Healthcare Records Need to Be Preserved

January 2026

James L. Weber, PhD

Department of Pediatrics

University of Wisconsin School of Medicine and Public Health

Madison, WI 53792

U.S. healthcare providers have made excellent progress over the last few decades in converting from paper to electronic healthcare records. In 2021, it was reported that 88% of American providers were using electronic health records (1), and the fraction is likely higher today. But there is still a big problem with retention of these records. In this opinion article, I make the case for permanent retention of all essential healthcare (also called medical) records.

Current U.S. regulations require healthcare records to be retained for surprisingly short times. With the exception of Massachusetts (20 years), individual state laws require retention of adult records after the patient has left a healthcare organization for only 5-11 years (2, 3). In some states, records for children must be retained until the children reach adulthood. No state requires indefinite retention. Some healthcare organizations are retaining records for longer than the minimum required times, but others are not (4). Since storage of records incurs cost, and since retention may increase liability risks, as long as retention requirements are short, many organizations will continue to destroy records.

Why Healthcare Records Should be Retained

Permanent retention of healthcare records, even after patients die, will improve both direct patient care and biomedical research. Regarding direct patient care, records of health events early in life can improve the effectiveness and decrease the cost of care later in life. Although the effects are modest, sharing of health records among organizations has been reported to reduce hospital readmissions, length of hospital stay, duplicate testing, medication errors, and to improve diagnoses (5-8). As an example of the value of retaining records, it's important to know whether a woman of reproductive age had measles or a measles vaccine when she was young, since measles during pregnancy entails substantial risks for both mother and fetus (9, 10). As another example, a person's genomic DNA sequence does not change over their lifetime and hence should need to be generated only once. But when this information, which is extremely useful for diagnosis, disease prevention, reproductive planning, and drug prescription, is destroyed, the only recourse is expensive resequencing.

There is great potential for preventing and ameliorating disease through the identification of individuals who are at particularly high risk (11). Such individuals can then be targeted for enhanced screening and preventive measures. Family history is one of the most important factors involved in estimating risk (12). Self-report of family history is often incomplete and/or inaccurate (13-15). Far better is to obtain disease information on family members through direct examination of their healthcare records. In clinical genetics, DNA sequence information from family members is useful in identifying de novo mutations, verifying or constructing family trees, determining the chromosomal phase of variants, and evaluating the clinical significance of variants (16). When records of family members, including deceased family members, are destroyed, this valuable information is lost forever.

Record retention also accelerates biomedical research. The digitization of healthcare records has enabled an explosion of research studies that use healthcare records (17). Among other applications, electronic record research has advanced our understanding of disease course and features, disease prevalence, treatment efficacy, healthcare delivery, adverse drug reactions, and drug development (18-21). However, all of this research is impeded by holes in the records due to record destruction.

Genetics research is another area that greatly benefits from record retention (22). The remarkable research successes of the Icelandic (23), UK Biobank (24), All of Us (25), and other databases that combine healthcare records with genomic information demonstrates the power of this approach. Combining clinical and genomic information catalyzes the development of algorithms that predict patient's disease risk (11). It also allows identification of DNA sequence variants that differ in frequency between cases and controls, and allows determination of the phenotypes that are associated with specific DNA variants.

Many DNA variants that impact human health have low population frequencies, often less than 1 per 10,000 individuals. Therefore, as shown in the table, databases must be quite large to establish the phenotypes of people who carry these rare variants. Databases can become large by combining records from different healthcare organizations and nations, but also by retaining records over multiple generations. The information within current healthcare records is valuable now for variant (and disease) characterization, and will continue to be valuable 1,000 years from now.

Database Sizes Needed to Characterize Rare Genotypes

	Required Database Sizes to Include 20 Individuals with the Indicated Genotypes (millions of people)	
Variant Allele Frequency	Single Variant	Two Variants*
1.0%	0.001	0.2
0.1%	0.01	20
0.01%	0.1	2,000
0.001%	1.0	200,000

* Assumes random mating and unlinked variants

How to Retain Healthcare Records

Retention of healthcare records in the U.S. is especially challenging because there is no national healthcare system and because the records are siloed among many separate hospitals, clinics, small practices, labs, and other organizations. Nevertheless, some progress toward permanent healthcare record retention has been achieved. Many U.S. healthcare organizations now participate in state or regional health information exchanges (also sometimes called Health Data Utilities) (5, 7). These exchanges permit records from multiple organizations for individual patients to be combined. In some cases, records from earlier in life may be saved from destruction by combining with records from later in life. Efforts are also underway to merge the state and regional health information exchanges into a truly national exchange (27, 28). In addition to the exchanges, clinical data warehouses that contain select patient data have been established to facilitate research (8, 29). Some data may be retained in the warehouses that would otherwise be lost through record destruction.

However, the record exchanges and the data warehouses are severely limited for the purpose of record retention. Not all healthcare organizations participate in the exchanges and not all data from participating organizations is contributed. Even when the exchanges save combined records, they are under little or no legal obligation to retain these records. We therefore need additional steps.

One such step is adoption of universal patient identification numbers. Universal identification numbers will make it easier to combine records from multiple organizations (30).

Another useful step would be to lengthen the individual state requirements for record retention to 20-30 years. This would prevent many records from being destroyed. Increasing state mandates to indefinite retention, however, would probably be impractical since many, and probably all, individual healthcare organizations will eventually cease operations.

The only way I can see to ensure permanent retention of healthcare records is to store the records in centralized, government-organized and government-regulated repositories (31, 32). All healthcare organizations would be required to send full copies of healthcare records for all patients to these centralized repositories. Perhaps existing health information exchanges could be adapted to this purpose; perhaps new repositories would need to be created.

There are many advantages to the centralized repositories. Complete cradle to grave health records would become available to providers, patients, and researchers. The healthcare efficiencies and cost savings realized by the sharing of records in the health information exchanges (see above) would also be attained by the centralized repositories. Information from family members could be combined. The availability of standardized data format would streamline research. The number of records in centralized databases would be larger than in individual organizations increasing the statistical power of the data. The centralized repositories would eliminate the need for at least many clinical data warehouses. Large centralized databases could also devote more resources to data security and privacy than many smaller healthcare organizations.

Financing centralized repositories is an obvious hurdle to their establishment. Although some taxpayer money will likely be required to at least initiate these repositories, the amounts need not be prohibitive. National healthcare record repositories of the type that I have outlined have already been established in Finland (33) and Singapore (34), and these nations spend less than half per capita on healthcare than in the U.S. Costs borne by the individual healthcare organizations in transmitting the data would be offset by eliminating the need for backing up the records. Even more importantly, the repositories could be partially, and perhaps completely, funded through carefully controlled sale of deidentified data to private companies (8). Many millions of deidentified healthcare records worth at least tens of billions of dollars are currently being bought and sold in the U.S. through established markets (35-37). Costs could also be held down by not retaining every byte of information. Expert panels could be convened to recommend retention times for different types of information. Essential information would need to be permanently retained, but other information might be able to be discarded after appropriate times without appreciable harm to patient care or research.

In the current U.S. political environment, it’s unlikely that a *national* healthcare record repository will be created, but individual U.S. states, or clusters of states can and should act (32). Eventually, data from multiple U.S. states and multiple countries can be merged.

Closing Remarks

People began constructing libraries thousands of years ago (38). The importance of retaining critical information was recognized by many. Unfortunately, very little of the information stored in the ancient libraries survived. Vast amounts of human progress, history, and literature were squandered. Today, we are doing better at retaining information, but still critical information, especially including healthcare information, is irretrievably lost. In my opinion, this is unacceptable waste.

We all benefit tremendously from modern healthcare. Average lifespans have approximately doubled over the last two centuries (39). Many people are alive today only because of modern healthcare. I think it is fair that we give back to an activity that provides us with so much. Through permanent retention of records, each patient and each provider will contribute meaningfully to the continued improvement of healthcare.

References

1. Centers for Disease Control, National Electronic Health Records Survey. Available at https://www.cdc.gov/nchs/data/nehrs/2021NEHRS-PUF-weighted-estimates-508.pdf. 2021.

2. Office of the National Coordinator for Health Information Technology, State record retention. Available at https://www.healthit.gov/sites/default/files/appa7-1.pdf. Accessed Nov. 2025.

3. Cariend, Medical records retention laws. Available at https://www.cariend.com/wp-content/uploads/2022/08/Cariend-Medical-Records-Retention-by-State.pdf. Current as of 2022.

4. L. A. Rinehart-Thompson, Record retention practices among the nation's "most wired" hospitals. *Perspect Health Inf Manag*. **5**, 8 (2008).

5. N. Menachemi, S. Rahurkar, C. A. Harle, J. R. Vest, The benefits of health information exchange: an updated systematic review. *J Am Med Inform Assoc*. **25**, 1259-1265 (2018).

6. R. Janakiraman, E. Park, E. M. Demirezen, S. Kumar, The effects of health information exchange access on healthcare quality and efficiency: an empirical investigation. *Management Science* **69**, 791-811 (2022).

7. S. Dupont, J. Nemeth, S. Turbow, Effects of health information exchanges in the adult inpatient setting: a systematic review. *J Gen Intern Med*. **38**, 1046-1053 (2023).

8. K. S. Williams, S. Rahurkar, S. J. Grannis, T. K. Schleyer, B. E. Dixon, Evolution of clinical health information exchanges to population health resources: a case study of the Indiana network for patient care. *BMC Med Inform Decis Mak*. **25**, 97 (2025).

9. A. Khalil, A. Samara, C. Campbell, S. N. Ladhani, Pregnant women and measles: we need to be vigilant during outbreaks. *EClinicalMedicine*. **72**, 102594 (2024).

10. M. Science, J. Hutson, M. Barton, Measles in pregnancy. *CMAJ*. **197**, E625 (2025).

11. Polygenic Risk Score Task Force of the International Common Disease Alliance, Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med*. **27**, 1876-1884 (2021).

12. N. Mars *et al*., Systematic comparison of family history and polygenic risk across 24 common diseases. *Am J Hum Genet*. **109**, 2152-2162 (2022).

13. A. Augustinsson, C. Ellberg, U. Kristoffersson, Å. Borg, H. Olsson, Accuracy of self-reported family history of cancer, mutation status and tumor characteristics in patients with early onset breast cancer. *Acta Oncol*. **57**, 595-603 (2018).

14. L. E. Hull, P. Natarajan, Self-rated family health history knowledge among All of Us program participants. *Genet Med*. **24**, 955-961 (2022).

15. A. Harris *et al.*, Incomplete Family History and Meeting Algorithmic Criteria for Genetic Evaluation of Hereditary Cancer. *JAMA Netw Open* **8**, e2539870, (2025).

16. J. L. Weber, Genomes and the Gene Pool. Ten Generations Press. (2026).

17. M. A. Gianfrancesco, N. D. Goldstein, A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Methodol*. **21**, 234 (2021).

18. J. Chishtie *et al.*, Use of Epic electronic health record system for health care research: scoping review. *J Med Internet Res*. **25**, e51003 (2023).

19. A. S. Tang *et al.*, Harnessing EHR data for health research. *Nat Med*. **30**, 1847-1855 (2024).

20. E. V. Minikel, M. R. Nelson, Human genetic evidence enriched for side effects of approved drugs. *PLoS Genet*. **21**, e1011638 (2025).

21. Y. Li *et al.*, Cell-type-directed network-correcting combination therapy for Alzheimer's disease. *Cell* **188**, 5516-5534 (2025).

22. J. E. Linder, L. Bastarache, J. J. Hughey, J. F. Peterson, The role of electronic health records in advancing genomic medicine. *Annu Rev Genomics Hum Genet*. **22**, 219-238 (2021).

23. D. F. Gudbjartsson *et al.*, Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet*. **47**, 435-444 (2015).

24. V. Y. Tan, N. J. Timpson, The UK Biobank: a shining example of genome-wide association study science with the power to detect the murky complications of real-world epidemiology. *Annu Rev Genomics Hum Genet*. **23**, 569-589 (2022).

25. T. Dutka *et al.*, All of Us research program year in review: 2024. *Am J Hum Genet*. **112**, 1983-1987 (2025).

26. M. K. Kim, C. Rouphael, J. McMichael, N. Welch, S. Dasarathy, Challenges in and opportunities for electronic health record-based data analysis and interpretation. *Gut Liver* **18**, 201-208 (2024).

27. A. B. Abbasi, L. H. Curtis, R. M. Califf, The promise of real-world data for research - what are we missing? *N Engl J Med*. **393**, 318-321 (2025).

28. J. Everson, W. Chang, V. Patel, J. Adler-Milstein, The state of health information organizations and plans to participate in the federal exchange framework. *Health Aff Sch*. **2**, qxae098 (2024).

29. Z. Wang *et al.*, Clinical data warehousing: a scoping review. *J Soc Clin Data Manag.* **4**, (2024).

30. Patient ID Now. Available at https://patientidnow.org/. Accessed Nov. 2025.

31. E. J. Topol, The big medical data miss: challenges in establishing an open medical resource. *Nat. Rev. Genet*. **16**, 253–254 (2015).

32. A. Burt, D. Boffa, State governments could revolutionize health-care data. *Nature* **623**, 916 (2023).

33. V. Männikkö *et al*., Overview of Finnish national patient data repository for research on medical risk assessment. *Finnish Journal of eHealth and eWelfare* **16**, 254-268 (2024).

34. C. Chen *et al*., Economic evaluation of the next generation electronic medical records in Singapore: cost-utility analysis. *J Med Internet Res*. **27**, e70484 (2025).

35. A. Tanner, Our Bodies, Our Data. 2017. Beacon Press.

36. K. D. Mandl, E. D. Perakslis, HIPAA and the leak of "deidentified" EHR data. *N. Engl. J. Med*. **384**, 2171–2173 (2021).

37. A. Telenti, X. Jiang, Treating medical data as a durable asset. *Nat Genet*. **52**, 1005-1010 (2020).

38. A. Pettegree, A. der Weduwen, The Library. Basic Books. 2021.

39. J. Oeppen, J. W. Vaupel, Broken limits to life expectancy. *Science* **296**, 1029-1031 (2002).