The HIPAA Privacy Rule is intended to protect individually identifiable health information by limiting its use and disclosure. But the Privacy Rule expressly permits the de-identification of that information, and in doing so recognizes the usefulness of that information for “secondary purposes” such as comparative effectiveness studies, policy assessment and life sciences research. The Privacy Rule prescribes two methods by which individually identifiable information can be de-identified; one method requires the involvement of an expert, and the other does not.

Almost three years following the workshop convened by the U.S. Department of Health and Human Services Office of Civil Rights (OCR) to develop guidelines for implementing the two methods, OCR has published its Guidance Regarding Methods for De-Identification of Protected Health Information (the Guidance), which clarifies the two methods.

**Expert Determination**

Under the Privacy Rule, a covered entity may conclude that health information is not individually identifiable if “a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods ... (i) applying such principles and methods, determines that the risk is very small that the information could be used ... 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.” 45 CFR §164.514(b)(1).

The Guidance clarifies that the “expert” need possess no particular professional degree or certification in order to de-identify health information. It also notes that while the Privacy Rule does not require that determinations of de-identification be time-limited, changes in technology, social conditions and the availability of information over time have led de-identification practitioners in fact to impose time limits on their determinations. The result is that a covered entity may need to consult its expert upon expiration of the determination period, to assess whether the data requires additional or different de-identification efforts to remain sufficiently de-identified.

Interestingly, the Guidance does not attempt to quantify the level of risk that qualifies as “very small,” noting that the risk of identification for one data set in one context may not be the same as the risk of identification for the same data set in another context. Similarly, the Guidance acknowledges that there may be multiple “solutions” for de-identifying a particular data set. Some commenters think these guidelines should be more specific.

The great bulk of the Guidance regarding Expert Determination is devoted to a detailed and instructive discussion of the principles used by experts in determining the identifiability of health information (see the Guidance’s useful Table 1); the degree to which a data set containing health information about individuals can be linked to a data set that discloses the identity of the same individuals (see the Guidance’s Figure 3); and the approaches used by experts to de-identify individual health information, such as suppression, generalization and perturbation of patient values.

**Safe Harbor**

Under the Privacy Rule, a covered entity may also conclude that health information is not individually identifiable if (i)
specific identifiers (name, telephone number, social security number, etc.) of the individual and certain others are redacted and (ii) “the covered entity does not have actual knowledge that the information could be used ... to identify an individual who is a subject of the information.” 45 CFR §164.514(b)(2).

The Guidance clarifies when zip codes can be used in de-identified information and explains when dates are and are not permitted to be included. It also gives examples of the eighteenth identifier: “any other unique identifying number, characteristic or code.”

The Guidance addresses “actual knowledge,” offering examples of knowledge on the part of a covered entity that would constitute actual knowledge (where, for example, a patient is the former president of a state university.) Importantly, it points out that medical records exist in both structured and unstructured or “free text” form and that protected health information may well appear in free-text documents or fields. The Guidance provides that whether or not information in free-text documents or fields must be redacted in order to de-identify a particular record will depend on whether or not the information in question gives the covered entity actual knowledge of the identity of the individual.

Finally, the Guidance clarifies that, whether the Expert Determination or the Safe Harbor method is used to de-identify its health information, the covered entity need not require that the recipient of de-identified health information enter into a data use agreement, thus distinguishing the sharing of de-identified information from the sharing of a “limited data set” under the Privacy Rule.

**Bottom Line**

If compliant with the Guidance, de-identification of patient health data allows organizations that do not need to link the medical information to one unique patient to do valuable analyses and research outside the regulatory purview of the HIPAA Privacy Rule. While the OCR Guidance allows such de-identification of medical data, both de-identification methods approved by OCR, “even when properly applied, yield de-identified data that retains some risk of identification ... [and that] although the risk is very small, it is not zero.” For this reason, organizations embarking on the de-identification of patient health information will need to pay special attention to OCR’s Guidance.