Expert's De-identification Assessment of National Association of Epilepsy Centers Surgical Data Collection Program

Prepared by Privasense, LLC
Evaluation as of October 30, 2025

To whom correspondence should be addressed:
Bradley Malin, Ph.D.*
brad.malin@gmail.com

*This document reflects the work and viewpoint of the preparer of this work and is representative of neither his primary employer (Vanderbilt University Medical Center of Nashville, TN) nor of any granting agency through which his work at his primary employer is supported, including the National Institutes of Health, the National Science Foundation, or the Advanced Research Projects Agency for Health.

Document History

Version	Effective Date	Summary of Reasons for Changes
1	October 30, 2025	Original

Table of Contents

1. Introduction

A non-profit association, the National Association of Epilepsy Centers[1] (NAEC) assures quality epilepsy care by accrediting level 3 and level 4 specialized epilepsy centers in the US. NAEC plans to initiate a surgical data collection program in 2026 to collect data on epilepsy surgical procedures and to track corresponding patient outcome data. Through this program, NAEC will receive data from various covered entities, who are subject to oversight by the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA). The NAEC aims to receive data from such covered entities in a manner that meets the de-identification definition of the HIPAA Privacy Rule, which is overseen by the Office for Civil Rights at the U.S. Department of Health and Human Services (HHS).

The HIPAA Privacy Rule provides a definition of de-identified data, as well as two implementation options to meet the definition: i) Safe Harbor and ii) Expert Determination. Due to the fact that de-identified data is no longer subject to the Privacy Rule (or HIPAA more generally), it can be processed without oversight by HHS. NAEC hired this de-identification Expert to ensure that the data would be transformed in a manner that satisfies the HIPAA Expert Determination model. As will be shown, all of the data received by NAEC is consistent with the Safe Harbor implementation, save for a unique ID that is used for tracking and updating a patient's record over time. This report documents why this ID is considered to be de-identified as well.

The following sections of this document review relevant components of HIPAA, provide a high-level overview of the patient-derived data to be disseminated, and describe how the protections instituted for the data relate to the regulatory requirements of de-identification.

[1] https://naec-epilepsy.org/

2. Privacy Rules and Regulations

In the United States, there is no centralized legal statute for data protection. Rather, a patchwork of laws are tailored to oversee the handling and disclosure of specific types of personal data, such as the Graham-Leach-Bliley Act for financial data, the Federal Educational Rights and Privacy Act for student data, and HIPAA, which is the most pertinent to this report. Specifically, under HIPAA, the Privacy Rule protects all "individually identifiable health information" held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral. The Privacy Rule calls this information "protected health information" (PHI). By definition, "individually identifiable health information" is information, including demographic data that relates to:

- the individual's past, present or future physical or mental health or condition,
- the provision of health care to the individual, or
- the past, present, or future payment for the provision of health care to the individual,
- and that identifies the individual or for which there is a reasonable basis to believe it can be used to identify the individual.

Protected health information includes many common identifiers (e.g., name, address, birth date, Social Security Number), but there are also potential "quasi-identifiers" which may permit a recipient of the data to determine the identity of the corresponding subject.

When health information does not identify an individual, and there is no reasonable basis to believe that it can be used to identify an individual, it is said to be "de-identified" and is not protected by the Privacy Rule. More specifically, 45 C.F.R., section 164.514(a) of the Privacy Rule provides the standard for de-identification of individually identifiable health information:

> **§ 164.514 Other requirements relating to uses and disclosures of protected health information.**
> (a) *Standard: de-identification of protected health information.* Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information.

Section 164.514(b) of the Privacy Rule contains the implementation specifications that a covered entity, or affiliated business associate, must follow to meet the de-identification standard. In particular, the Privacy Rule outlines two routes by which health data can be designated as de-identified. The first route is the "Expert Determination" method.

(b) *Implementation specifications: requirements for de-identification of protected health information.* A covered entity may determine that health information is not individually identifiable health information only if:

(1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:

(i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and

(ii) Documents the methods and results of the analysis that justify such determination; or

The alternative de-identification implementation defined by the HIPAA Privacy Rule is what is typically referred to as the "Safe Harbor" method.

(2)(i) The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

(A) Names	
(B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census: (1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and (2) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000	
(C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older	
(D) Telephone numbers	(M) device identifiers and serial numbers
(E) Fax numbers	(N) Web Universal Resource Locators (URLs)
(F) Email addresses	(O) Internet Protocol (IP) addresses
(G) Social security numbers	(P) Biometric identifiers, including finger and voice prints
(H) Medical record numbers	
(I) Health plan beneficiary numbers	(Q) Full-face photographs and any comparable images

(J) Account numbers	(R) Any other unique, identifying number, characteristic, or code
(K) Certificate / license numbers	
(L) vehicle identifiers and serial numbers, including license plate numbers	

(ii) The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information

The term (R) is particularly notable, as this refers on to §164.514(c), which defines the notion of a re-identification code. It should be noted that the term "re-identification" has been used to indicate both a unique ID that can be used to track a record in a de-identified dataset, as well as commit an unapproved attack on a patient's privacy. In this report, we use the term as the regulation does and consider the tracking ID:

(c) Implementation specifications: re-identification. A covered entity may assign a code or other means of record identification to allow information de-identified under this section to be re-identified by the covered entity, provided that:

(1) Derivation. The code or other means of record identification is not derived from or related to information about the individual and is not otherwise capable of being translated so as to identify the individual; and

(2) Security. The covered entity does not use or disclose the code or other means of record identification for any other purpose, and does not disclose the mechanism for re-identification.

It is further important to recognize that while the re-identification provision does not permit assignment of a code or other means of record identification that is derived from identifying individual information, a covered entity may disclose such derived information if an expert determines that the data meets the de-identification requirements at §164.514(b)(1). This is particularly the case if the resulting information cannot be translated to identify the individual. Along these lines, the Office for Civil Rights [OCR 2012] stated that, to resolve confusion about what constitutes a code and how it relates to PHI, it was providing guidance similar to that from the National Institutes of Standards and Technology [McCallister 2010], which states:

De-identified information can be re-identified (rendered distinguishable) by using a code, algorithm, or pseudonym that is assigned to individual records. The code, algorithm, or pseudonym should not be derived from other related information* about the individual, and the means of re-identification should only be known by authorized parties and not disclosed to anyone without the authority to re-identify

> records. A common de-identification technique for obscuring PII [Personally Identifiable Information] is to use a one-way cryptographic function, also known as a hash function, on the PII.
>
> *This is not intended to exclude the application of cryptographic hash functions to the information.

Thus, codes derived from PHI as part of a de-identified data set may be disclosed if an expert determines that the data meets the de-identification requirements at §164.514(b)(1).

3. Protection Model for Structured Data

The set of data fields that will be provided to NAEC for the surgical data collection program are shown in Table 1. All of the fields that are labeled as "retained as is" (shaded in green) are compliant with Safe Harbor regardless of the values they communicate. All of the fields that are shaded in yellow are limited in some manner to be consistent with the de-identification requirements. For the two age fields, "age at diagnosis (years)" and "age at surgery (years)", these fields are limited to ages that are no greater than 89, after which they are reported as a single label indicative of an 89+ age group.

Table 1. Data fields sent to NAEC and their relationship to the HIPAA Safe Harbor implementation of the Privacy Rule.

Data Field	Management to Ensure Safe Harbor Compliance
NAEC Patient ID	Unique ID confirmed to not be used for purposes beyond de-identified data management
Race	Retained as is
Ethnicity	Retained as is
Sex	Retained as is
Primary Payer	Retained as is
Epilepsy Type	Retained as is
Etiology (multiselect)	Retained as is
Structural etiology (multiselect)	Retained as is
Seizure types (multiselect)	Retained as is
Syndrome	Retained as is
Referral to level 4	Retained as is
Previous treatment surgery (including neuromodulation)	Retained as is
Age at Diagnosis (years)	Top recoded as 89+
Age at Surgery (years)	Top recoded as 89+
Year of Surgery	Retained as is
Intent of Surgery	Retained as is
Intracranial EEG (multiselect)	Retained as is
Intraoperative ECoG at resection	Retained as is
Location of treatment or target (multiselect)	Retained as is
Type of resection, disconnection, or ablation	Retained as is
Type of device (multiselect)	Retained as is
Complications (multiselect)	Retained as is
12-month Engel Outcome	Retained as is
12-month ILAE Score	Retained as is
24-month Engel Outcome	Retained as is

24-month ILAE Score	Retained as is
Comments (see instructions)	Retained as is (confirmed to be clinical information only)

For the "NAEC Patient ID", it was confirmed that this value is not used for any purposes outside of the communication of information about a record between a covered entity and the de-identified dataset at NAEC and for NAEC to be able to link data on the same patient. Moreover, it was confirmed that this value is not correlated in any way with patient identifiers, such that, in the event the value was to be disclosed to some entity beyond NAEC, the value would not be meaningful in its own right, nor would it be reasonable to assume that it could be linked back to a patient identity without the assistance of the originating covered entity. Thus, it was determined that this unique ID met the requirements for a re-identification code under the HIPAA Privacy Rule and that it could be included in a de-identified dataset at NAEC that was updated over time by a covered entity.

4. Determination

It is my determination that the data sent by covered entities to the National Association of Epilepsy Centers, as described in this document, meets the de-identification requirements set forth in the HIPAA Privacy Rule and other related laws and regulations. This is inclusive of the unique ID, referred to as the NAEC Patient ID, that is used by a covered entity to update a de-identified record at NAEC over time. This determination is based on the representation of the NAEC data management practices as of October 30, 2025 and will be considered void in the event that these practices deviate in a manner that would render the data into a less protected state.

DocuSigned by:
Bradley Malin
3451567A92564D6...

10/30/2025

Bradley A. Malin, Ph.D. Date

Expert Determination Consultant

Privasense, LLC

5. References

[McCallister 2010] McCallister E, Grance T, Scarfone K. Guide to protecting the confidentiality of personally identifiable information (PII): recommendations of the National Institute of Standards and Technology. Special Publication 800-122, National Institute of Standards and Technology. 2010.

[OCR 2012] Office for Civil Rights. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. U.S. Department of Health and Human Services. November 2012. Available at: http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html. Last accessed: October 26, 2025.

6. Information About the Expert

Bradley Malin, Ph.D. is the founder and principal consultant at Privasense, LLC. He is currently a Professor of Biomedical Informatics, Biostatistics, and Computer Science at Vanderbilt University, where he founded and currently directs the Vanderbilt Health Information Privacy Laboratory. He is an internationally recognized expert on data privacy and has served on national advisory committees for the National Academy of Medicine at the National Academy of Sciences regarding the management of data from electronic medical record systems and biorepositories, as well as the Technical Anonymisation Group of the European Medicines Agency. His research has been published through over two hundred peer-reviewed articles, portions of which have been cited in the Federal Register, Congressional briefings, and popular media outlets such as Nature News, Scientific American, and MIT Technology Review. Among various honors, he received the prestigious Presidential Early Career Award for Scientists and Engineers (PECASE) is an elected fellow of the American College of Medical Informatics (ACMI), International Academy of Health Sciences Informatics (IAHSI), American Institute for Medical and Biological Engineering (AIMBE), the Institute for Electrical and Electronics Engineers (IEEE), and the National Academy of Medicine (NAM). He received a bachelor's in biological sciences, master's in public policy and management, and doctorate in computer science, all from Carnegie Mellon University. Additional information can be found at http://www.hiplab.org/people/malin.

Dr. Malin has consulted on de-identification for HIPAA compliance for numerous companies, governmental agencies, and academic medical centers, including, but not limited to Abridge, Amazon, Ambience Health, Briya, Amerisource Bergen, Boston Scientific, Caris Life Sciences, ConcertAI, COTA Healthcare, CVS, Definitive Health, Duke University Medical Center, Exact Sciences, GE Healthcare, Guardant Health, HealthVerity, Illinois Healthcare Association, Intuitive Surgical, Komodo Health, IQVIA, LexisNexis, Loopback Analytics, Massachusetts Center for Health Information and Analysis, Mayo Clinic, Memorial Sloan Kettering Cancer Center, Merative (formerly IBM Watson Health), Natera, Quest Diagnostics, PROCEPT BioRobotics, Prognos AI, PurpleLab, Sanofi-Aventis, SiteRx, Tempus AI, United Healthcare, Verana Health, and

Walgreens Boots Alliance. From 2009-2013, he served as a consultant to the Office for Civil Rights at the U.S. Department of Health and Human Services, where he assisted in the drafting of the agency's HIPAA de-identification guidance. He was also a co-founder of HealthDataLink, a company commercializing privacy-preserving record linkage software, which was acquired by Datavant.