HealthBlawg :: David Harlow's Health Care Law Blog

OCR releases HIPAA privacy rule guidance on de-identifying PHI

Posted: 27 Nov 2012 07:57 AM PST

Just two and a half years after hosting a workshop on the HIPAA Privacy Rule's de-identification standard, OCR has issued its "**Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule**." Like they say, it's not rocket surgery -- and there are few surprises here. One area worth reviewing is the expert determination section -- for those of you using, or considering the use of, expert opinions to guide your de-identification programs. Reproduced below is a table describing some of the principles used by experts in determining whether information has been de-identified:

Table 1. Principles used by experts in the determination of the identifiability of health information.

Principle	Description	Examples
Replicability	Prioritize health information features into levels of risk according to the chance it will consistently occur in relation to the individual.	*Low:* Results of a patient's blood glucose level test will vary
		High: Demographics of a patient (e.g., birth date) are relatively stable
Data source Availability	Determine which external data sources contain the patients' identifiers and the replicable features in the health information, as well as who is permitted access to the data source.	*Low:* The results of laboratory reports are not often disclosed with identity beyond healthcare environments.
		High: Patient name and demographics are often in public data sources, such as vital records -- birth, death, and marriage registries.
Distinguishability	Determine the extent to which the subject's data can be distinguished in the health information.	*Low:* It has been estimated that the combination of *Year of Birth, Gender,* and *3-Digit ZIP Code* is unique for approximately 0.04% of residents in the United States. This means that very few residents could be identified through this combination of data alone.

		High: It has been estimated that the combination of a patient's *Date of Birth, Gender,* and *5-Digit ZIP Code* is unique for over 50% of residents in the United States. This means that over half of U.S. residents could be uniquely described just with these three data elements.
Assess Risk	The greater the replicability, availability, and distinguishability of the health information, the greater the risk for identification.	*Low:* Laboratory values may be very distinguishing, but they are rarely independently replicable and are rarely disclosed in multiple data sources to which many people have access.
		High: Demographics are highly distinguishing, highly replicable, and are available in public data sources.

One element of the expert determination worth noting is the notion that a determination should perhaps be time-limited. Since that which is de-identified today may not be de-identified tomorrow (thanks in part to the rapid growth in the volume of data that is made available to the public on the internet). Here is the relevant FAQ:

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 such a data set, is de-identified information. However, experts have recognized that technology, social conditions, and the availability of information changes over time. Consequently, certain de-identification practitioners use the approach of time-limited certifications. In this sense, the expert will assess the expected change of computational capability, as well as access to various data sources, and then determine an appropriate timeframe within which the health information will be considered reasonably protected from identification of an individual.

Information that had previously been de-identified may still be adequately de-identified when the certification limit has been reached. When the certification timeframe reaches its conclusion, it does not imply that the data which has already been disseminated is no longer sufficiently protected in accordance with the de-identification standard. Covered entities will need to have an expert examine whether future releases of the data to the same recipient (e.g., monthly reporting) should be subject to additional or different de-identification processes consistent with current conditions to reach the very low risk requirement.

It is also worth noting that the guidelines suggest that a data use agreement is not required to be put in place in connection with the sharing of data de-identified in accordance with an expert determination.

However, use of such agreements is common, whether or not data has been de-identified, and may contain other provisions of value to the parties.

(I was also tickled to learn the identity of the seventeen ZIP code tabulation areas -- identified by the first three digits of their ZIP codes-- that include fewer than 20,000 residents each per the 2000 Census, and therefore must be listed as 000 in order for a record containing one of them to be condidered de-identified.)

When it comes to **HIPAA compliance**, these guidelines provide a greater measure of certainty regarding the **privacy** rule for folks in the secondary use of health data market. It remains to be seen whether the market has anticipated the content of these guidelines or whether there will be an uptick in the secondary use market, and further growth of "big data" in health care and/or an increase in the proliferation of health management tools (including mHealth apps using this population health data), as a result of the guidelines' release.

David Harlow
The Harlow Group LLC
Health Care Law and Consulting