TMA Privacy Office Information Paper

New Guidance on De-Identification Methods under the HIPAA Privacy Rule

HIPAA Privacy ◆ February 2013

PURPOSE

On November 26, 2012, the Department of Health and Human Services (HHS) Office for Civil Rights (OCR) released new guidance on methods for de-identifying protected health information (PHI) in accordance with the HIPAA Privacy Rule. Mandated under section 13424(c) of the Health Information Technology Economic and Clinical Health (HITECH) Act, this guidance provides useful clarifications, details and best practices on de-identification. The purpose of this paper is to provide a summary of this HHS / OCR guidance.

The guidance is available at http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html

SUMMARY

The HIPAA Privacy Rule provision addressed by the guidance is 45 CFR 164.514(a)-(c). DoD implemented this provision in paragraph C8.1 of the DoD Health Information Privacy Regulation, DoD 6025.18-R.

Health information is not protected by the HIPAA Privacy Rule if it does not identify an individual and there is no reasonable basis to believe that the information can be used to identify an individual. Thus HIPAA Privacy Rule compliance obligations are eliminated once health information is properly de-identified. The HIPAA Privacy Rule establishes two de-identification methods known as the Safe Harbor and Expert Determination methods and addresses re-identification of de-identified information.

With respect to both methods, the new guidance confirms that a covered entity may, but is not required to enter into a data use agreement with a data set recipient, similar to the data use agreements required for limited data sets. A data use agreement might, for example, prohibit re-identification or prohibit sharing of the de-identified data with any other party. The guidance cautions that a data use agreement may not substitute for satisfying the Safe Harbor or Expert Determination method.

In addition, the new guidance covers a variety of other points, as summarized in the following.

A. Safe Harbor Method

The Safe Harbor method requires deleting 18 specified identifiers from PHI. In addition, the covered entity must "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." 45 CFR 164(b)(2)(ii).

With respect to deleting the 18 specified identifiers, the new guidance addresses the following:

- *Zip code data*: The HIPAA Privacy Rule permits including the first three digits of zip codes in de-identified data only if the geographic area covered by all zip codes beginning with those three digits has a population greater than 20,000 or the zip codes for those areas are changed to 000 in the data set. For this purpose, the guidance explains the Zip Code Tabulation Area (ZCTA) concept, created by the Census Bureau. Covered entities must use updated information from the 2010 Decennial Census, once the Census Bureau makes it available.
- *Dates*: The new guidance explains how to handle dates that imply age.
- *Other identifiers*: One of the 18 specified identifiers that must be deleted from a data set is "any other unique identifying number, characteristic or code." The new guidance provides examples and clarifies that this requirement does not prevent the covered entity from creating codes for re-identification purposes.
- *Actual knowledge*: The new guidance provides examples of failure to satisfy the rule that the covered entity must not have "actual knowledge" of re-identifiability. The examples involve: occupation data; familial relation between an anticipated recipient of the data and an individual whose information is included in the data; rare clinical events; and knowledge that a data recipient has a readily available mechanism to identify an individual included in the data set.
- *Personal names*: The new guidance clarifies that not all personal names (e.g. physician names) need to be deleted.
- *Free text data*: The new guidance clarifies that the "actual knowledge" standard determines when free text data must be redacted (for example, discharge summaries, progress notes, lab test interpretations, and clinical narratives in medical records). The guidance cautions that identifiers are not always clearly labeled. It references external sources for best practices in documentation and standards for creating de-identified data sets.

B. Expert Determination Method

De-identification of PHI applying the Expert Determination method requires a determination by an appropriate expert that the risk of re-identification is "very small" when the anticipated recipients of the data set use it alone or in combination with other reasonably available information. The expert must document the methods and results of such analysis. A covered entity may assign codes to data to enable re-identification if certain conditions are satisfied.

The new guidance addresses the following:

- *Who is an expert*: The new guidance clarifies that no specific professional degree or certification program defines who is an expert in de-identification. The guidance references expertise outside the health field.
- *Expiration*: The guidance approves the practice of time-limited expert determinations that a data set is de-identified. The practice allows an expert to take into account the fact that technology, social conditions, and the availability of information changes over time. Expiration of the time period determined by the expert does not imply that disseminated data is no longer protected, according to the guidance. Instead, an expert should determine whether future releases of the data to the same recipient need additional or different de-identification processes. (The guidance does not address whether a time-limited data set should be subject to a data use agreement providing that the data set will be destroyed at expiration of the time limit absent a new expert determination.)
- *Multiple solutions*: The new guidance explains that experts may tailor de-identification solutions to reflect the information expected to be available to the data set recipient. In crafting measures to protect against re-identification, experts should ensure that data sets cannot be combined to

evade those measures. For example, a covered entity might approve sharing of two data sets with a recipient after an expert determines that the data sets cannot be merged to identify an individual (or the expert could require a data use agreement with additional safeguards).

- *Explanation of expert methods*: Sections 2.6-2.8 of the guidance discuss processes and techniques employed by de-identification experts. This discussion should be helpful to covered entities who need to work with experts.
- *Re-identification codes*: Clarifying the HIPAA Privacy Rule provision on using codes for re-identification, the new explains that PHI may be converted to values derived by cryptographic hash functions, if the associated keys are not disclosed. A glossary defines a hash function as a mathematical function converting binary data into condensed representations.