Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act

On November 26, 2012, the Department of Health and Human Services ("HHS") Office for Civil Rights ("OCR") issued guidance regarding the methods and approaches for achieving de-identification in accordance with the Health Insurance Portability and Accountability Act ("HIPAA") Privacy Rule. This guidance is intended to assist covered entities with understanding the meaning of de-identification, the general process by which de-identified information is created, and the options available for performing de-identification in compliance with the de-identification standard set forth in Section 164.514(a) of the HIPAA Privacy Rule. Under this standard, health information is not individually identifiable if it fails to identify an individual and the covered entity has no reasonable basis to believe it can be used to identify the individual.

This standard may be implemented using one of two methods: (i) a determination by a qualified expert that health information is de-identified upon using certain scientific principles and methods and then documenting the justification for such determination ("Expert Determination"); or (ii) by removal of 18 specified identifiers, in combination with the covered entity's lack of actual knowledge that the de-identified information, alone or in combination with additional information, could be used to re-identify any individuals ("Safe Harbor"). The guidance explains the approaches that covered entities may use to de-identify protected health information ("PHI") under these two methods.

Expert Determination Method

With regard to de-identification via Expert Determination, the guidance provided detail regarding several questions, including (i) the length of time an Expert Determination is valid for a given data set; (ii) how an expert assesses the risk of identification of information; (iii) the approaches by which an expert assesses the risk that health information can be re-identified; (iv) the approaches by which an expert mitigates the risk of identification of an individual in health information; and (v) use of a data agreement when sharing de-identified data.

Time Limits on Expert Determinations. The guidance does not mandate an expiration date for Expert Determinations, but suggests that there may be time limits on the length of time an Expert Determination is valid for a given data set given changes in technology and information availability over time. Accordingly, covered entities may wish to implement time-limited certifications by the expert.

Assessment of the Risk of Identification of Information. According to the guidance, experts can use the following principles as a starting point when assessing the risk of re-identification of an individual's data in a particular set of health information: (i) the extent to which any features of the individual's health information are independently replicable (*e.g.,* birth date); (ii) whether external data sources are available that contain the individual's identifiers and replicable features (*e.g.,* public data sources such as birth or marriage registries); and (iii) the extent to which the individual's data is distinguishable in the health information (*e.g.,* combination of date of birth, gender and five-digit zip code). A greater risk of identification of an individual's data exists when there is greater replicability, availability, and distinguishability of the health information (*e.g.*, patient demographics are considered high-risk).

Approaches by which an Expert Assesses the Risk of Re-identification. An 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 guidance indicates that an expert may use external sources such as population data in order to assess this risk, but may rely upon statistics derived

from the data set if population statistics are unavailable or unknown. In the latter case, the guidance suggests that the expert would need to make a conservative estimate regarding the uniqueness of the record.

Approaches to Mitigate Risk of Re-identification. The guidance suggests that if an expert determines the risk of re-identification of an individual's health information is too high, the expert may modify the health information in order to mitigate the re-identification risk to a very small level. Some broad methods include: (i) suppression of an entire feature of the health information or of specific values within a record (*e.g.,* removal of the zip code feature or only those zip codes that may be unique); (ii) generalization of the health information into more abstract representations (*e.g.*, changing a five digit zip code to a three digit zip code); (iii) perturbation of specific values within the health information (*e.g.,* replacing specific values with different values that are equally specific); and (iv) limitation of the distribution of records through a data use agreement or restricted access agreement.

Use of a Data Agreement. The guidance indicates that a covered entity is not required to use a data use agreement when sharing de-identified data under either the Expert Determination or Safe Harbor method. However, a covered entity may require the recipient of de-identified information to enter into a data use agreement with provisions prohibiting re-identification to access files with a known disclosure risk.

Safe Harbor Method

With regard to de-identification via the Safe Harbor method, the guidance provided detail regarding several common questions, including (i) what constitutes "any other unique identifying number, characteristic, or code"; (ii) what is "actual knowledge" that the remaining information could be used to identify the individual; and (iii) whether a covered entity must remove PHI from free text fields in order to satisfy the Safe Harbor method.

Unique Identifying Number, Characteristic or Code. This category corresponds to any unique features not explicitly described in the Safe Harbor that could be used to identify a particular individual. Although the guidance notes that there are many potential identifying numbers, codes or characteristics, it specifically names a few, including clinical record numbers, barcodes in patient records and medications, and potentially the patient's occupation (*e.g.*, "current President of State University.").

Actual Knowledge Standard. The guidance defines "actual knowledge" as "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." It provides several examples of actual knowledge including a covered entity's awareness that (i) the occupation of a patient is listed in the record and would lead to the identification of the patient when combined with almost any additional data or (ii) the anticipated recipient of the information had a family member in the data and the data would provide sufficient context for the recipient to recognize the relative.

Removal of PHI from Free Text Fields. The guidance acknowledges that PHI exists in a multitude of forms, from highly structured database tables (*e.g.*, billing records) to documents written in natural language (*e.g.*, discharge summaries, progress notes). The guidance cautions covered entities that although the format of PHI may vary, the de-identification standard does not distinguish between data entered into standardized fields and information listed as free text. In either case, covered entities must take care to remove all identifiers outlined by the Safe Harbor, regardless of their form or format.

Overall, the guidance offers detailed practical advice for covered entities that de-identify PHI under either the Expert Determination or the Safe Harbor method. In the future, covered entities that de-identify data should take care both to follow the guidance and to appraise the risks and benefits of using each method.