A Significant Increase in The Risk for Exposure Of Health Information In The United States. Result from Analysing the US Data Breach Registry

Johan Gustav Bellika [1,2], Alexandra Makhlysheva [1], Per Atle Bakkevoll[1]

[1] Norwegian Centre for e-health research, University hospital of North Norway, Tromsø, Norway

[2] Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway

Abstract

The study surveys the probability and consequences of protected health information (PHI) data breaches. We analysed the development of data breaches in the US data breach registry available online in 2010-2016 by focusing on two PHI breach categories: theft and loss, and hacking and unauthorised use. 79% of all analysed PHI breaches was the result of hacking or unauthorised use versus 19% caused by loss or theft. Totally over 171 million persons were affected by PHI breaches during the analysed period, which corresponds to 54% of the US population. On average, 4.6 million persons are annually affected by theft or loss of PHI versus 19.4 million affected by hacking and unauthorised use of PHI. The number of hacking attacks increased by 15 times from 2010 to 2016. The largest single loss of PHI so far is 78.8 million records. The analysis has shown the risk of PHI breaches in the US is high and significantly increasing. In Scandinavian settings, such a risk would imply measures to reduce both probability and consequence of breaches.

Keywords

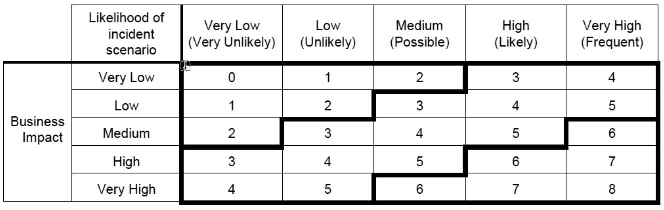
Computer Security, Cybersecurity, Risk Assessment

1 INTRODUCTION

Risk is a measure that combines the probability and impact of an undesired event(“ISO/IEC 27005 risk management standard,” n.d.). In risk analysis, risk can be estimated by computing the product of the probability that the event will occur and the consequence of the event: risk (x) = probability (x) • consequence (x). In addition, in risk analysis, consequence and probability are normally divided into specific categories, which provides the basis for graduation of risk levels. Estimates of risk levels form the basis for the evaluation of measures to reduce the risk.

One way of estimating the consequence of an event, is to look at the number of persons affected by the event. Normally, the more people affected by an undesired event, the more severe the consequence of the event will be, given that other aspects of the unwanted event are kept constant. Another approach could be to take into consideration the degree of sensitivity and amount of information exposed. For instance, the risk matrix used by Sykehuspartner, one of the reginal health authority information technology support institutions in Norway, grade consequence from catastrophic to small consequence (HSØ RHF, 2017). Similarly, the frequency of an event is a way of classifying the probability of an event. If an event occurs every fifth year, it is less likely than events occurring weekly or daily. According to (“ISO/IEC 27005 risk management standard,” n.d.), likelihood must be classified into distinct categories, for instance ranging from very unlikely through very high or frequent.

The combination of probability and impact can be expressed in a risk matrix as shown in Figure 1.



	Likelihood of incident scenario	Very Low (Very Unlikely)	Low (Unlikely)	Medium (Possible)	High (Likely)	Very High (Frequent)
Business Impact	Very Low	0	1	2	3	4
	Low	1	2	3	4	5
	Medium	2	3	4	5	6
	High	3	4	5	6	7
	Very High	4	5	6	7	8

Figure 1. Example of risk matrix from (“ISO/IEC 27005 risk management standard,” n.d.).

Risk could then for example be classified into the levels of low (0-2), medium (3-5) or high (6-8) risk.

The combination of a frequent event with very high consequences would imply maximum risk (8 in the example above). No health IT system should pass a risk analysis stage that involves maximum risk. Such a system should never make its way into production and usage by health personnel or patients.

In the United States, breaches of privacy in the health sector, which are regulated by the Health Insurance Portability and Accountability Act (HIPAA) (“Privacy | HHS.gov,” n.d.), are reported to the U.S. Department of Health and Human Services, Office for Civil Rights. Events that represent a breach of this Act and involve more than 500 persons, are published in a breach registry on the

Paper from the 15th Scandinavian Conference on Health Informatics SHI2017, Kristiansand, Norway, 29 - 30 August.

Conference Proceedings published by Linköping University Electronic Press..

Department’s Breach Portal (“U.S. Department of Health & Human Services - Office for Civil Rights,” n.d.).

Liu et. al. analysed the US Department of Health and Human Services data breach registry for the years 2010 to 2013 (Liu, Musen, & Chou, 2015). Liu et al. identified 949 breaches affecting 29 million records. Six of the events involved more than 1 million records. Lui et al. expressed concern for the increasing number of data breaches involving cloud based systems.

While the National Institute of Health in the US allows the use of cloud computing services for storage and analysis of genomics data sharing (“NOT-OD-15-086: Notice for Use of Cloud Computing Services for Storage and Analysis of Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy,” n.d.), Filkins et. al. provides a long list of advices on how a medical researcher can ensure that the cloud service provider is able to provide the necessary information security measures. At the same time, the authors note that “cloud computing represent significant unknowns, such as lack of direct control over hardware and software, lack of visibility into audit/system activities, physical location of data, and impact of different jurisdictions where the data may be held” (Filkins et al., 2016).

In May 2017, an incident occurred in Norway when a major health trust had to revoke privileges granted to employees of an international company. The employees of the international company should not have been granted access to patient data on 2.3 million Norwegian citizens, as the data was accessible to the employees globally [6]. Whether similar incidents have occurred in the past is unknown, since Norway does not currently have a publicly available data breach registry accessible to researchers.

As we do not have a breach registry available in Norway, one possibility to assess the risk for health-related data is to make estimations of probability and consequence based on data breaches registered in the publicly available US data breach registry. However, the remaining unresolved question then is whether the information security situation in Norway is comparable to the situation in the US.

2 MATERIALS AND METHODS

To survey the probability and consequence of the breaches to information security, we analysed the data from the US Department of Health and Human Services Breach Portal (“U.S. Department of Health & Human Services - Office for Civil Rights,” n.d.). As stated on the Breach Portal site, “as required by section 13402(e)(4) of the HITECH Act, the Secretary must post a list of breaches of unsecured protected health information affecting 500 or more individuals”. First PHI breaches on this portal were registered in October 2009.

To have comparable periods with full years of available data, we downloaded and analysed the reported breach events in the US health sector for the period 1st of January 2010 to 31st of December 2016 amounting to 1780 registered events. According to the Breach Portal, all attacks on protected health information can be roughly divided into several categories, such as 1) hacking/IT incident, 2) unauthorised access/disclosure, 3) loss, 4) theft, and 5) improper disposal. We grouped the identified breaches into two higher-level event categories: 1) physical theft and loss of PHI, and 2) hacking and unauthorised use of PHI. We compared breach events frequency for the years 2010 to 2016. Further, for the analysed period, we compared cyber theft (hacking and unauthorised use of PHI) and physical theft (physical theft and loss of PHI) in terms of 1) number of PHI breaches annually, 2) shares of breach events of both categories in percentage of total number PHI breaches annually, and 3) number of individuals affected by PHI breach events annually.

However, the information about the types of the health information affected by each breach event was not provided on the US Breach Portal, which could influence the conclusions about the PHI breach risk consequences. Also, the breach registry do not contain a systematic grading of the sensitivity of the information exposed, which could have been used as input to estimation of consequence of the breaches.

3 RESULTS

According to data reported to the Breach Portal between 1st of January 2010 and 31st of December 2016, protected health information for 171,074,016 persons was exposed. Although some may have had their health information exposed more than once or have health information in several health care institutions (Liu et al., 2015), this indicates that approximately 54% of the United States population of 318.9 million have had their protected health information exposed.

This number includes 135,775,362 persons who have been affected by hacking or unauthorised use of protected health information, and 31,908,209 persons affected by theft or loss of PHI during this period. On average, protected health information for 19,396,480 persons has been affected by hacking or unauthorised use annually. However, this number is heavily affected by a single event where 78.8 million health records were exposed in a single event. Excluding this event, the average yearly expose was 8,139,337 persons affected by hacking or unauthorised use of PHI.

As a comparison, 4,558,316 persons were affected by theft or loss of protected health information. This means that in the period from 2010 until the end of 2016, the probability of being exposed by cyber theft was 4.26 times larger compared to physical theft.

The distribution of observed breaches presented in Figure 2 shows that 52% of the events were classified as hacking or unauthorised use. Theft and loss of data accounted for 42% of the data breach events.

The number of events in the category hacking or unauthorised use of PHI increased from 16 cases in 2010 to 240 cases in 2016. During the same period, the number of cases of thefts or losses decreased from 154 to 78, as shown in Figure 4. As Figure 3 shows, the average number of protected health data breaches per day in the United States in 2016 was 0.89. This means that on average a breach of privacy regulations occurred more often than every second day.

The graph in Figure 5 shows the relative frequency of theft/loss and hacking/unauthorised use incidents as a percentage of all attacks on protected health information annually in the period 2010-2016. There is a steadily increasing trend for hacking and unauthorised use, and a decreasing trend for theft/loss incidents. In 2016, for instance, 10.3 times as many persons were involved in hacking/unauthorised use of health information compared to theft/loss.

Figure 6 shows the development of number of individuals affected by PHI breach incidents. From 2013, we see a rapidly increasing trend (dotted blue line) in number of individuals affected by hacking and unauthorised use of their health information. The historical development in number of this type of breach events (shown in Figure 4) is also increasing.

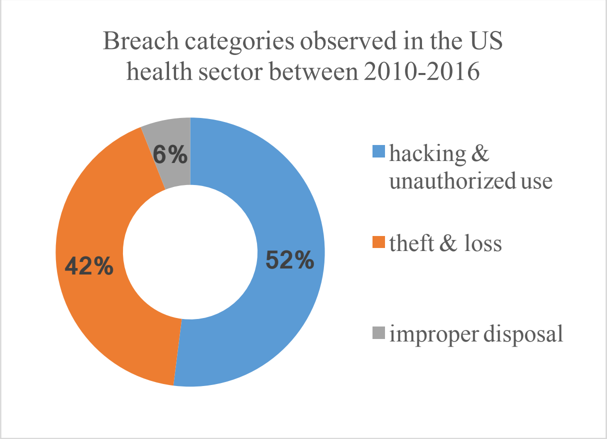


Figure 2. Observed health data breach categories.

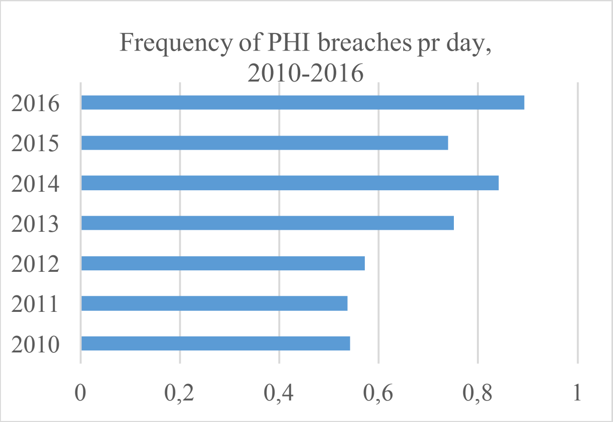


Figure 3. Frequency of PHI breaches pr day for the years 2010 to 2016.

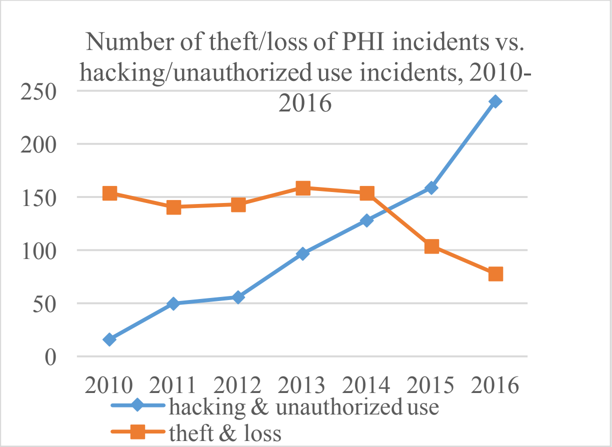


Figure 4. Historical development in number of PHI breaches for the years 2010 to 2016.

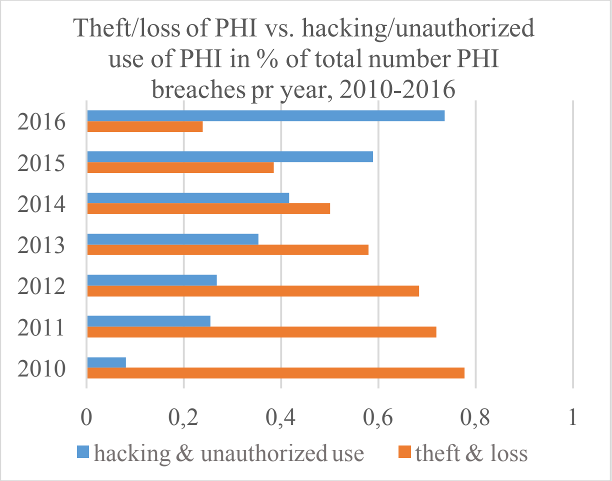


Figure 5. Historical development of the distribution of PHI breach categories for the years 2010 to 2016.

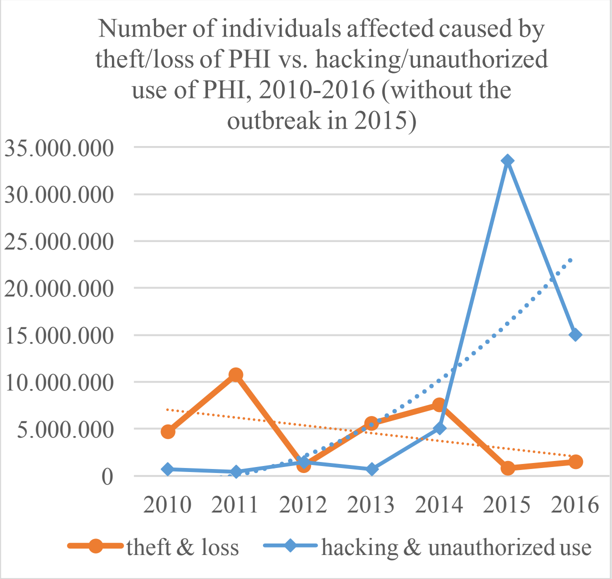


Figure 6. Historical development of number of people affected by PHI breach events for the years 2010 to 2016, excluding the event involving 78.8 million records affected in a single breach in 2015.

This trend corresponds to the increasing use of electronic health record systems in the US, as predicted by Lui et al.

At the same time, the trend for theft and data loss (dotted orange line in Figure 6) is decreasing. Based on the data reported to the US Breach Portal, hacking or unauthorised use of protected health data have to be defined as highly likely events. For the consequence measure, the annual average number of individuals affected is more than 12 million citizens, which is, using a conservative estimate, more than twice the size of the population in Norway, every year. And, the trend for number of persons affected by hacking or unauthorised use of protected health data (shown in Figure 6) is increasing.

4 DISCUSSION

Lessons learned by analysing the US data breach registry are twofold. First, a data breach registry is a very good tool to uncover and follow the development of PHI breaches across time. It provides a resource available to both patients, researchers, media and health IT managers to uncover the true probability and consequence of data breaches. Without a systematic approach to handle the reality of cyber security and failing to implement prevention measures, more data breaches may occur. Therefore, Norway should as soon as possible establish a publicly available data breach registry following the model established in the US.

The second lesson learned by analysing the data in the US data breach registry is the estimation of risk for PHI breaches we currently experience. In the United States, it is currently high, and it is increasing. If the situation in Norway is comparable, it is alarming, and measures should be implemented to protect the privacy of Norwegian citizens.

While the true risk for data breaches in Norway is currently unknown, Norwegian national health authorities are making health data about every citizen in Norway available through national web portals. A single data breach of these systems may expose sensitive health information about the majority of the Norwegian population. The decision on exposing the Norwegian population to this data breach risk has been taken without asking each individual whether they want or need to have their health records available online.

If we hypothetically assume that the risk for data breaches in Norway is high, what risk treatment can be taken to reduce the risk for data breaches to Norwegian citizens' health records? According to ("ISO/IEC 27005 risk management standard," n.d.), we can do risk modification, risk retention and risk avoidance. First, we can try to build security barriers that prevent breaches from happening. Secondly, we can do risk avoidance by reducing the consequences of data breaches. This can be done by decreasing the number of persons affected, if a data breach should occur. Alternatively, we can do risk retention by making the population to accept the risk of data breaches. This can be done by asking everyone to consent to have their data exposed to the risk of data breaches or switch to an opt-in solution where the citizen must actively ask for data to be available. As is, the current risk for data breaches may very well have consequences for the relation between patients and health workers by patients withholding sensitive health information, as noted by Blumenthal et al. (Blumenthal & McGraw, 2015) and many others. If a data breach to Norwegian national health data portals should occur, the consequences for the trust relation between the patient and health workers on one side and Norwegian national health authorities on the other will be severely negatively affected.

However, a longer historical perspective of analysed data from the US Breach Portal, together with the available types of accessed protected health information could influence the inferred conclusions about breaches of unsecured protected health information in the United States.

5 CONCLUSION

The review of the United States Breach Portal shows that the probability of PHI breaches is increasing, and is close to becoming a daily event. The extent or consequence of breaches also shows an increasing trend. In sum, this means that the risk of breaches to the privacy legislation for protected health information in the United States is very high and increasing. In Scandinavian setting, such a risk would require measures to reduce both the probability and consequence of breaches. The extent of breaches of privacy legislation causes concern among health professionals that patients will withhold health related information from health workers, and, thereby, undermine opportunities to improve their health, as well as health services (Blumenthal & McGraw, 2015).

6 REFERENCES

[1] HSØ RHF. 2017, May 24. Foreløpig redegjørelse iMod V1.pdf. Retrieved June 23, 2017, from https://www.helse-sorost.no/Documents/Styret/Styrem%C3%B8ter/2017/20170524/2017-05-24%20HS%C3%98%20RHF%20-%20Forel%C3%B8pig%20redegj%C3%B8relse%20iMod%20V1.pdf

[2] Blumenthal, D., & McGraw, D. 2015. Keeping personal health information safe: the importance of good data hygiene. JAMA, 313(14), 1424. https://doi.org/10.1001/jama.2015.2746

[3] Filkins, B. L., Kim, J. Y., Roberts, B., Armstrong, W., Miller, M. A., Hultner, M. L., … Steinhubl, S. R. 2016. Privacy and security in the era of digital health: what should translational researchers know and do about it? American Journal of Translational Research, 8(3), 1560–1580.

[4] ISO/IEC 27005 risk management standard. (n.d.). Retrieved June 23, 2017, from http://www.iso27001security.com/html/27005.html

[5] Liu, V., Musen, M. A., & Chou, T. (2015). Data Breaches of Protected Health Information in the United States. JAMA, 313(14), 1471. https://doi.org/10.1001/jama.2015.2252

[6] NOT-OD-15-086: Notice for Use of Cloud Computing Services for Storage and Analysis of Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS)

Policy. (n.d.). Retrieved May 15, 2017, from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-086.html

[7] Privacy | HHS.gov. (n.d.). Retrieved May 15, 2017, from https://www.hhs.gov/hipaa/for-professionals/privacy/

[8] U.S. Department of Health & Human Services - Office for Civil Rights. (n.d.). Retrieved May 15, 2017, from https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf

7 ACKNOWLEDGEMENT

We thank the Norwegian centre for e-health research for funding this study.