Thomas D. Bixby
Law Office LLC
Legal Advice for Health Plans

PRIVACY LAW ALERT
November 26, 2012

OCR Issues Guidance on PHI De-Identification

Agency Explains Appropriate De-Identification Methods and Approaches

Today, the Department of Health and Human Services' Office for Civil Rights (OCR), which enforces the HIPAA Privacy and Security Rules, published guidance on how covered entities may de-identify protected health information so that the information is no longer subject to the Privacy Rule. Under the Privacy Rule, a covered entity may de-identify protected health information under a "safe harbor" by removing specified identifiers.[1] Alternatively, a covered entity may have a "person with appropriate knowledge of and experience with [appropriate] statistical and scientific principles and methods" determine whether protected health information is de-identified. The more significant aspects of the OCR guidance address the latter method of de-identification—de-identification by expert determination.

The guidance was published pursuant to the HITECH Act, which requires OCR to "issue guidance on how best to implement the requirements for the de-identification of protected health information under [the Privacy Rule]." Click here for the guidance, posted on the OCR's website.

De-Identification Using "Expert Determination"

Much of the guidance provides little new information. For example, OCR explains that "[t]here is no specific professional degree or certification program for designating who is an expert at rendering health information de-identified." Similarly, OCR admits that "[t]here is no explicit numerical level of identification risk that is deemed to universally meet the [Privacy Rule's standard]." But, the guidance does address some significant issues.

First, the OCR explains that experts who adopt a de-identification standard that a covered entity will use repeatedly over a period of time, such as with monthly or quarterly reports, should consider approving the standard with a "time-limited certification." Such certification would require that the process be reviewed after a specified period of time. OCR believes this is appropriate because "technology, social conditions, and the availability

[1] Protected health information is not de-identified if the covered entity has "actual knowledge that the information could be used alone or in combination with other information to identify an individual," notwithstanding that the specified identifiers have been removed.

of information changes over time." Thus, a de-identification process that is effective today may not be effective one-year (or five years) from now.

Although "OCR does not require a particular process for an expert to use" in determining that protected health information is de-identified, the agency explains that persons seeking to use de-identified information and the expert may engage in a process that requires "several iterations" before the participants "agree on an acceptable solution." Thus, the expert and data users may go back and forth several times trying to determine an appropriate combination of data elements that both meet the data users' needs and qualify as being de-identified under the Privacy Rule's standards.

The OCR lists three "principles for considering the identification risk of health information[, which] should serve as a starting point" for the de-identification process.

- **Replicability**: The likelihood a data element "will consistently occur in relation to [an] individual." Whereas an individual's birth date is highly replicable—it will always be the same—the individual's blood pressure varies, and so is less replicable.

- **Data source availability**: The availability of external sources of information that contain data elements with "replicable features" that are also in the health information the covered entity is trying to de-identify. While lab results are unlikely to be available in an external data source, an individual's date of birth or zip code may be readily available in public records, such as a voter registration list.

- **Distinguishability**: The extent to which an individual's data can be distinguished in the health information the covered entity is trying to de-identify. The OCR explains that over 50% of U.S. residents can be uniquely identified using date of birth, 5-digit zip code, and gender. Using only the first three digits of a zip code, however, reduces the likelihood of being uniquely identifiable to 0.4%. Thus, health information that contains 5-digit zip codes is much more "distinguishable" than health information that contains only the first three digits of zip codes.

An expert may assess the risk of health information remaining identifiable by evaluating these factors together. Data elements with low replicability, low availability, and low distinguishability create less risk of identification than data elements with high replicability, high availability, and high distinguishability.

* * * * * * * * * *

For more information, please contact Tom Bixby at (608) 661-4310 or
TBixby@tbixbylaw.com

Thomas D. Bixby Law Office LLC
(608) 661-4310 | www.tbixbylaw.com