Health Care ADVISORY

December 7, 2012

New Guidance on De-Identification of Protected Health Information Released by Office of Civil Rights

On November 26, 2012, the United States Department of Health and Human Services' (HHS) Office of Civil Rights (OCR) published a guidance document discussing methods and approaches for de-identification of protected health information (PHI) as permitted under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule.[1] The "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule" was published in response to a provision in the Health Information Technology for Economic and Clinical Health Act (HITECH) that required the Secretary of HHS to provide guidance on the best ways to satisfy the HIPAA Privacy Rule de-identification requirements within 12 months of enactment.[2] To meet this requirement, OCR held a workshop on March 8-9, 2010, to discuss various issues and concerns surrounding the de-identification of PHI. The De-identification Guidance summarizes a number of topics and issues discussed at the workshop. HHS published the Guidance as a tool to assist covered entities in understanding the process of de-identification and the appropriate uses of de-identified information.

The HIPAA Privacy Rule established two acceptable methods for the de-identification of PHI: (1) formal determination by a qualified expert ("Expert Determination") or (2) the removal of specified individual identifiers as well as absence of actual knowledge by the covered entity that the remaining information could be used alone or in combination with other information to identify the individual ("Safe Harbor"). The De-identification Guidance does not establish new de-identification methods;[3] rather, the document provides detailed explanations and answers on how covered entities may better satisfy the two established methods. Additionally, the Guidance provides direction on how covered entities may use each of the methods when engaging in the de-identification of PHI maintained in paper or electronic records. This advisory summarizes a number of those questions and explanations regarding the Expert Determination and Safe Harbor de-identification methods.

---

1 Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule ("De-Identification Guidance"), dated September 4, 2012, can be located at http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html.

2 42 U.S.C. § 17953 (Section 13424(c) of the HITECH Act).

3 The establishment of new methods to de-identify PHI would require rulemaking.

What Is De-Identified Protected Health Information?

The HIPAA Privacy Rule permits covered entities and their business associates to engage in the de-identification of PHI; once de-identified, covered entities and/or business associates can freely use and disclose such information that does not identify the individual who is the subject of the information.[4] Protected health information is health information, including demographic information, created or received by a covered entity that "relates to (1) the past, present, or future physical or mental health or condition of an individual; (2) the provision of health care to an individual; or (3) the past, present, or future payment for the provision of health care to an individual; and that identifies the individual or with respect to which there is a reasonable basis to believe the information can be used to identify the individual."[5] Demographic information associated with PHI may include certain identifying information, such as name, address, Social Security number or medical record number. PHI is considered de-identified, and therefore no longer PHI, when identifying information is removed so that the individual is no longer identifiable.

The HIPAA Privacy Rule establishes a standard for the de-identification of PHI. The standard states "[h]ealth information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information (PHI)."[6] Sections 164.514(b) and (c) provide the implementation specifications or requirements for de-identification of PHI. The implementation specifications set forth the two acceptable methods of de-identification of PHI: (1) Expert Determination and (2) Safe Harbor.

Expert Determination

The first acceptable method of de-identification is Expert Determination. The Privacy Rule provides that:

> [a] covered entity may determine that health information is not individually identifiable health information only if: (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable; (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient 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 C.F.R. § 164.514(b). OCR's De-identification Guidance addresses a number of questions regarding compliance with the Expert Determination method.

On the issue of who is an expert, OCR advises that there is no specific or required professional degree or certification required for a person to be an expert at determining whether PHI is de-identified. Experts may come from statistical, mathematical or other scientific fields, and expertise may be gained through various means of education and experience. When considering whether a person constitutes an expert as part of an enforcement action, OCR will consider the relevant professional experience and academic training of the expert, as well as the actual experience of the expert, using PHI de-identification methodologies.

---

4 45 C.F.R. § 164.502(d).

5 45 C.F.R. § 160.103.

6 45 C.F.R. § 164.514(a).

The implementation standard requires a "very small" risk that the de-identified information could be used only or together with other available information to identify the individual. The De-identification Guidance provides that "[t]here is no explicit level of identification risk that is deemed to universally meet the 'very small' level indicated by the method."[7] This results from the fact that the risk of identification for one data set in a specific environmental context may not apply to the same data set in a different environment or a different data set in the same environment. The Privacy Rule requires covered entities to document the methods and results of the analysis justifying the de-identification determination—such documentation must be made available to OCR upon request.

The De-identification Guidance addresses the question of how long an expert determination is valid for a given data set. The HIPAA Privacy Rule does not require an expiration date for an expert determination of de-identification of a data set. Given the nature of technology and other conditions changing over time, the De-identification Guidance, however, notes many experts use the approach of time-limited certifications. OCR advises covered entities using time-limited certifications to have their expert examine the data set at the end of the certification to determine if information remains de-identified, or whether future releases of the data will require an additional or different de-identification process.

Stakeholders attending the workshop suggested that the determination of identification risks can be achieved through a process consisting of a series of steps.

- **Step 1:** The expert will evaluate the extent to which the health information can (or cannot) be identified by the anticipated recipients.
- **Step 2:** The expert often will provide guidance to the covered entity or business associate on which statistical or scientific methods can be applied to the health information to mitigate the anticipated risk. The expert will execute such methods as deemed acceptable by the covered entity or business associate's data managers.
- **Step 3:** The expert will evaluate the identifiability of the resulting health information to confirm that the risk is no more than very small when disclosed to the anticipated recipients.[8]

The De-identification Guidance discusses principles used by experts to determine the identifiability of health information.[9] It also discusses the approaches by which an expert can assess the risk that health information can be identified. The HIPAA Privacy Rule de-identification standard does not require a particular method for experts to assess such risks. "A qualified 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 compared."[10]

OCR also recommends a number of approaches experts may use to mitigate the risk of identification of an individual in health information. If an expert determines that the risk of identification is greater than very small, the expert may modify the information to mitigate the identification risk to the very small risk level,

---

7 OCR, De-identification Guidance at 10.

8 *Id.* at 12.

9 *Id.* at 13-15.

10 *Id.* at 16.

as required by the de-identification standard. The recommended mitigation approaches discussed in the guidance document include:

- **Suppression techniques:** An approach involving the removal or elimination of certain features about the data prior to dissemination.
- **Generalization:** An approach involving the transformation of specific information/data into more abstract representations.
- **Perturbation:** An approach involving the replacement of specific data values with equally specific, but different, values.

OCR notes that there are many different disclosure risk reduction approaches for health information and that no particular method is universally considered the best option for every covered entity and health information set.

The De-identification Guidance provides clarification on the issue of what constitutes a code and how a code relates to PHI. OCR uses an approach similar to that taken by the National Institutes of Standards and Technology (NIST). A covered entity is permitted to disclose codes derived from PHI as part of a de-identified data set, if an expert determines that the data meets the HIPAA Privacy Rule de-identification standard at § 164.514(b)(1). Moreover, OCR notes that the Privacy Rule's re-identification provision[11] does not preclude an expert from using cryptographic hash functions to de-identify PHI, provided the keys associated with it are not disclosed.

Covered entities are not required to use a data use agreement when sharing de-identified data in order to satisfy the Expert Determination method because the Privacy Rule does not limit how covered entities may disclosed de-identified information. Covered entities may, however, use a data use agreement for recipients of de-identified information that access files with known disclosure risk, similar to that required for release of a limited data set under the HIPAA Privacy Rule.

Safe Harbor

The second acceptable method of de-identification is the Safe Harbor method. The Safe Harbor method is defined as the removal of certain identifiers of the individual or of relatives, employers, or household members of the individual, where the covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify the individual who is the subject of the information.[12] OCR's De-identification Guidance addresses a number of questions regarding compliance with the Safe Harbor method.

---

[11] 45 C.F.R. § 164.514(c) (a code or other means of record; re-identification can be assigned to permit re-identification, provided that the code or means of re-identification (1) is not derived from or related to information about the individual and is not capable of being translated to identify the individual, and (2) is not used or disclosed for any other purpose, and the mechanism for re-identification is not disclosed).

[12] 45 C.F.R. § 164.514(b)(2). Identifiers include: names, geographic subdivisions smaller than a state, all elements of dates (except year) for dates that are directly related to an individual, telephone numbers, fax numbers, email addresses, Social Security numbers, medical record numbers, health plan beneficiary numbers, account numbers, certificate/license numbers, vehicle identifiers and serial numbers, device identifiers and serial numbers, URLs, IP addresses, biometric identifiers, full-face photographs and any comparable images, and any other unique identifying number, characteristic or code.

OCR addresses the question of when ZIP codes can be included in de-identified information. OCR advises that "covered entities may include the first three digits of the ZIP code if, according to current publicly available data from the Bureau of the Census: (1) the geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; or (2) the initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000."[13] Accordingly, covered entities may include the initial three digits of ZIP codes in de-identified information except in the following restricted ZIP codes that must be listed as 000: 036, 059, 063, 102, 203, 556, 692, 790, 821, 823, 830, 831, 878, 879, 884, 890, and 893.[14] OCR notes that covered entities should refer to the Bureau of the Census on a regular basis for updates regarding ZIP code populations.

The guidance states that the disclosure of parts or derivatives of the listed identifiers (such as a data set containing patient initials or the last four digits of a Social Security number) is not consistent with the Safe Harbor method. Additionally, dates that include the day, month and any other information more specific than the year of an event are not permitted under the Safe Harbor method. Further, dates associated with test measures, such as those from a laboratory report, are protected health information and do not meet the requirements of the Safe Harbor method.

OCR also provides guidance on what constitutes of "any other unique identifying number, characteristic, or code"[15] as this phrase is included in the Safe Harbor list of identifiers that must be deleted. The following are examples of unique features that fall into that category, and must also be stripped from a data set:

- **Identifying Number:** There are a number of potential identifying numbers. For example, the preamble to the Privacy Rule discusses clinical trial record numbers as being such an identifying number.
- **Identifying Code:** A code relating to a value derived from a non-secure encoding mechanism. Potential identifying codes include barcodes embedded in electronic medical records and electronic prescribing systems.
- **Identifying Characteristic:** Characteristics are anything that distinguishes an individual and allows identification. For example, a patient's occupation may be an identifying characteristic if the patient has a unique occupation.

The De-identification Guidance discusses what constitutes "actual knowledge" so 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 as required under the Safe Harbor method. In this context, "actual knowledge" means clear and direct knowledge. A covered entity has actual knowledge if it concludes that the remaining information could be used to identify the individual, or the covered entity is aware that the information is not actually de-identified information—OCR gives several examples of instances where an entity would have "actual knowledge." Importantly, however, OCR notes that, without more, knowledge of specific studies concerning methods to re-identify health information or to use de-identified information (alone or in combination with other information) to identify an individual does not constitute "actual knowledge" that such methods would be used in connection with the de-identified information.

---

[13] OCR, De-identification Guidance at 23.

[14] This information regarding restricted ZIP codes is based on population data from the 2000 Census.

[15] *See* 45 C.F.R. § 164.514(b)(2)(i)(R).

OCR advises that a covered entity is not required to suppress all personal names—such as a physician's name—from health information for it to be considered de-identified. The Safe Harbor method only requires that names of the individuals associated with the corresponding health information and their relatives, employers and household members be suppressed. While there is no requirement to remove the names of providers or other workforce members, there is also no requirement to retain the information in the de-identified data set.

As discussed under the Expert Determination method, covered entities are not required to use a data use agreement when sharing de-identified information in order to satisfy the Safe Harbor method, because the Privacy Rule does not limit how covered entities may disclose de-identified information. Covered entities may, however, use a data use agreement in connection with the disclosure of information de-identified under the Safe Harbor method and, for example, prohibit the recipient from re-identifying the information.

According to OCR, the de-identification standard does not distinguish between data entered into a standardized field or structured databases and information maintained as unstructured free text. Under the Safe Harbor method, listed identifiers must be removed regardless of location in a standardized or free text field if it is recognizable as an identifier. "Whether additional information must be removed falls under the actual knowledge provision; the extent to which the covered entity has *actual knowledge* that residual information could be used to individually identify a patient."[16]

Conclusion

The OCR guidance document provides a great deal of information on satisfying the two acceptable de-identification methods, Expert Determination and Safe Harbor. Covered entities and business associates should review this document as it addresses a number of questions from industry stakeholders regarding compliance with the two methods.

For more information regarding the HIPAA Privacy Rule, the de-identification standards and acceptable methods, or any of the information discussed in the OCR guidance document, please do not hesitate to contact a member of the Alston & Bird Health Care Regulatory Group.

---

[16] OCR, De-identification Guidance at 29.

ALSTON+BIRD LLP

If you would like to receive future *Health Care Advisories* electronically, please forward your contact information including e-mail address to **healthcare.advisory@alston.com**. Be sure to put "**subscribe**" in the subject line.

For further information, please do not hesitate to contact any of the following:

Donna P. Bergeson
404.881.7278
donna.bergeson@alston.com

Cathy L. Burgess
202.239.3648
cathy.burgess@alston.com

Angela T. Burnette
404.881.7665
angie.burnette@alston.com

Jennifer L. Butler
202.239.3326
jennifer.butler@alston.com

Brendan Carroll
202.239.3216
brendan.carroll@alston.com

Guillermo Cuevas
202.239.3205
guillermo.cuevas@alston.com

Peter Fise
202.239.3842
peter.fise@alston.com

Joyce Gresko
202.239.3628
joyce.gresko@alston.com

Elinor A. Hiller
202.239.3401
elinor.hiller@alston.com

William H. Jordan
404.881.7850
bill.jordan@alston.com

Peter M. Kazon
202.239.3334
peter.kazon@alston.com

Blanche L. Lincoln
202.239.3601
blanche.lincoln@alston.com

Dawnmarie R. Matlock
404.881.4253
dawnmarie.matlock@alston.com

Kim McWhorter
404.881.4254
kim.mcwhorter@alston.com

Raad S. Missmar
202.239.3034
rudy.missmar@alston.com

William (Mitch) R. Mitchelson, Jr.
404.881.7661
mitch.mitchelson@alston.com

D'Andrea J. Morning
404.881.7538
dandrea.morning@alston.com

Elise N. Paeffgen
202.239.3939
elise.paeffgen@alston.com

Michael H. Park
202.239.3630
michael.park@alston.com

Earl Pomeroy
202.239.3835
earl.pomeroy@alston.com

Steven L. Pottle
404.881.7554
steve.pottle@alston.com

J. Mark Ray
404.881.7739
mark.ray@alston.com

Mark H. Rayder
202.239.3562
mark.rayder@alston.com

Colin Roskey
202.239.3436
colin.roskey@alston.com

Marc J. Scheineson
202.239.3465
marc.scheineson@alston.com

Thomas A. Scully
202.239.3459
thomas.scully@alston.com

Donald E. Segal
202.239.3449
donald.segal@alston.com

Robert G. Siggins
202.239.3836
bob.siggins@alston.com

Carolyn E. Smith
202.239.3566
carolyn.smith@alston.com

Paula M. Stannard
202.239.3626
paula.stannard@alston.com

Robert D. Stone
404.881.7270
rob.stone@alston.com

W.J. "Billy" Tauzin
202.684.9844
billy.tauzin@alston.com

Julie Klish Tibbets
202.239.3444
julie.tibbets@alston.com

Timothy P. Trysla
202.239.3420
tim.trysla@alston.com

Michelle A. Williams
404.881.7594
michelle.williams@alston.com

Marilyn K. Yager
202.239.3341
marilyn.yager@alston.com

Esther Yu
212.210.9568
esther.yu@alston.com

ATLANTA
One Atlantic Center
1201 West Peachtree Street
Atlanta, GA 30309-3424
404.881.7000

BRUSSELS
Level 20 Bastion Tower
Place du Champ de Mars
B-1050 Brussels, BE
Phone: +32 2 550 3700

CHARLOTTE
Bank of America Plaza
Suite 4000
101 South Tryon Street
Charlotte, NC 28280-4000
704.444.1000

DALLAS
2828 N. Harwood St.
Suite 1800
Dallas, TX 75201
214.922.3400

LOS ANGELES
333 South Hope Street
16th Floor
Los Angeles, CA 90071-3004
213.576.1000

NEW YORK
90 Park Avenue
New York, NY 10016-1387
212.210.9400

RESEARCH TRIANGLE
4721 Emperor Boulevard
Suite 400
Durham, NC 27703-8580
919.862.2200

SILICON VALLEY
275 Middlefield Road
Suite 150
Menlo Park, CA 94025-4004
650.838.2000

VENTURA COUNTY
Suite 215
2801 Townsgate Road
Westlake Village, CA 91361
805.497.9474

WASHINGTON, D.C.
The Atlantic Building
950 F Street, NW
Washington, DC 20004-1404
202.239.3300

www.alston.com