Big Data Analytics Under HIPAA

Kevin Coy and Neil W. Hoffman, Ph.D.

Privacy laws and regulations such as the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule can have a significant impact on the development of health care data analytics in cases where those analytics rely upon, or are developed using, HIPAA-regulated protected health information. Data analytics, also referred to as "big data" when done on a large scale, are often viewed as having the potential of transforming health care, enabling providers to use population data in identifying and preventing diseases, developing treatments, and reducing costs of health care.[1] As recently noted by the Privacy and Security Workgroup (PSWG),[2] benefits of data analytics could include "safer treatments, the ability to target communities and individuals with tailored interventions, and the ability to respond to the spread of diseases more rapidly."

The many potential benefits from data analytics for the health care system and to the health of individuals must be balanced with protecting the privacy of individuals whose health information is used in those analytics. In noting the potential benefits of data analytics, the PSWG also stated that "[r]apid growth in the volume of health-related information increases the risk of privacy violations, particularly when data sets are combined. Data anonymization tools such as de-identification are useful, but cannot eliminate risks to re-identification."[3] Such comments illustrate the inherent tension between the development of big data and privacy concerns under HIPAA.

Applicability of HIPAA

The HIPAA Privacy Rule only applies to health information from health care providers, health plans, and healthcare clearinghouses (HIPAA "covered entities") and their business associates. Data analytics dependent upon health information from these sources must take the requirements of the HIPAA Privacy Rule into account. Analytics based on information from sources not subject to HIPAA still may present important privacy issues, but such analytics are not regulated by HIPAA.

Large vendors capable of providing data analytics are often able to do so based on their HIPAA-defined, business associate relationships with health care providers and other covered entities.[4] For HIPAA purposes, a "business associate" is a person or entity that uses or discloses protected health information in providing services on behalf of covered entities. Accordingly, the context in which such vendors provide data analytics services must satisfy the HIPAA rules governing business associates. This article discusses those rules that are most applicable to data analytics under business associate arrangements, as well as HIPAA rules governing de-identification and "limited data sets." As discussed below, however, a general lack of clarity regarding these rules makes it uncertain how far these vendors can go in using protected health information for their own commercial, data-analytics purposes.

1 *A Policy Forum on the Use of Big Data in Health Care*. Bipartisan Policy Center (Dec. 3, 2013).
2 *Health Big Data Recommendations.* Privacy and Security Workgroup (PSWG) of the Health Information Technology Policy Committee (HITPC) (Aug. 2015), at p. 3. Available online at: http://www.healthit.gov/sites/faca/HITPC_Health_Big_Data_Report_FINAL.pdf.
3 *Id.* at 4.
4 *See* R. Hirsh & H. Deixler: *HIPAA Business Associates and Health-Care Big Data: Big Promise, Little Guidance. Bloomberg Law (Feb. 21, 2014).*

Business Associates

Under HIPAA, a "business associate" is a person or entity, other than a member of the covered entity's work force, that performs certain functions or activities that involve the use or disclosure of protected health information on behalf of, or provides services to, a covered entity.[5] The covered entity must obtain satisfactory assurances from its business associates that they will appropriately safeguard the protected health information they receive or create on behalf of the covered entity. These assurances must be in the form of a written contract that contains requirements specified under the Privacy Rule.[6]

Data Aggregation under HIPAA

Business associates may, under the Privacy Rule, perform "data aggregation services" if related to the covered entity's "health care operations" and if specifically permitted in the underlying business associate agreement. Data aggregation is a permissible term for a business associate agreement; a covered entity is not required to permit its business associate(s) to engage in data aggregation activities. In its commentary to the final Privacy Rule, the U.S. Department of Health and Human Services (HHS) stated that it was permitting data aggregation under business associate arrangements so that covered entities may contract with business associates "to undertake quality assurance and comparative analyses that involve the protected health information of more than one contracting covered entity."[7] However, the HHS definitions of "data aggregation" and "health care operations" are key to understanding the applicability of this permitted use to data analytics.

"Data aggregation," under the Privacy Rule, is the act of a business associate combining protected health information from multiple covered entities in order "to permit data analyses that relate to the health care operations of the respective covered entities."[8] "Health care operations," under the Privacy Rule, is broadly defined to include—

> [c]onducting quality assessment and improvement activities, including outcomes evaluation and development of clinical guidelines, provided that the obtaining of generalizable knowledge is not the primary purpose of any studies resulting from such activities; population-based activities relating to improving health or reducing health care costs, protocol development, case management and care coordination, contacting of health care providers and patients with information about treatment alternatives; and related functions that do not include treatment; . . . .[9]

Certain of these activities, such as "population-based activities" appear consistent with the concept of business associates providing data analytics services. Unfortunately, there remains a lack of guidance by HHS and the Office of Civil Rights (OCR), the branch of HHS that directly regulates HIPAA, as to the scope of activities considered to be "data aggregation,"[10] and whether business associates can undertake such activities for their own data-analytics purposes—in other words, the extent to which such activities can or must be linked to the covered entity's health care operations.

De-Identification of Protected Health Information

Under the Privacy Rule, health information that does not identify an individual and for which there is no reasonable basis for a covered entity to believe that it can be used to identify an individual is not protected health information under HIPAA.[11] Accordingly, health information that has been "de-identified" pursuant to applicable HIPAA standards is no longer subject to protection under HIPAA. Note that OCR has recognized two methods for de-identifying protected health

5 45 C.F.R. § 160.103.
6 45 C.F.R. § 164.504(e).
7 *Standards for Privacy of Individually Identifiable Health Information.* 65 Fed. Reg. 82462, 82644 (Dec. 28, 2000).
8 45 C.F.R. § 164.501.
9 *Id.*
10 *See* R. Hirsh & H. Deixler, supra note 4.
11 45 C.F.R. § 164.514.

information under the Privacy Rule.[12] These two methods sometimes are referred to the "safe harbor" method and the "statistical" or "expert" method of de-identification.

The "safe harbor" method of de-identification requires the removal of the following identifiers of the individual or the individual's relatives, employers, or household members:

- names;
- all geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code and equivalent geocodes (with certain limited exceptions);
- all elements of dates other than year for dates directly related to an individual, including birth date, admission date, and discharge date, date of death, and all ages over 89 and all elements of dates (including year indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older;
- telephone numbers;
- fax numbers;
- e-mail addresses;
- social security numbers;
- medical record numbers;
- health plan beneficiary numbers;
- account numbers;
- certificate/license numbers;
- vehicle identifiers and serial numbers, including license plate numbers;
- device identifiers and serial numbers;
- Web Universal Resource Locators (URLs);
- Internet protocol (IP) address numbers;
- biometric identifiers, including finger and voice prints;
- full face photographic images and any comparable images; and
- any other unique identifying number, characteristic, or code.[13]

Also, the covered entity must not have actual knowledge that the remaining information could be used, either by itself or in combination with other information, to identify the individual to whom such information pertains.[14]

The "statistical" or "expert" method involves an expert determination that the information is not individually identifiable. A covered entity may rely on such determination when a person with appropriate knowledge and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable—

- applies such principles and methods, determines that the risk is very small that the information could be used, alone or in conjunction with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
- documents the methods and results of the analysis that justify such determination.[15]

What does it mean to be a "person with appropriate knowledge and experience"? In its November 26, 2012 guidance, OCR addressed this question as follows:

> There is no specific professional degree or certification program for designating who is an expert at rendering health information de-identified. Relevant expertise may be gained through various routes of education and experience. Experts may be found in the statistical, mathematical, or other scientific domains. From an enforcement perspective,

12 *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.* OCR (Nov. 26, 2012). Available online at: http://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf.
13 45 C.F.R. § 164.514(b)(2)(i).
14 45 C.F.R. § 164.514(b)(2)(ii).
15 45 C.F.R. § 164.514(b)(1).

> OCR would review the relevant professional experience and academic or training of the expert used by the covered entity, as well as actual experience of the expert using health information de-identification methodologies.[16]

With respect to data analytics, however, de-identified health information may have limited use. As the PSWG observed, "data de-identified pursuant to HIPAA as the enabling mechanism for data use often significantly reduces the potential for valuable uses of information even where the risk associated with the use of more identifiable information is very low."[17]

Whatever its ultimate value, and depending on the purpose, if a business associate wants to be able to de-identify protected health information it would be prudent to expressly address this in the relevant business associate agreement. A business associate may, however, use de-identified health information for any purpose since such information is no longer protected health information. But what is less clear is the extent to which a business associate may receive protected health information from a covered entity in order to de-identify such information for the business associate's own commercial purposes, as in providing data analytics to others for financial gain. Such use, at least arguably, would not be for or on behalf of the covered entity in keeping with the definition of business associate under the Privacy Rule. Again, clear guidance from OCR on this point would be helpful.

Limited Data Sets

Covered entities can disclose "limited data sets" (protected health information from which many direct identifiers have been removed as discussed below) in accordance with a data use agreement between the covered entity and a third party. Unlike de-identified data, a limited data set is still considered to be protected health information and may only be used and disclosed for purposes of research, public health, or health care operations. Limited data sets must be governed by a data use agreement that meets HIPAA requirements, but these requirements are more limited than what a business associate agreement would require.

A limited data set is protected health information that excludes the following direct identifiers of the individual or of relatives, employers, or household members of the individual:

(i) Names;
(ii) Postal address information, other than town or city, State, and zip code;
(iii) Telephone numbers;
(iv) Fax numbers;
(v) Electronic mail addresses;
(vi) Social security numbers;
(vii) Medical record numbers;
(viii) Health plan beneficiary numbers;
(ix) Account numbers;
(x) Certificate/license numbers;
(xi) Vehicle identifiers and serial numbers, including license plate numbers;
(xii) Device identifiers and serial numbers;
(xiii) Web Universal Resource Locators (URLs);
(xiv) Internet Protocol (IP) address numbers;
(xv) Biometric identifiers, including finger and voice prints; and
(xvi) Full face photographic images and any comparable images.[18]

The list of identifiers that must be removed is similar to, but not as extensive as, the list of identifiers that must be removed to meet the HIPAA de-identification safe harbor, allowing the inclusion of information about dates and geographic subdivisions that would need to be removed for the data to be considered de-identified under HIPAA.

16 *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.* OCR (Nov. 26, 2012) at p. 10. Available online at: http://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf.
17 *Health Big Data Recommendations.* PSWG (draft as of Aug. 6, 2015), at p. 14.
18 45 C.F.R. § 164.514(e)(2).

If the covered entity wants to provide a data set that includes any of the above identifiers to a third party so that the third party can create a limited data set for the covered entity, a business associate agreement would be required for the work related to the creation of the limited data set.[19]

For a covered entity to provide a limited data set to a third party (or to allow a third party to use a limited data set that the third party created for the covered entity pursuant to a business associate agreement) a data use agreement is required. The data use agreement, which is different from a business associate agreement, must:

A. Establish the permitted uses and disclosures of such information by the limited data set recipient. The data use agreement may not authorize the limited data set recipient to use or further disclose the information in a manner that would be impermissible if done by the covered entity.
B. Establish who is permitted to use or receive the limited data set; and
C. Provide that the limited data set recipient will:

1. Not use or further disclose the information other than as permitted by the data use agreement or as otherwise required by law;
2. Use appropriate safeguards to prevent use or disclosure of the information other than as provided for by the data use agreement;
3. Report to the covered entity any use or disclosure of the information not provided for by its data use agreement of which it becomes aware;
4. Ensure that any agents to whom provides the limited data set agree to the same restrictions and conditions that apply to the limited data set recipient with respect to such information; and
5. Not identify the information or contact the individuals.[20]

Whether a limited data set would be useful (or preferable to de-identified data) for data analytics purposes would depend upon the analytic project and whether this additional geographic or date information added value to the analytics project.

Business Associates' Management and Administration

One final aspect of the Privacy Rule should be mentioned in the context of data analytics. If included in the business associate agreement, a business associate may use protected health information received from the covered entity for its management and administration purposes if—

- the disclosure is required by law; or
- the business associate obtains reasonable assurances from the person to whom the information is disclosed that it will be held confidentially and used or further disclosed only as required by law or the purpose for which it was disclosed to the person; and the person notifies the business associate of any instances of which it is aware in which the confidentiality of the information has been breached.[21]

Again, this is a permissive term for a business associate agreement; a covered entity is not required to include such language in its business associate agreements. While the clause provides additional flexibility for the business associate, the full scope of the provision is unclear. Unfortunately, OCR has provided little guidance as to the meaning of "management" and "administration" with respect to this provision, and what guidance it has provided may not be particularly favorable to the development of data analytics. In its comments to the proposed Privacy Rule, OCR stated the following: "[a]side from disclosures for data aggregation and business associate management, the business associate contract cannot authorize any uses or disclosures that the covered entity itself cannot make. Therefore, data mining by the business associate for any purpose not specified in the contract is a violation of the contract and grounds for termination of the contract by the covered entity."[22] Covered entities frequently limit use of their protected health

19 45 C.F.R. § 164.514(e)(3)(ii).
20 45 C.F.R. § 164.514(e)(4).
21 45 C.F.R. § 164.504(e)(4)(ii).
22 65 Fed. Reg. at 82644.

information to supporting the work being done for that covered entity or, for data aggregation purposes, as listed above. The extent to which a covered entity could choose to authorize additional data analytics under the HIPAA's "management and administration" provision is an open question.

Conclusion

Data analytics hold considerable promise for improving healthcare, but the privacy interests of the subjects of the data must also be addressed. Adding to this tension are uncertainties as to how HIPAA may restrict or permit use of protected health information in data analytics, where large vendors capable of providing such services gain access to such information through their business associate relationships with covered entities. Additional guidance from OCR specific to this area would be a welcome development. In the meantime, data analytics firms should carefully consider how they structure their activities from a HIPAA standpoint, and covered entities likewise should carefully consider the extent to which they choose to authorize their business associates to use protected health information for these purposes.

Authors and Contributors

Kevin Coy
Partner, DC Office
202.677.4034
kevin.coy@agg.com

Neil W. Hoffman, Ph.D.
Partner, Atlanta Office
404.873.8594
neil.hoffman@agg.com

not *if*, but *how*.®

About Arnall Golden Gregory LLP

Arnall Golden Gregory, a law firm with more than 150 attorneys in Atlanta and Washington, DC, employs a "business sensibility" approach, developing a deep understanding of each client's industry and situation in order to find a customized, cost-sensitive solution, and then continuing to help them stay one step ahead. Selected for The National Law Journal's prestigious 2013 Midsize Hot List, the firm offers corporate, litigation and regulatory services for numerous industries, including healthcare, life sciences, global logistics and transportation, real estate, food distribution, financial services, franchising, consumer products and services, information services, energy and manufacturing. AGG subscribes to the belief "not if, but how.®" Visit www.agg.com.

Atlanta Office
171 17th Street, NW
Suite 2100
Atlanta, GA 30363

Washington, DC Office
1775 Pennsylvania Avenue, NW
Suite 1000
Washington, DC 20006

To subscribe to future alerts, insights and newsletters: **http://www.agg.com/subscribe/**