Brought to you by Ronstadt Insurance, Inc.

HHS Provides Guidance on Methods for De-identifying PHI

On Nov. 26, 2012, the Department of Health and Human Services (HHS) released technical guidance and a related webpage regarding methods for de-identification of protected health information (PHI) in accordance with the HIPAA Privacy Rule. HHS identified the **Expert Determination Method** and **Safe Harbor Method** as the two available methods for satisfying the Privacy Rule's de-identification standard.

HHS' guidance explains both de-identification methods and provides information in a question and answer (Q&A) format to help covered entities, such as health plans, understand the available options for performing de-identification.

DE-IDENTIFICATION STANDARD

The HIPAA Privacy Rule protects individually identifiable health information by permitting only certain uses and disclosures of PHI. Because health information can be useful even when it is not individually identifiable (for example, in comparative effectiveness studies, policy assessments or life sciences research) the Privacy Rule permits a covered entity to create information that is not individually identifiable by following a de-identification standard.

The Privacy Rule allows a covered entity to freely use and disclose information that neither identifies nor provides a reasonable basis to identify an individual. The Privacy Rule's standard for de-identifying PHI recognizes the following de-identification methods:

- A formal determination by a qualified expert (Expert Determination Method); or
- The removal of specified individual identifiers as well as absence of actual knowledge by the covered entity that the remaining information could be used alone or in combination with other information to identify the individual (Safe Harbor Method).

Regardless of the de-identification method used, the Privacy Rule does not restrict the use or disclosure of de-identified health information, as it is no longer considered PHI.

EXPERT DETERMINATION METHOD

Under the Expert Determination Method for de-identifying PHI, a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:

- Applies these principles and methods and 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
- Documents the methods and results of the analysis that justify this determination.

HHS' guidance contains numerous Q&As explaining the Expert Determination Method, some of which provide very technical guidance. The following paragraphs contain components of this guidance that health plan sponsors may find useful. The full Q&As are available on the HHS' website.

HHS Provides Guidance on Methods for De-identifying PHI

Who is an "expert?"

There is no specific professional degree or certification program for designating who is an expert at rendering health information de-identified. From an enforcement perspective, HHS' Office for Civil Rights (OCR) would review the relevant professional experience and academic or other training of the expert used by the covered entity, as well as actual experience of the expert using health information de-identification methodologies.

What is an acceptable level of identification risk for an expert determination?

There is no explicit numerical level of identification risk that is deemed to universally meet the "very small" level indicated by the method. The ability of a recipient of information to identify an individual (that is, subject of the information) is dependent on many factors, which an expert will need to take into account while assessing the risk from a data set.

How long is an expert determination valid for a given data set?

The Privacy Rule does not explicitly require that an expiration date be attached to the determination that a data set, or the method that generated the data set, is de-identified information. However, certain de-identification practitioners use the approach of time-limited certifications. Information that had previously been de-identified may still be adequately de-identified when the certification limit has been reached. Covered entities will need to have an expert examine whether future releases of the data to the same recipient (for example, monthly reporting) should be subject to additional or different de-identification processes consistent with current conditions to reach the very low risk requirement.

Can an expert derive multiple solutions from the same data set for a recipient?

Yes. Experts may design multiple solutions, each of which is tailored to the covered entity's expectations regarding information reasonably available to the anticipated recipient of the data set. In these cases, the expert must take care to ensure that the data sets cannot be combined to compromise the protections set in place through the mitigation strategy. (Of course, the expert must also reduce the risk that the data sets could be combined with prior versions of the de-identified data set or with other publically available data sets to identify an individual.)

How do experts assess the risk of identification of information?

No single universal solution addresses all privacy and identifiability issues. Rather, a combination of technical and policy procedures are often applied to the de-identification task. OCR does not require a particular process for an expert to use to reach a determination that the risk of identification is very small. However, the Rule does require that the methods and results of the analysis that justify the determination be documented and made available to OCR upon request.

The determination of identification risk can be a process that consists of the following series of steps.

- The expert evaluates the extent to which the health information can (or cannot) be identified by the anticipated recipients.
- The expert often will provide guidance to the covered entity or business associate on which statistical or scientific methods can be applied to the health information to mitigate the anticipated risk.
- The expert will then execute the methods as deemed acceptable by the covered entity or business associate data managers (that is, the officials responsible for the design and operations of the covered entity's information systems).
- The expert will evaluate the identifiability of the resulting health information to confirm that the risk is no more than very small when disclosed to the anticipated recipients.

The process may require several iterations until the expert and data managers agree upon an acceptable solution. Regardless of the process or methods employed, the information must meet the very small risk specification requirement.

This Ronstadt Insurance, Inc. Legislative Brief is not intended to be exhaustive nor should any discussion or opinions be construed as legal advice. Readers should contact legal counsel for legal advice.

Design © 2012 Zywave, Inc. All rights reserved.

What are the approaches by which an expert assesses the risk that health information can be identified?

The de-identification standard does not mandate a particular method for assessing risk. A qualified expert may apply generally accepted statistical or scientific principles to compute the likelihood that a record in a data set is expected to be unique, or linkable to only one person, within the population to which it is being compared.

The expert may consider different measures of "risk," depending on the concern of the organization looking to disclose information. The expert will attempt to determine which record in the data set is the most vulnerable to identification. However, in certain instances, the expert may not know which particular record to be disclosed will be most vulnerable for identification purposes. In this case, the expert may attempt to compute risk from several different perspectives.

What are the approaches by which an expert mitigates the risk of identification of an individual in health information?

The Privacy Rule does not require a particular approach to mitigate, or reduce to very small, identification risk. The following provides a survey of potential approaches. An expert may find all or only one appropriate for a particular project, or may use another method entirely.

- A first class of identification risk mitigation methods corresponds to **suppression techniques**. These methods remove or eliminate certain features about the data prior to dissemination.
- A second class of methods that can be applied for risk mitigation are based on **generalization** (sometimes referred to as abbreviation) of the information. These methods transform data into more abstract representations.
- A third class of methods that can be applied for risk mitigation corresponds to **perturbation**. In this case, specific values are replaced with equally specific, but different, values.

Using these methods, the expert will prove that the likelihood an undesirable event (for example, future identification of an individual) will occur is very small.

There are many different disclosure risk reduction techniques that can be applied to health information. However, it should be noted that there is no particular method that is universally the best option for every covered entity and health information set. Each method has benefits and drawbacks with respect to expected applications of the health information, which will be distinct for each covered entity and each intended recipient. The determination of which method is most appropriate for the information will be assessed by the expert on a case-by-case basis and will be guided by input of the covered entity.

Finally, as noted in the preamble to the Privacy Rule, the expert may also consider the technique of limiting distribution of records through a **data use agreement or restricted access agreement** in which the recipient agrees to limits on who can use or receive the data, or agrees not to attempt identification of the subjects. Of course, the specific details of such an agreement are left to the discretion of the expert and covered entity.

Can an expert determine a code derived from PHI is de-identified?

There has been confusion about what constitutes a code and how it relates to PHI. A covered entity may disclose codes derived from PHI as part of a de-identified data set if an expert determines that the data meets the de-identification requirements of the HIPAA Privacy Rule.

Must a covered entity use a data use agreement when sharing de-identified data to satisfy the Expert Determination Method?

No. The Privacy Rule does not limit how a covered entity may disclose information that has been de-identified. However, a covered entity may require the recipient of de-identified information to enter into a data use agreement to access files with known disclosure risk, such as is required for release of a limited data set under the Privacy Rule. This agreement may contain a number of clauses designed to protect the data, such as prohibiting re-identification. Of

This Ronstadt Insurance, Inc. Legislative Brief is not intended to be exhaustive nor should any discussion or opinions be construed as legal advice. Readers should contact legal counsel for legal advice.

Design © 2012 Zywave, Inc. All rights reserved.

course, the use of a data use agreement does not substitute for any of the specific requirements of the expert determination method. Further information about data use agreements can be found on the OCR website. Covered entities may make their own assessments whether such additional oversight is appropriate.

SAFE HARBOR METHOD

Under the Safe Harbor Method for de-identification, a covered entity is protected if it does not have actual knowledge that the information could be used alone or in combination with other information to identify the subject of the information and specific identifiers are removed, including:

- Names, telephone numbers and social security numbers;
- All geographic subdivisions smaller than a state;
- All dates, except years, that are directly related to an individual; and
- Any other unique identifying number, characteristic or code, except those that are otherwise permitted by the Privacy Rule.

HHS' guidance contains numerous Q&As explaining the Safe Harbor Method, some of which provide very detailed guidance and examples. The following paragraphs contain components of this guidance that health plan sponsors may find useful. The full Q&As are available on HHS' website.

When can ZIP codes be included in de-identified information?

Covered entities may include the first three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census:

- The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; or
- The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000.

The Bureau of the Census provides information regarding population density in the United States. Covered entities are expected to rely on the most current publicly available Bureau of Census data regarding ZIP codes. This information can be downloaded from, or queried at, the American Fact Finder website.

May parts or derivatives of any of the listed identifiers be disclosed consistent with the Safe Harbor Method?

No. For example, a data set that contained patient initials, or the last four digits of a Social Security number, would not meet the requirement of the Safe Harbor Method for de-identification.

What are examples of dates that are not permitted according to the Safe Harbor Method?

Elements of dates that are not permitted for disclosure include the day, month and any other information that is more specific than the year of an event. For instance, the date "Jan. 1, 2009" could not be reported at this level of detail. However, it could be reported in a de-identified data set as "2009".

Many records contain dates of service or other events that imply age. Ages that are explicitly stated or implied as over 89 years old must be recoded as 90 or above. For example, if the patient's year of birth is 1910 and the year of healthcare service is reported as 2010, then in the de-identified data set the year of birth should be reported as "on or before 1920." Otherwise, a recipient of the data set would learn that the age of the patient is approximately 100.

Can dates associated with test measures for a patient be reported in accordance with Safe Harbor?

This Ronstadt Insurance, Inc. Legislative Brief is not intended to be exhaustive nor should any discussion or opinions be construed as legal advice. Readers should contact legal counsel for legal advice.

Design © 2012 Zywave, Inc. All rights reserved.

No. Dates associated with test measures, such as those derived from a laboratory report, are directly related to a specific individual and relate to the provision of health care. Dates such as these are PHI. As a result, no element of a date (except as described above) may be reported to adhere to the Safe Harbor Method.

What constitutes "any other unique identifying number, characteristic, or code" with respect to the Safe Harbor Method of the Privacy Rule?

This category corresponds to any unique features that are not explicitly enumerated in the Safe Harbor list, but could be used to identify a particular individual. Thus, a covered entity must ensure that a data set stripped of the explicitly enumerated identifiers also does not contain any of these unique features. The following are examples of these features:

- *Identifying Number* - There are many potential identifying numbers. For example, the preamble to the Privacy Rule noted that "Clinical trial record numbers are included in the general category of "any other unique identifying number, characteristic or code."
- *Identifying Code* - A code corresponds to a value that is derived from a non-secure encoding mechanism. For instance, a code derived from a secure hash function without a secret key (for example, "salt") would be considered an identifying element. This is because the resulting value would be susceptible to compromise by the recipient of the data. As another example, an increasing quantity of electronic medical record and electronic prescribing systems assign and embed barcodes into patient records and their medications. These barcodes are often designed to be unique for each patient, or event in a patient's record, and thus can be easily applied for tracking purposes.
- *Identifying Characteristic* - A characteristic may be anything that distinguishes an individual and allows for identification. For example, a unique identifying characteristic could be the occupation of a patient, if it was listed in a record as "current President of State University." Many questions have been received regarding what constitutes "any other unique identifying number, characteristic or code" in the Safe Harbor approach. Generally, a code or other means of record identification that is derived from PHI would have to be removed from data de-identified following the Safe Harbor Method.

What is "actual knowledge" that the remaining information could be used either alone or in combination with other information to identify an individual who is a subject of the information?

In the context of the Safe Harbor Method, actual knowledge means clear and direct knowledge that the remaining information could be used, either alone or in combination with other information, to identify an individual who is a subject of the information. This means that a covered entity has actual knowledge if it concludes that the remaining information could be used to identify the individual. The covered entity, in other words, is aware that the information is not actually de-identified information. Examples illustrating when a covered entity would fail to meet the "actual knowledge" provision are available on HHS' website under this Q&A.

If a covered entity knows of specific studies about methods to re-identify health information or use de-identified health information alone or in combination with other information to identify an individual, does this necessarily mean a covered entity has actual knowledge under the Safe Harbor Method?

No. Much has been written about the capabilities of researchers with certain analytic and quantitative capacities to combine information in particular ways to identify health information. A covered entity may be aware of studies about methods to identify remaining information or using de-identified information alone or in combination with other information to identify an individual. However, a covered entity's mere knowledge of these studies and methods, by itself, does not mean it has "actual knowledge" that these methods would be used with the data it is disclosing. OCR does not expect a covered entity to presume these capacities of all potential recipients of de-identified data. This would not be consistent with the intent of the Safe Harbor Method, which was to provide covered entities with a simple method to determine if the information is adequately de-identified.

This Ronstadt Insurance, Inc. Legislative Brief is not intended to be exhaustive nor should any discussion or opinions be construed as legal advice. Readers should contact legal counsel for legal advice.

Design © 2012 Zywave, Inc. All rights reserved.

HHS Provides Guidance on Methods for De-identifying PHI

Must a covered entity suppress all personal names, such as physician names, from health information for it to be designated as de-identified?

No. Only names of the individuals associated with the corresponding health information (that is, the subjects of the records) and of their relatives, employers and household members must be suppressed. There is no explicit requirement to remove the names of providers or workforce members of the covered entity or business associate. At the same time, there is also no requirement to retain such information in a de-identified data set.

Beyond the removal of names related to the patient, the covered entity would need to consider whether additional personal names contained in the data should be suppressed to meet the actual knowledge specification. Additionally, other laws or confidentiality concerns may support the suppression of this information.

Must a covered entity use a data use agreement when sharing de-identified data to satisfy the Safe Harbor Method?

No. The Privacy Rule does not limit how a covered entity may disclose information that has been de-identified. However, nothing prevents a covered entity from asking a recipient of de-identified information to enter into a data use agreement, such as is required for release of a limited data set under the Privacy Rule. This agreement may prohibit re-identification. Of course, the use of a data use agreement does not substitute for any of the specific requirements of the Safe Harbor Method. Further information about data use agreements can be found on the OCR website. Covered entities may make their own assessments whether such additional oversight is appropriate.

Must a covered entity remove protected health information from free text fields to satisfy the Safe Harbor Method?

PHI may exist in different types of data in a multitude of forms and formats in a covered entity. This data may reside in highly structured database tables, such as billing records. Yet, it may also be stored in a wide range of documents with less structure and written in natural language, such as discharge summaries, progress notes and laboratory test interpretations. These documents may vary with respect to the consistency and the format employed by the covered entity.

The de-identification standard makes no distinction between data entered into standardized fields and information entered as free text (that is, structured and unstructured text)—an identifier listed in the Safe Harbor standard must be removed regardless of its location in a record if it is recognizable as an identifier.

Whether additional information must be removed falls under the actual knowledge provision; the extent to which the covered entity has actual knowledge that residual information could be used to individually identify a patient. Clinical narratives in which a physician documents the history and/or lifestyle of a patient are information rich and may provide context that readily allows for patient identification.

MORE INFORMATION

Detailed information about the Privacy Rule and how it protects the privacy of health information is available at: www.hhs.gov/ocr/privacy/hipaa/understanding/index.html.

Source: Department of Health and Human Services

This Ronstadt Insurance, Inc. Legislative Brief is not intended to be exhaustive nor should any discussion or opinions be construed as legal advice. Readers should contact legal counsel for legal advice.

Design © 2012 Zywave, Inc. All rights reserved.