Published in final edited form as:
JAMA. 2015 April 14; 313(14): 1471–1473. doi:10.1001/jama.2015.2252.

Data Breaches of Protected Health Information in the United States

Vincent Liu, MD, MS, **Mark A. Musen, MD, PhD**, and **Timothy Chou, PhD**
Kaiser Permanente Division of Research, Oakland, California (Liu); Stanford Center for Biomedical Informatics Research, Stanford, California (Musen); Department of Computer Science, Stanford University, Stanford, California (Chou)

Reports of data breaches have increased during the past decade.[1,2] Compared with other industries, these breaches are estimated to be the most costly in health care; however, few studies have detailed their characteristics and scope.[1]

Methods

We evaluated an online database maintained by the US Department of Health and Human Services describing data breaches of unencrypted protected health information (ie, individually identifiable information) reported by entities (health plans and clinicians) covered under the Health Insurance Portability and Accountability Act (HIPAA).[3] Under the Health Information Technology for Economic and Clinical Health Act of 2009, breaches involving the acquisition, access, use, or disclosure of protected health information and thus posing a significant risk to affected individuals must be reported.[4]

When data breaches affect 500 individuals or more, the report must include the name and state of the entity breached, the number of records affected, the type and source of the breach, and the involvement of any external vendor using protected health information. Examples include the theft of unsecured laptops, dissemination of data in emails, and improper disposal of patient records. Reports are made online via form templates.[3]

We included breaches affecting 500 individuals or more reported as occurring from 2010 through 2013, accounting for 82.1% of all reports.[3] We quantified the frequency and

Corresponding Author: Vincent Liu, MD, MS, Kaiser Permanente Division of Research, 2000 Broadway, Oakland, CA 94612 (vincent.x.liu@kp.org)..

Author Contributions: Dr Liu had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: All authors.
Acquisition, analysis, or interpretation of data: Liu.
Drafting of the manuscript: Liu.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Liu.
Obtained funding: Liu.
Administrative, technical, or material support: Musen.
Study supervision: Chou.

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Correction: This article was corrected on May 21, 2015, to fix wording and data errors in the text and Table.

geographic locations of breaches, adjusting for 2013 population estimates from the US Census Bureau.

Based on categorical templates, we grouped breaches as occurring via theft, loss or improper disposal of data, unauthorized data access or disclosure, hacking or information technology incidents, or other and missing (n = 2). We described the media through which breaches occurred as electronic (including network server; desktop computer, email, and electronic medical records; or laptop computer and electronic portable devices), paper, or other.

We compared annual data with χ^2 tests and linear regression using Stata version 13.1 (StataCorp) with a 2-sided significance level of $P < .05$. The Kaiser Permanente Northern California institutional review board determined that this study did not qualify as human subjects research.

Results

We evaluated 949 breaches affecting 29 million records between 2010 and 2013. Six breaches involved more than 1 million records each and the number of reported breaches increased over time, although the trend using linear regression did not reach statistical significance (*P* = .07; Table). Breaches were reported in every state, the District of Columbia, and Puerto Rico. Five states (California, Texas, Florida, New York, and Illinois) accounted for 34.1% (95% CI, 31.2%-37.2%) of all breaches. However, when adjusted by population estimates, the states with the highest adjusted number of breaches and affected records varied (Figure).

Most breaches occurred via electronic media (67.4%; 95% CI, 64.4%-70.4%; Table), frequently involving laptop computers or portable electronic devices (32.7%; 95% CI, 29.7%-35.7%). Most breaches also occurred via theft (58.2%; 95% CI, 55.0%-61.3%). The combined frequency of breaches resulting from hacking and unauthorized access or disclosure increased during the study period (12.1% in 2010 to 27.2% in 2013; *P* = .003). Breaches involved external vendors in 28.8% (95% CI, 25.9%-31.7%) of reports.

Discussion

Between 2010 and 2013, data breaches reported by HIPAA-covered entities involved 29 million records. Most data breaches resulted from overt criminal activity. The persistent threat of theft and the increase in hacking raise serious security concerns.

Our study was limited to breaches that were already recognized, reported, and affecting at least 500 individuals. Therefore, our study likely underestimated the true number of health care data breaches occurring each year. Some entities or patients may have been involved in more than 1 breach.

We were unable to assess the costs or the effect on operations caused by these breaches and the accompanying increased data security measures. We were also unable to calculate the rates at which breaches occurred based on the number of total US records or entities at risk.

Given the rapid expansion in electronic health record deployment since 2012, as well as the expected increase in cloud-based services provided by vendors supporting predictive analytics, personal health records, health-related sensors, and gene sequencing technology, the frequency and scope of electronic health care data breaches are likely to increase.[2,5,6] Strategies to mitigate the risk and effect of these data breaches will be essential to ensure the well-being of patients, clinicians, and health care systems.

Acknowledgments

Funding/Support: Dr Liu was supported by the Permanente Medical Group and grant K23 GM112018 from the National Institutes of Health.

Role of the Funder/Sponsor: The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript, and decision to submit the manuscript for publication.

References

1. Symantec Corporation. [Accessed December 8, 2014] Cost of data breach study: global analysis. 2013. https://www4.symantec.com/mktginfo/whitepaper/053013_GL_NA_WP_Ponemon-2013-Cost-of-a-Data-Breach-Report_daiNA_cta72382.pdf
2. Blumenthal D. Wiring the health system—origins and provisions of a new federal program. N Engl J Med. 2011; 365(24):2323–2329. [PubMed: 22168647]
3. US Department of Health and Human Services Office for Civil Rights. [Accessed December 4, 2014] Breaches affecting 500 or more individuals. https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf
4. 45 CFR Parts 160 and 164.
5. Schneeweiss S. Learning from big health care data. N Engl J Med. 2014; 370(23):2161–2163. [PubMed: 24897079]
6. Adler-Milstein J, Jha AK. Sharing clinical data electronically: a critical challenge for fixing the health care system. JAMA. 2012; 307(16):1695–1696. [PubMed: 22535851]

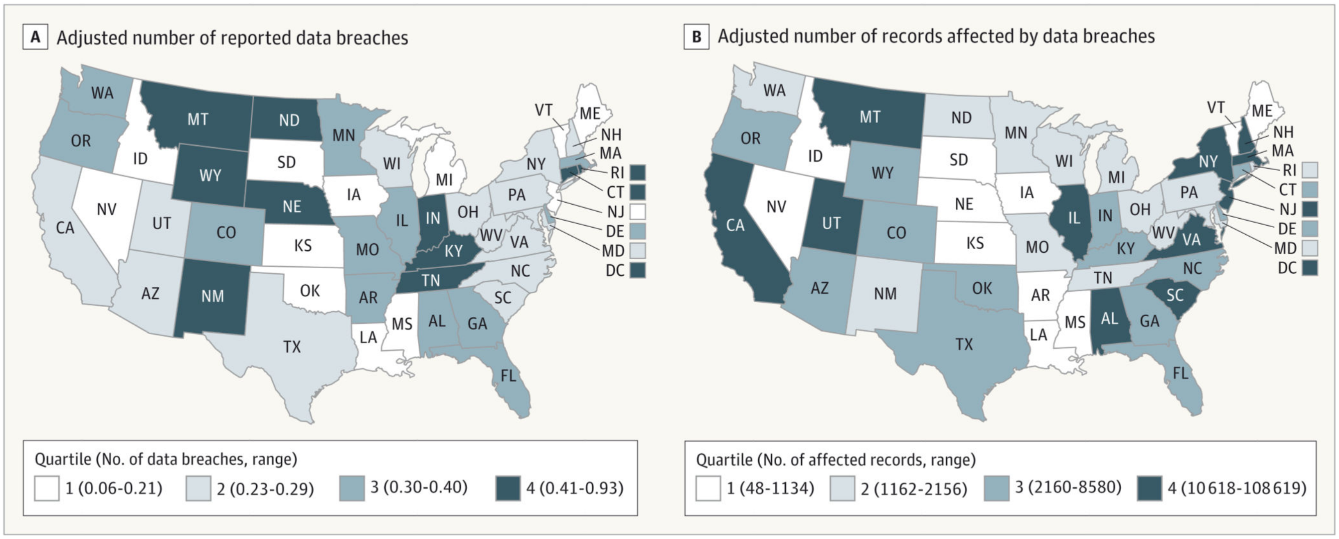

Figure. Adjusted Number of Data Breaches and Affected Records Between 2010 and 2013 by State and Quartile

Adjusted values were calculated by dividing the number of breaches and the affected records by 2013 population estimates from the US Census Bureau based on the state in which the breach was reported. The data quartiles are per 100 000 residents. The Figure does not display data for Hawaii, Alaska, or Puerto Rico.

Table

Characteristics of Data Breaches of Protected Health Information Affecting at Least 500 Individuals Reported by Entities Covered by the Health Insurance Portability and Accountability Act

		Year of Data Breach				***P*** **Value**[a]
	Overall	**2010**	**2011**	**2012**	**2013**	
Total No. of data breaches reported	949	214	236	234	265	.07
Total No. of records affected, in millions	29.0	5.1	11.6	3.4	9.0	.88
No. of data breaches affecting at least 1 million records	6	1	3	0	2	.37
Data breach by media type, No. (%) [95% CI]						
Portable electronic device or laptop	310 (32.7) [29.7-35.7]	77 (36.0) [29.8-42.7]	72 (30.5) [24.9-36.7]	78 (33.3) [27.5-40.0]	83 (31.3) [26.0-37.2]	
Desktop, email, or EMR	148 (15.6) [13.4-18.0]	32 (15.0) [10.7-20.4]	25 (10.6) [7.2-15.2]	43 (18.4) [13.9-23.9]	48 (18.1) [13.9-23.3]	.09
Paper	212 (22.3) [19.8-25.1]	50 (23.4) [18.1-30.0]	55 (23.3) [18.3-29.2]	52 (22.2) [17.3-28.0]	55 (20.8) [16.3-26.1]	
Network server	101 (10.6) [8.8-12.8]	16 (7.5) [4.6-11.9]	25 (10.6) [7.2-15.2]	29 (12.4) [8.7-17.3]	31 (11.7) [8.3-16.2]	
Other	178 (18.8) [16.4-21.4]	39 (18.2) [13.6-24.0]	59 (25.0) [19.9-31.0]	32 (13.7) [9.8-18.7]	48 (18.1) [13.9-23.3]	
Data breach category, No. (%) [95% CI]						
Theft	552 (58.2) [55.0-61.3]	139 (65.0) [58.3-71.1]	142 (60.2) [53.7-66.3]	141 (60.3) [53.8-66.4]	130 (49.1) [43.0-55.1]	
Loss or improper disposal	105 (11.1) [9.2-13.2]	24 (11.2) [7.6-16.2]	21 (8.9) [5.9-13.3]	28 (12.0) [8.4-16.8]	32 (12.1) [8.6-16.6]	
Unauthorized access or disclosure	140 (14.8) [12.6-17.2]	16 (7.5) [4.6-11.9]	39 (16.5) [12.3-21.9]	36 (15.4) [11.3-20.6]	49 (18.5) [14.2-23.7]	.003
Hacking or IT incident	67 (7.1) [5.6-8.9]	10 (4.7) [2.5-8.5]	20 (8.5) [5.5-12.8]	14 (6.0) [3.6-9.9]	23 (8.7) [5.8-12.8]	
Other	85 (9.0) [7.3-11.0]	25 (11.7) [8.0-16.8]	14 (5.9) [3.5-9.8]	15 (6.4) [3.9-10.4]	31 (11.7) [8.3-16.2]	
Data breach involved external vendor, No. (%) [95% CI]	273 (28.8) [25.9-31.7]	54 (25.2) [19.8-31.5]	76 (32.2) [26.5-38.5]	70 (29.9) [24.4-36.1]	73 (27.6) [22.5-33.3]	.39

Abbreviations: EMR, electronic medical record; IT, information technology.

[a] Calculated using linear regression or χ^2 tests.