. Examples of Health Care Client Identifiers include 1) Person Identifier 2) Medical Record Number (MRN) 3) Local Client Identifier 4) Health (care client) Identification Number 5) Unit Record (UR) Number 6) Enterprise Identifier 7) Area Identifier 8) State/Territory Health Identifier 9) Unique Identifier (UID) 10) Unique Health Identifier (UHID) 11) National Health Identifier (NHI).

The five most important criteria and characteristics (Table 1) of the Health Care Client Identifier (HCCI) Designation are adapted from the AS5017-2006 and provided here as a guide to assigning a unique identifier [6].

<table>
<thead>
<tr>
<th>Atomic</th>
<th>Content-free</th>
<th>Longevity</th>
<th>Permanent</th>
<th>Unambiguous</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>(the HCCI Designation should be a single data item. It should not contain sub-elements that have meaning outside the context of the entire HCCI Designation. Nor should the HCCI Designation consist of multiple items that must be taken together to constitute an identifier).</td>
<td>(the HCCI Designation should not depend on possibly changing or possibly unknown information pertaining to the health care client. Including content in the HCCI Designation will make it impossible to assign the ‘correct’ identifier if that information is not known. It also leads to invalid situations if the information changes: for example, what happens to an identifier based on sex if the health care client has a sex change procedure).</td>
<td>(an HCCI system should be designed to function for the foreseeable future. It should not contain known limitations that will force the system to be restructured or revised radically).</td>
<td>(once assigned, an HCCI Designation should remain with the health care client. It should never be reassigned to another client, even after the health care client’s death).</td>
<td>(whether represented in automated or handwritten form, an HCCI Designation should minimize the risk of misinterpretation. Where using alphanumeric identifiers, be aware of possible confusion with the number ‘0’ with the letter ‘O’ and the number ‘1’ with the letter ‘I’).</td>
<td>(a valid HCCI Designation should identify one and only one health care client. A health care client should have only one primary HCCI Designation).</td>
</tr>
</tbody>
</table>

These criteria are designed to support four basic functions of a universal health care identifier:

1. Positive identification of patients when clinical care is rendered,
2. Automated linkage of various computer-based records on the same patient for the creation of lifelong electronic health care files,
3. Provision of a mechanism to support data security for the protection of privileged clinical information (does not attempt to address all safety concerns, however)
4. Use of technology for patient record handling to keep health care operating costs at a minimum.
The founding principles of the French Health Identification Number

In September 2005, the French government defined the founding principles of the Health Identification Number (INS in France) of the personal medical file (PMR). This number must be:

- **unique**: to ensure that two people cannot have the same INS, thus avoiding erroneous attribution of a diagnosis. An identifier is therefore associated with a person whose identity can be validated.
- **content-free**: the number reveals no information about the holder (sex, age, place of birth…).
- **public**: the number can be legally used by all health care professionals authorized by the patient. It can be stored in the health care professional’s information system. The number is therefore not secret.
- **permanent**: will remain the same for the life of the patient and, possibly, beyond.
- **irreversible**: impossible to determine the identifier by calculating backwards from the social security number. However, if lost, the same number can be recreated.

First proposal for the creation of the INS in the PMR

To meet the “uniqueness” condition, the French patient identification working group recommended that the identifier be associated with the social security number. The solution proposed by this working group involves the creation of a Health Identification Number (INS in France) for each patient managed by the patient’s request by an independent organization known as a “trusted authority”. The role of the “trusted authority”, which may not be a host, is to guarantee that the identifier is not and will not be duplicated. The trusted authority will also be in charge of the secure access, the respect of confidentiality and the integrity of all the identification workflows.

To preserve the anonymity of the identifier, the trusted authority creates the INS through the following process:

- The patient chooses an approved host and advises the PMR office using a request form and the host using a membership form.
- By contacting the National Health Insurance organization the PMR office confirms the requester’s affiliation and that it is a first-time request.
- The trusted authority creates the INS using the request form number that it transmits along with patient attributes to the host.
- The host then informs the trusted authority that the patient’s INS is operational.

The use of the request form number and not the social security number to generate the INS is designed to guarantee that the number is content-free and the INS creation process mathematically irreversible, i.e. no mathematical operation would permit backward calculation from the INS to lead to the social security number. The guarantee of INS anonymity would rely on the absence of a table of correspondence directly linking INS and the social security number. Detection of collisions and duplicates would occur by way of the request form number. Nonetheless, the relationship between the social security number and INS would be indirectly maintained at both sides. One and only one request form number would be associated with each social security number and one and only one INS would be associated with each request form number. The trusted authority would not require knowledge of the social security number to prevent collisions; they would need only to ensure that the INS had been assigned to no other request form number. Similarly, by checking that each form number is given to only one INS, they would prevent duplicates provided that no duplicate form number existed further up the chain.

The trusted authority would send the INS to the host along with other patient identifiers, such as first and last name and date of birth, but never the social security number. The host would thus possess only those identifiers that could not be used to reconstitute the social security number.

Criticism of this method of generation

The principal criticism [7] addresses the second main function, as defined in the ASTM Standard Guide for Properties of a Universal Healthcare Identifier [5]: Automated linkage of various computer-based records on the same patient for the creation of lifelong electronic health care files.

Thus, the fact that the proposed system leads to an institutional disconnect between the INS as a central health system identifier and the social security number. The main reason cited to justify this disconnect is the desire to prevent the creation of a population table of correspondence between the social security number and the INS. This argument is, however, open to criticism insofar as all health professionals and in particular large institutions whose patient databases cover a large proportion of the population of a region must have access to the correspondence between the social security number and the INS.

If the link between the Social Security Number (SSN) and the INS is broken, relating PMR data with other patient data for research purposes becomes impossible. For example, relating PMR with other sources is necessary not only to complete their data but also to ensure multi-source validation in conformity with the recommendations of the report of the academy of sciences regarding epidemiology. If research using PMR data is not possible in the short term because of the time necessary to create databases that are sufficiently structured, exhaustive and validated, it can be expected that the data will be of use in less than a decade, if an acceptable identifier is adopted. That’s why, like our British colleagues who complain that overzealous interpretation of UL laws stifles epidemiological research [8-11], we think that the current French identifier may cause particular difficulties regarding epidemiological and clinical research.

Moreover, it must be possible to reconstitute the INS, which represents the true signature of a patient, whatever the nationality or the country of origin starting from features of identification that are always available. These methodological principles must guarantee the secure and perennial identification of all the patients. These principles must apply whatever the future use of the information. Indeed, the technical and methodological choices must be able to follow evolutions in the legislation, which may be modified according to the evolution of mentalities and cultures.
Our proposal

It would be perfectly possible to preserve the confidentiality due to the patient and desired by patient associations by setting up anonymous procedures [12-14] such as those adopted by the Institut de Veille Sanitaire (Health Surveillance Institute) on the recommendation of the French National Commission for Data protection and the Liberties (CNIL), in the context of the follow-up procedures for the 30 diseases subject to mandatory reporting (including AIDS).

Unlike encryption methods that must be reversible to allow the legitimate recipient to decode the message, unidirectional hashing techniques are irreversible. Hashing produces a perfectly anonymous code (it is not possible to retrace the patient’s identity) that is always the same for a given individual so that patient data can be linked. There are many medical applications, which include the creation of national databases (such as those relating to the national follow-up of infected subjects - approximately 100,000 patients – an excellent example of what can be put to the service of epidemiological research, with complete patient approval) as well as regional and inter-regional databases in many areas (cancer, perinatality, genetic diseases). This system has also allowed (based on the hash-coding of the Social Security Number (SSN), the gender and the date of birth), standardized hospital discharge abstracts to be linked, classified into French Diagnosis Related Groups at the French national level and linked to the data of the national medical insurance information system. An anonymous procedure based on hash coding is also used for chaining patient files in Switzerland [15].

In the case of the PMR the situation is not, in fact, more complex because several requirements must similarly be met in a similar way:

- demands of the CNIL, patient and healthcare professional associations regarding confidentiality of personal information contained in the PMR, and respect of the law of August 13, 2004 concerning the use of data,
- needs in public health or for individuals to have access to these data, particularly when the patient has given express consent.

Ideally, hashing the social security number would help meeting these requirements (Figure 1).

Regarding confidentiality, insofar as the social security number could not be reconstituted using the INS, the link between them would be broken. Another advantage of using hash-coding is to meet the criterion of being focused (created and maintained solely for supporting health care cf. Table 1). As a consequence, using different keys for hash-coding will allow the creation of distinct identifiers according to different purposes (health administration, diagnosis and healthcare, epidemiological research). The same solutions also derived through an irreversible encryption of the unique Social Security Number (SSN) have been proposed in Belgium [16] and New Zealand.

Regarding data access for public health research, the use of the Social Security Number (SSN), after hash-coding, would thereby allow the linkage of the main national databases. However this solution raises many questions. Firstly, in France, foreigners do not have a complete social security number. For instance, the last characters, corresponding to the country of birth, are not systematically filled. Moreover, as the date of birth in some foreign countries is not known precisely, two different patients may have the same date and the same country of birth, resulting in collisions, and linkage errors. Secondly, as the Social Security Number (SSN), has a different structure according the country, it can not be a solution for a unique patient identifier, interoperable at the European level.

One solution would be to add personal patient characteristics such as family name, first name, date of birth (separately hashed) to the social security number, which would help to conform to the recommendations of the International Association of Medical Information Technology and where possible to ensure interoperability of this identifier with a European identifier [17]. In fact, in the national medical insurance information system like in the standardized discharge abstracts, gender and date of birth have already been associated with the social security number, in the same identification field, before hash-coding, resulting in a single code. We could thus propose to build the INS through the simultaneous hash-coding of gender and date of birth associated with the social security number.

However, due to a simple error in the field for gender, linkage of data from the same patient can be definitively refused. It is important to note that this field is highly unreliable (numerous data entry errors have been observed in hospital settings). In addition, with regard to the quality of the linkage this variable is undiscriminating insofar as the probability of two people being the same gender is close to 50% and it does not discriminate for twins. It would therefore be preferable to replace this variable with one that is more sensitive, such as the first and last name of the beneficiary. As a consequence, our first proposal is to separately hash the social security number (removing the last digits), the date of birth and the last and first names, then merge this into a single signature of patient identity (Figure 2).

Regarding confidentiality, insofar as the social security number could not be reconstituted using the INS, the link between them would be broken. Another advantage of using hash-coding is to meet the criterion of being focused (created and maintained solely for supporting health care cf. Table 1). As a consequence, using different keys for hash-coding will allow the creation of distinct identifiers according to different purposes (health administration,
We could also propose, if possible to add biometric components (separately hashed). Biometric technologies are sometimes proposed in solving the problem of associating patients with their medical data, as they do not require the patient to bring any documents or remember information. Though this technology represents a real progress both in the identification and in the authentication of the patient, it raises many questions [18]. First, the accuracy and reliability of each biometric technology (for instance: finger print, iris scan, retinal scan and DNA) are not optimal in all circumstances. For example, fingers are frequently injured. Though retina scans and DNA analyses are the most accurate (with the exception of twins for DNA) they are also quite intrusive, which render even more difficult the possibility of combining these technologies, as proposed in some commercialized biometric systems. Moreover, the costs of the biometric solutions have to be considered. But the main problem lies in their acceptance by ethical organizations such as patients associations, national committees on ethics, human rights associations and national committees for data protection. In order to render this solution compatible with the current national procedures, it would be preferable to add the biometric component (separately hashed) to the hashed social security number, in order to ensure (at the European level) the interoperability of this identifier with national ones.

**Conclusion**

Our proposal for a French Health Identification Number will make it possible to uniquely identify and link a patient to his specific medical data. By hashing the social security number it will be possible to link the information of the personal medical file to other national health information sources, with the aim to complete or validate Personal Medical Data (PMR) or conduct epidemiological research. Adding personal patient characteristics such as first and second names and date of birth and/or biometric identifiers (separately hashed, then merged) to the hashed social security number would also contribute to the establishment of European public health statistics by matching healthcare data of the patients' records with other administrative data (mortality, social information …). This data linkage would thus meet the requirements of anonymous data of the European directive on data protection and health related services. Of course, even if anonymised, the use of biometric components would require, the authorization of the National commissions for data protection.

**Acknowledgements**

This research was supported by The French National Agency for Research (ANR).

**References**


[14] ISO TS 25237 "Health informatics - Pseudonymisation practices for the protection of personal health information and health related services”


**Address for correspondence**

Professeur Catherine QUANTIN
Service de Biostatistique et Informatique Médicale
Centre Hospitalier Universitaire - BP 77908
21079 DIJON CEDEX - Tel: 33 3 80 29 36 29
Fax: 33 3 80 29 39 73
Email: catherine.quantin@chu-dijon.fr