

Don't take your EHR to heaven, donate it to science: legal and research policies for EHR post mortem

Vojtech Huser, James J Cimino

Laboratory for Informatics Development, NIH Clinical Center, Bethesda, Maryland, USA

Correspondence to

Dr Vojtech Huser, Laboratory for Informatics Development, NIH Clinical Center, 10 Center drive (Rm 10/6-5561), Bethesda, MD 20892, USA; vojtech.huser@nih.gov

Received 31 May 2013

Revised 2 July 2013

Accepted 23 July 2013

Published Online First

21 August 2013

ABSTRACT

Recently, an important public debate emerged about the digital afterlife of any personal data stored in the cloud. Such debate brings also to attention the importance of transparent management of electronic health record (EHR) data of deceased patients. In this perspective paper, we look at legal and regulatory policies for EHR data post mortem. We analyze observational research situations using EHR data that do not require institutional review board approval. We propose creation of a deceased subject integrated data repository (dsIDR) as an effective tool for piloting certain types of research projects. We highlight several dsIDR challenges in proving death status, informed consent, obtaining data from payers and healthcare providers and the involvement of next of kin.

INTRODUCTION

In recent years, electronic health records (EHRs) and large healthcare integrated data repositories (IDRs) have had significant impacts on observational clinical research.¹⁻³ Technologies for 'big data'⁴ enable analyses that were previously technologically impossible.

Despite these promises, there are also challenges and barriers to using EHR data.⁵ Existing legal policies require ethical approval by an institutional review board (IRB) for every research analysis meant for publication. Existing policies specifically prevent broad research topics (eg, retrospective analysis of EHR data) and require separate IRB approval for each individual research analysis (eg, retrospective analysis of EHR data in diabetic patients taking rosiglitazone).⁶

To eliminate the requirement for individual IRB approval, we suggest a research framework that uses data of deceased patients and we refer to it as deceased subject integrated data repository (dsIDR). Such a repository, either local or federated, would enable limited research use of data by qualified researchers. In this perspective paper, we look at legal and research policies for decedents' EHR data. In our legal analysis, we focus on federal regulations and do not analyze situations which are modified by more stringent state laws.

FACILITATING RESEARCH ACCESS TO EHR DATA

In principle, there are two approaches that enable research on EHR data without IRB review of each individual project. One approach is removal of protected health information (PHI) through de-identification.⁷ Examples of the de-identification approach are the Columbia University RedX platform⁸ or Vanderbilt University synthetic derivative

platform.⁹ Research analyses using de-identified IDRs are considered non-human subject research and consequently do not require IRB review. Significant advantage of de-identified IDRs is the ability to analyze an entire patient population in a healthcare data warehouse (given proper infrastructure). However, their disadvantage is possible distortion of the data during de-identification,¹⁰ especially if unstructured data from clinical documents are included in such repositories.

The second, less commonly used approach that also eliminates IRB review uses data of deceased subjects. The privacy rule of the Health Insurance Portability and Accountability Act (HIPAA) [45 CFR 164.512(i)(1)(iii)] allows disclosure of PHI of decedents without the need to seek IRB or privacy board approval.¹¹ Such use can occur if three conditions are met: (1) use is sought solely for research on the PHI of decedents; (2) the researcher can provide, on request, documentation of the death for subjects used in the study; and (3) the PHI is necessary for the research. Throughout this article, we refer to this HIPAA provision as the 'decedent research clause'.

For general, non-research use of EHR data of note is also a recent change in January 2013 to the HIPAA privacy rule (also known as the omnibus rule) that shortened the protection of PHI from an indefinite interval to a period of 50 years after death. However, state law can extend this federal law provision. For example, the state of Hawaii over-rides this federal law and mandates indefinite protection (Haw. Rev. Stat. section 323C-43).

VALUE OF RESEARCH ON DECEDENTS

It is important to examine whether dsIDRs provide significant research value, especially in light of the increased availability of de-identified EHR or claims data. For example, depending on the institution and the research topic, researchers can access locally established de-identified EHR data or external de-identified EHR databases such as Humedica's dataset of 30 million patients or the University Healthcare Consortium's Clinical Database-Pharmacy. Such datasets contain very detailed EHR data but may not cover the entire lifespan of the patients. External claims databases, in comparison, may offer larger sample sizes and long lifespan; however, they are limited to billing data. Examples of external claims databases are the Agency for Healthcare Research and Quality's (AHRQ) Healthcare Cost and Utilization Project data,¹² CMS's Chronic Condition Data Warehouse¹³ or United Healthcare claims database (Optum Labs Real World Evidence dataset).

So why would a researcher, who wants to eliminate the IRB step in a preliminary research, use a

To cite: Huser V, Cimino JJ. *J Am Med Inform Assoc* 2014;**21**:8-12.

dsIDR when de-identified datasets are available? First, not all institutions have established de-identified IDRs with streamlined IRB processes. Establishing a high-quality de-identified IDR can be complex and expensive, mainly owing to de-identification costs to eliminate PHI from unstructured clinical documents and costs associated with modification of coded data (eg, diagnoses, demographics) that minimize re-identification risk (data suppression, generalization and perturbation).¹⁴ Second, many of the external datasets can be expensive to license. In such cases, using the decedent research provisions in HIPAA is the only way to access EHR data that does not require IRB approval. Another advantage is that there is no requirement for de-identification, although, for unstructured clinical documents data, this depends on the type of clinical document. Whereas EHR records always contain what we refer to as *primary PHI* (information about the patient in question) they may also contain *secondary PHI* (information about individuals other than the primary patient, such as relatives or caregivers). Although not dealt with by the HIPAA decedent clause, privacy principles require removal of secondary PHI of any living individuals. We hypothesize that the number of secondary PHI elements is much smaller than that of primary PHI. Hence, the resulting distortion of clinical text data after secondary PHI removal is smaller than with de-identified, non-decedent EHR data where both primary and secondary PHI must be removed.

More research is needed on the exact occurrence of secondary PHI and on the different complexity of de-identification methods that distinguish between primary and secondary PHI. For example, any mention of geographic location or date may be preserved in dsIDR (unless it is part of secondary PHI). Most likely clear guidance will be needed on how, or whether at all, the traditional list of 18 private elements applies to decedent records. In the meantime, creation of phase I dsIDRs limited to only administrative billing data integrated with highly structured EHR data, which is guaranteed to be free of secondary PHI data, should be relatively easy, and it should offer a highly accessible resource of significant value to some researchers into comparative effectiveness and outcomes.

In addition to less need for de-identification and potentially smaller record distortion, another advantage of decedent clause research is the increased ability of researchers to integrate decedents' EHRs from multiple entities (multiple health plans or providers). Because the patient is deceased this permits dsIDR researchers to request the record using the decedent clause rather than research project informed consent.

dsIDRs have two significant disadvantages. The first is smaller sample size compared with analyzing entire populations. The second disadvantage is the inability to analyze recent EHR data. Most deceased patients will be older patients (2011 USA average life expectancy is 78.6 years) and analysis of EHR

events occurring at mid-life stage will thus be working with data that may be decades old. For some medical conditions or research projects in which clinical documentation and treatment patterns change rapidly this may be a limitation. However, we envisage the dsIDR to be a piloting platform that leads to full IRB-reviewed research proposals investigating complete IDR populations as the next step. In many informatics projects (eg, named entity recognition methods) or clinical research (eg, natural history studies) these two limitations may be acceptable. Table 1 compares de-identified EHR repositories with the proposed dsIDR.

DECEASED SUBJECTS' IDRS

We see significant value in explicitly defining the concept of dsIDRs and developing a set of best informatics practices for researchers who need to pilot an EHR analysis project without seeking complex IRB approval.⁶

We envision the dsIDR to be available for limited use in research (not in the same sense as the HIPAA limited dataset on living patients), guarded by a login mechanism allowing access only to eligible researchers. Despite the earlier discussion of lesser de-identification, we envisage that the final dsIDR data extracts released to researchers would have erased patient names, medical record numbers and exact street addresses where they are recorded in structured form since those have limited research value. The terms of use of the dsIDR would ensure research use only, forbid users from re-identifying patients or living secondary PHI parties or conduct analyses related to malpractice. Federation of dsIDRs from multiple institutions would enable the variability of EHR records across institutions to be studied. Existing credentialing systems, such as electronic Research Administration commons¹⁵ could even be leveraged for such systems.

The possible sample size of a national scale dsIDR is indicated by the fact the Central Intelligence Agency Factbook estimates that each month about 221 thousand deaths occur in the USA. This implies that over a period of 2 years the dsIDR may accumulate 2.6 million patients in a national dsIDR if a 50% accrual rate is achieved.

A long-term perspective further emphasizes the possible value of dsIDRs, since data on deceased subjects in any healthcare research warehouse will at some point prevail over data on living individuals. This rise can be already be seen in current repositories and is also evident from demographic estimates comparing the total number of dead versus living people.^{16 17}

CHALLENGES

In addition to the limitation of sample size, we discuss several additional dsIDR challenges that require legislative, regulatory, or technical solutions (summarized in box 1).

Table 1 Comparison of research databases that do not require IRB approval

EHR data (de-identified)	Claims (de-identified)	dsIDR (deceased patients)
<ul style="list-style-type: none"> ▶ Detailed EHR data ▶ Free-text, unstructured data are distorted by comprehensive de-identification (primary and secondary PHI) ▶ Whole patient population accessible to research 	<ul style="list-style-type: none"> ▶ Limited to claims data only (eg, no laboratory results) ▶ Claims data may be distorted by techniques that try to mitigate re-identification risk (data suppression, generalization and perturbation) ▶ No clinical documents ▶ Whole patient population accessible to research 	<ul style="list-style-type: none"> ▶ Detailed EHR data ▶ Smaller de-identification distortion of free-text data in clinical documents (removal of only secondary PHI of living people) ▶ dsIDR research protocol allows consolidation of data from multiple payers and providers (after death data retrieval) ▶ Smaller sample size; only deceased patients

dsIDR, deceased subject integrated data repository; EHR, electronic health record; IRB, institutional review board; PHI, protected health information.

Box 1 Summary of immediate regulatory and policy challenges

- ▶ Rules clarifying necessary documents proving death status (when exercising HIPAA decedent clause)
- ▶ Obtaining EHR data after death from health plans and small providers (financial remuneration issues, authorizing a research team to be HIPAA 'personal representative')
- ▶ Ability of next of kin to donate EHR of deceased family members, disagreement management
- ▶ Clarification of legal claims resulting from secondary PHI against researchers using the decedent clause
- ▶ Preventing destruction of the EHR when minimum record retention period elapsed

EHR, electronic health record; HIPAA, Health Insurance Portability and Accountability Act; PHI, protected health information.

1. *Proving death status:* Many informatics projects benefit from the largest possible sample size and thus the potential dsIDR should try to capture all deceased patients. Maintaining a reliable record of all deceased patients is a key infrastructure consideration for a dsIDR. A centralized informatics infrastructure would greatly simplify the need for individual researchers intending to use the HIPAA decedent clause to prove deceased status of all patients involved in the research question. The decedent research clause places the burden of proof of death on the data requestor rather than on the holder of the data (eg, a health insurance company). Death status can be obtained using local sources, such as death observed during a hospital stay. A 2007 AHRQ study indicates that around 32% of deaths occur in hospitals. The remaining 68% have to be captured from external sources, such as nursing homes or health plan data, or national sources, such as the Social Security Death Master File (DMF). However, the DMF may be incomplete or contain false-positive entries. A study in 2001 showed that the DMF contains 94.5% of deaths of people aged ≥ 65 .¹⁸ A government audit in 2008 estimated that the DMF true-positive rate is 99.97%. HIPAA does not state whether DMF data are sufficient proof of death for researchers using the decedent clause.

Creating an informatics infrastructure that would support researchers' ability to prove deceased status when requesting their EHR records has significant costs. The official DMF copy distributed by the National Technical Information Service costs US\$1825 in initial costs plus US\$3650 for quarterly updates.¹⁹ The National Death Index from the National Center for Health Statistics, which is another source of death data, can be even more expensive.²⁰ Many healthcare institutions—for example, members of the Health Maintenance Organization Research Network, have mechanisms for importing externally provided death status data. Since the DMF represents a public resource, possibly made more widely available owing to the Freedom of Information Act,²¹ we envision national efforts to streamline the determination of death status, perhaps by the Clinical and Translational Science Award consortium or other research networks.

Health plans (eg, Medicare advantage plans) that clearly register lack of monthly premium for an extended period of

time (eg, 12 months) are also reliable sources of death status. This, combined with the absence of any unsubscribe request, is a very likely indicator of patient's death.

2. *Informed consent:* Traditionally, patient enrollment into patient registries and biobanks is accomplished with informed consent. Similarly, donation of corpses to research is governed by the 'declaration of intent' form defined by the Federal Uniform Anatomical Gift Act signed by either the patient while alive or by the surviving spouse or next of kin. The HIPAA decedent clause, however, enables research use of EHR data without such individual consent and with no rights of individuals to opt out of research use of their EHR post mortem. We refer to this situation as *decedent clause dsIDR research*. We envision this mode as the prevailing use case for most dsIDRs and such a repository would be much easier to create and administer. However, this research mode may be viewed by some privacy advocates as too permissive. To initiate a wider discussion, we also consider a second situation, in which a patient would somehow indicate, while they were alive, their willingness to donate their EHR after their death. In this case the dsIDR would, in addition to using the decedent clause, pursue formal individual informed consent and thus go beyond existing minimum legal requirements. We refer to this second scenario as *explicit consent dsIDR research*. Such explicit consent might be helpful during the collection of EHR data from multiple data holders post mortem.

For non-research access to an EHR after death, HIPAA defines the concept of a 'personal representative' as an executor, administrator or other person who has authority under applicable law (which varies by state) to act on behalf of the decedent. Following this provision, patients willing to donate their EHR post mortem may designate a researcher or a research team to be their personal representative. There are very few or no precedents of a research team approaching holders of EHR data (providers, payers, especially third-party entities unaffiliated with the research team institution) and acting as patient personal representative in order to retrieve the complete EHR record. Again, such a document would be analogous to a 'declaration of intent' defined under the Federal Uniform Anatomical Gift Act governing donation of a corpses to medical education. This Act allows surviving spouse or next of kin to make such donation on behalf of the deceased; however, they cannot reverse any intent expressed by the patient while alive, even if they disagree with it.

Management of online records after death is a recognized problem that is not limited to the healthcare domain. Some experts suggest the establishment of a 'digital executor' who is given a list of relevant online accounts. Creating a designated executor account is preferred to direct sharing of access credentials. For example, the terms of service of Microsoft's HealthVault personal health record (PHR) define a role of a record 'custodian'; however, the HealthVault terms²² do not specify whether custodian accounts are deactivated post mortem or how HealthVault would detect or verify a patient's death.

3. *Loss of data:* Another important consideration for dsIDR infrastructure is to consider the case in which the EHR data holder may delete data. In the era of big data, it is paradoxical to see data owners destroying records; however, providers may want to limit their legal liability by destroying such records at the earliest legal opportunity. Although some large academic medical centers understand the research value of EHR data, this may not be true for other large institutions or

small or rural providers. Minimum record-keeping requirements for EHRs vary by state and range between 5 and 20 years. After the January 2013 change to HIPAA, the ability to provide data to commercial entities is allowed 50 years after death. Some data holders may keep the data for potential use when the HIPAA limit of 50 years elapses, but others may destroy them. The regulator (Department of Health and Human Services) made clear that the 50-year boundary is not a recommended or mandatory retention period but simply a temporal privacy threshold. From a small-provider perspective, lack of any EHR record activity for 5–20 years (depending on the state) may be a reason to delete the whole EHR. Providers are not obliged to keep EHR records while the patient is alive. Neither are they motivated to seek the death status of all their patients.

We see a need for dsIDR researchers to advocate the value of EHR data and prevent such deletion. Another option is to use the patient as the transfer point of the data. The patient, while alive, would regularly import EHR data into a health record bank, a PHR or the dsIDR, and would be able to use a ‘donate EHR’ button. Post mortem, the data would be available for research use and no data would be lost because of deletion. In this situation the patient acts as an owner of a copy of the data. Inspection of the complete record before donation would clearly show how much primary and secondary private information it truly contains and can even express preference about the level of EHR donation (eg, diagnoses, laboratory results and medications but no clinical documents of a chosen type). We are limiting our focus on deceased patient data, but initiatives where patients can consent to continuous research use of their EHR data (during their life) in a clear opt-in fashion are relevant.²³

4. *Involvement of next of kin:* Similar to donation of corpses, questions about involvement of surviving spouses or children may arise. On the one hand, in some research projects, surviving children may be benefactors from the outcome of any potential research that uses their deceased parent’s data. On the other hand, next of kin or secondary PHI parties may want to limit the use of certain types of decedents’ unstructured clinical documents. In the explicit consent dsIDR scenario, there are two possibilities. If the patient declared an intention to donate, this decision should probably be respected. Policy decisions arise only if such a declaration does not exist and not all involved parties agree on the level of EHR research donation. In the decedent clause research mode, technically no opt-out option is given to the patient or next of kin.
5. *Patient adoption:* Finally, there is no prior estimate or pilot study that would indicate what percentage of patients would be willing to make a post mortem donation of their EHR data. Different patient groups may have radically different rates of donation (eg, patients who are concerned about healthy or privacy versus patients with cancer, or high-burden or heritable disease). We believe that increasing public awareness of the research value of EHR data and increased discussion about privacy policies post mortem is necessary. The most integrative and transparent approach would be an explicit consent via a ‘donate EHR button’ in the PHR combined with a multiple-source approach for obtaining death status.

CONCLUSION

We believe that the dsIDR as a research piloting platform that does not require IRB review has significant value and the number of deceased patients is guaranteed to grow. Having an explicit approach that is transparent to research participants is

an ethical imperative. Whereas for living subjects the society has ‘defaulted’ to a position of tight control of EHR data, even when evidence of harm from research using EHR data is scant, this is not completely true for deceased patient data. The HIPAA decedent clause represents a powerful and underused research option that simplifies ethical review and increases research access to EHR data. Deceased status simplifies some of the ethical and privacy concerns, especially the ability to federate and integrate data from multiple sources. Creation of a formal informatics infrastructure for decedents’ EHR data (based either on existing regulations or explicit consent) extends the portfolio of directly accessible tools available to clinical researchers and also contributes to the debate about privacy protection versus research benefits of IDRs, in general, and IRB review procedures of IDR-based research.

Acknowledgements We thank Laritza M Rodriguez for comments on a draft of this manuscript; Aaron Miller for supportive analyses of deceased patient ratios within an exemplary integrated data repository, and Simon Lin and John Hurdle for discussion comments.

Contributors VH initially drafted the idea of a deceased subject integrated data repository. Both authors (VH and JJC) jointly critically revised the final manuscript.

Funding This research was supported by the Intramural Research Program of the National Institutes of Health Clinical Center and the National Library of Medicine.

Competing interests VH and JJC are supported by the Intramural Research Program of the National Institutes of Health Clinical Center and the National Library of Medicine. The ideas and opinions expressed are those of the authors. The content of this publication does not necessarily reflect the views or policies of the US National Institutes of Health.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 Mackenzie SL, Wyatt MC, Schuff R, *et al.* Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. *J Am Med Inform Assoc* 2012;19:e119–24.
- 2 Hripscak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc* 2013;20:117–21.
- 3 Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. *Ann Intern Med* 2012;157:207–10.
- 4 Murdoch TB, DAS. The inevitable application of big data to health care. *JAMA* 2013;309:1351–2.
- 5 Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ* 2009;338:b81.
- 6 Neuhauser D, Morrissey M, Votruba M. Your human subjects review process: a road block or a competitive advantage? *J Clin Res Bioeth* 2012;3:130.
- 7 Brothers KB, Clayton EW. “Human non-subjects research”: privacy and compliance. *Am J Bioeth* 2010;10:15–17.
- 8 Columbia University: Research Data Explorer (RedX). <http://ps.columbia.edu/CERS/toolkit-exploratory-research>
- 9 Roden DM, Pulley JM, Basford MA, *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 2008;84:362–9.
- 10 El Emam K, Jonker E, Arbuckle L, *et al.* A systematic review of re-identification attacks on health data. *PLoS One* 2011;6:e28071.
- 11 Electronic Code of Federal Regulations: Title 45: §164.512 Uses and disclosures for which an authorization or opportunity to agree or object is not required. <http://www.ecfr.gov/cgi-bin/retrieveECFR?SID=0c83756f3a487d6d70dd3232e084c0a0&n=45y1.0.1.3.78.5&r=SUBPART&ty=HTML#45.1.0.1.3.78.5.27.8>
- 12 Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. 2006. <http://www.ncbi.nlm.nih.gov/pubmed/21413206>.
- 13 CMS Chronic Condition Data Warehouse: Technical Guidance Documentation. <http://www.cdwdata.org/web/guest/technical-guidance-documentation>.
- 14 Office of Civil Rights: Guidance Regarding Methods for De-identification of PHI in Accordance with HIPAA. 2012. http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveridentities/De-identification/hhs_deid_guidance.pdf.
- 15 eRA Commons Overview. http://era.nih.gov/about_era/overview_eRA.cfm
- 16 How Many People Have Ever Lived on Earth? <http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx>.
- 17 Curtin C. Do living people outnumber the dead? *Sci Am* 2007;297:126.
- 18 Hill MA, Rosenwaite I. The social security administration’s death master file: the completeness of death reporting at older ages. *Soc Security Bull* 2001;64:45.

- 19 Office of Inspector General: Personally Identifiable Information Made Available to the General Public via the Death Master File (June 2008) (A-06-08-18042). http://web.timesrecordnews.com/online_forms/Deathfile.pdf
- 20 Lash TL, Silliman RA. A comparison of the National Death Index and Social Security Administration databases to ascertain vital status. *Epidemiology* 2001;12:259–61.
- 21 Publicly available and legally posted Social Security Death Master File (unmodified November 2011 copy). <http://ssdmf.info>
- 22 HealthVault Terms of Service. 2013. <https://account.healthvault.com/help.aspx?topicid=serviceagreement>
- 23 Critchley CR, Nicol D, Otlowski MF, *et al.* Predicting intention to biobank: a national survey. *Eur J Public Health* 2012;22:139–44.