# Data Collection: Sample Member Responses, Influences, and Perception of Data Security

Tamara L. Terry, [1] McKinlay Jeannis, [1] Dawn Thomas-Banks[1]
[1]RTI International, Research Triangle Park, NC 27709-2194

## Abstract

As identity theft and data security breaches occur in the United States, in government and private organizations alike, both public opinion and sample members' willingness to provide personal identifiable information are affected. Identity theft reported by households has increased by 23% from 2005 to 2007, according to the Bureau of Justice Statistics (BJS; Baum & Langton, 2010). BJS reports that in 2007 alone, 7.9 million households—which accounts for 6.6% of all households in the United States—were victims of one or more types of identity theft.

Public opinion is commonly formed by personal experience and events that are reported in the media. One of the most infamous security breaches occurred in 2006, when a U.S. Department of Veterans Affairs laptop was stolen. It contained the names, Social Security Numbers, and dates of birth of more than 26 million veterans (*Agency chief: Data on stolen VA laptop may have been erased*, 2006). With identity theft rates increasing and security breaches being reported in the media, we perceive that research study participants are conscious of the risk and skeptical about providing personally identifiable information.

In our presentation, we intend to demonstrate how identity theft and data security breaches in the United States may have influenced sample members' willingness to provide personally identifiable information on a cross-sectional project that collected data in 2004 and 2008. We will analyze sample members' responses to questions that attempted to obtain personal information such as Social Security Number, age, sex, income, employment status, and race. Both Web and telephone data collection modes were used.

**Key Words:** identity theft, data security breaches, public opinion, personally identifiable information (PII)

## 1. Data Security Breaches and Identity Theft Impacts

Data security is a means of ensuring that data is kept safe from corruption and that access to it is suitably controlled. In essence, it helps to ensure privacy of data. The Web site securitysearch.com describes a data breach as "an incident in which sensitive, protected or confidential data has potentially been viewed, stolen or used by an individual unauthorized to do so. Data breaches may involve personal health information (PHI), personally identifiable information (PII), trade secrets or intellectual property." Merriam Webster describes identity theft as "the illegal use of someone else's personal information in order to obtain money or credit." This paper explores how data security

breaches and identity theft potentially affect sample member responses and willingness to answer highly sensitive questions that collect personal information.

## 2. Thesis

For this research, we identified a national cross-sectional project that collected data in 2004 and 2008 using both Web and telephone data collection modes. We sought to analyze sample members' responses to questions that attempted to obtain personal information such as Social Security Number (SSN), age, sex, income, employment status, and race. Specifically, we compared sample members' responses from 2004 to those in 2008 to see whether the rise in security breaches affected nonresponse. We analyzed the following:

- Sample member responses to questions that collected highly sensitive and nonsensitive information for a survey conducted in 2004 and 2008
- Responses by particular demographic groups to highly sensitive and nonsensitive questions collecting personal information for a survey conducted in 2004 and 2008
- Potential mode effects in 2004 and 2008

### 2.1 Background Research Findings

To help evaluate our thesis, we reviewed legislation changes, data security breaches, and identify theft changes from 2004 to 2008. As of 2004, California was the only state with security breach notification legislation, but by 2008, 37 states had enacted such laws (National Conference of State Legislatures, 2010).

The 2003 yearly report from the Data Loss Database reported 16 publicly reported data breaches involving PII and affecting 7,061,950 records. The 2007 yearly report reported 511 publicly reported data breaches involving PII and affecting 165,262,288 records.

Finally, the number of households with at least one member who experienced one or more types of identity theft increased 23% from 2005 to 2007. From 2005 to 2007, the number of households that experienced credit card theft increased by 31%, and the number that experienced multiple types of identity theft during the same episode increased by 37% (Baum & Langton, 2010).

## 3. Responses to Highly Sensitive and Nonsensitive Questions

To depict sample members' willingness to provide PII on a cross-sectional project that collected data in 2004 and 2008, we analyzed their responses to questions asking for highly sensitive and nonsensitive information. The sample members' willingness to provide responses is indicated by "declined" if they decided to refuse or skip particular questions or by "provided" if they responded. We used sample members' willingness to provide their SSNs as a proxy for highly sensitive information. For nonsensitive information, we examined their willingness to provide a response to race, income, and employment questions.

### 3.1 Highly Sensitive: Social Security Number Comparison

The highly sensitive information comparison yielded some interesting findings in both data collection periods. The number of sample members who declined or provided responses to highly sensitive questions during the 2004 data collection period was almost

evenly split: 52% provided a response and 48% declined to provide a response to the SSN question. In the 2008 data collection period, 68% declined to respond and 32% provided a response to the SSN question (see Figure 1).

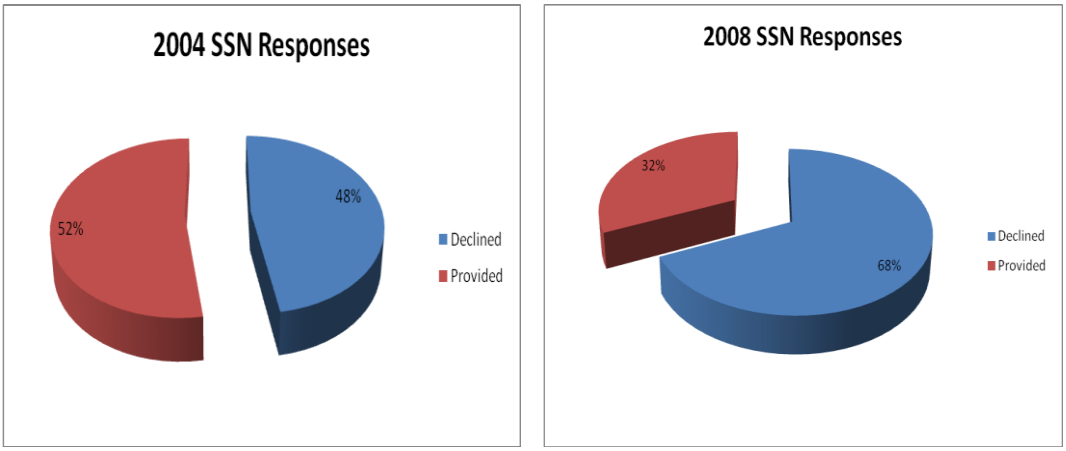


**Figure 1:** Highly sensitive information

### 3.2 Nonsensitive: Race, Income, and Employment Comparison

In examining several variables of nonsensitive information, we found that response rates followed a pattern similar to that for the highly sensitive information: the percentage of sample members who refused to provide a response increased from 2004 to 2008. During the 2004 data collection period, 96% provided a response to the race question, and only 4% declined. The 2008 data collection period did not show a huge variation, in that 87% provided a response to the race question and 13% declined (see Figure 2).

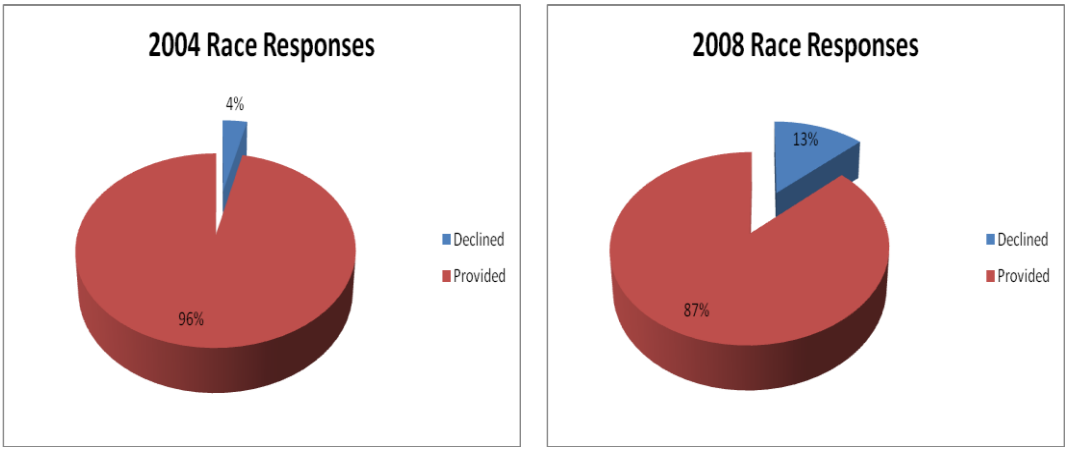


**Figure 2:** Nonsensitive information (race)

The second variable of the nonsensitive information we analyzed was a question pertaining to the subject's income. We found that 95% of the sample members in 2004 provided a response for the income question and 87% provided a response in 2008. Those who declined to provide a response to the income questions were 5% in 2004 and 13% in 2008 (see Figure 3).

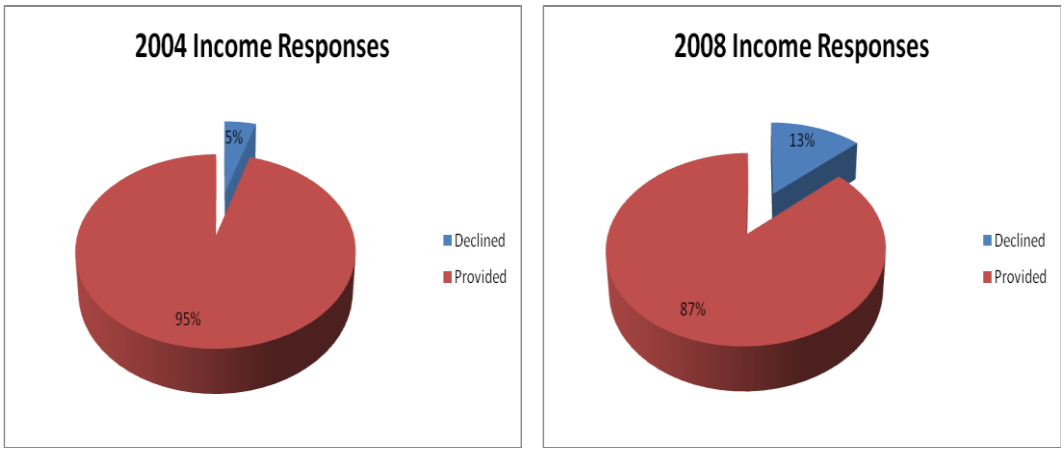


**Figure 3:** Nonsensitive information (income)

Last, the employment responses also illustrated similar trends across both data collection periods for the nonsensitive information. In 2004, 98% provided a response to the employment question in contrast to 2% who declined. In 2008, with the majority of the sample still willing to provide a response to employment, 88% provided a response and 12% declined (see Figure 4).

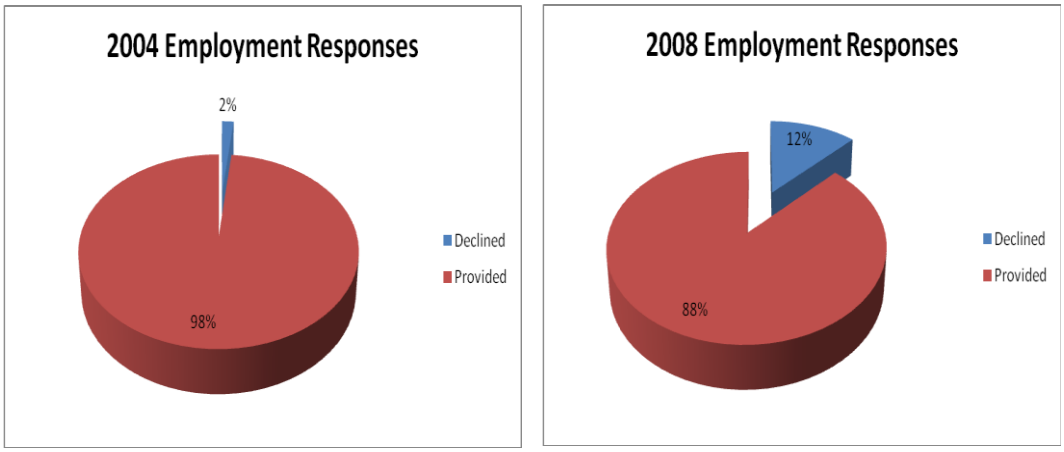


**Figure 4:** Nonsensitive information (employment)

### 3.3 Highly Sensitive and Nonsensitive Data Comparison Findings

Overall, from the 2004 to 2008 data collection periods, the percentage of participants who decline to provide both highly sensitive and nonsensitive information has increased. Examining the difference between highly sensitive information and nonsensitive information, we found that in both 2004 and 2008 a higher percentage of participants declined to provide a response to the SSN question, considered to be highly sensitive information, than to nonsensitive questions, including those about income, race, and employment.

Another trend we identified is the similarity of the percentage of participants who decline to provide all three pieces of nonsensitive information within each data collection period. For example, 5% or fewer declined to provide nonsensitive information in 2004, and 13%

or fewer declined to provide nonsensitive information in 2008. It appears that participants who declined to provide nonsensitive information for one question variable also refused to provide a response for the remaining nonsensitive question variables.

The most significant finding in the comparison of highly sensitive and nonsensitive information during the 2004 and 2008 data collection periods were the responses to the SSN question. The percentage of participants who declined to provide their SSNs increased considerably.

## 4. Highly Sensitive Responses by Demographic Groups

To further depict the impact of identity theft and data security breaches in the United States on sample member responses, we compared the various participant demographics with participants' willingness to provide their SSNs. We used the same cross-sectional project, which collected data in both 2004 and 2008. The purpose of evaluating the various demographics was to determine whether the participants' willingness to provide SSNs was limited to other factors, such as generation, gender, or race.

### 4.1 Social Security Number Provided by Generation

The first demographic variable that we examined was generation. The study participants were grouped into the following generational categories: Generation Y (ages 4–24 years at the time of completing the study), Generation X (ages 25–39), Baby Boomers (ages 40–58), and the Silent Generation (ages 59–79).

In 2004, 46% of the Generation Y participants declined to provide their SSN, compared with 51% of Generation X, 56% of Baby Boomers, and 63% of the Silent Generation. In 2008, participants classified within these same generations declined at the following rates: 68% Generation Y, 69% Generation X, 68% Baby Boomers, and 100% Silent Generation (see Figure 5).

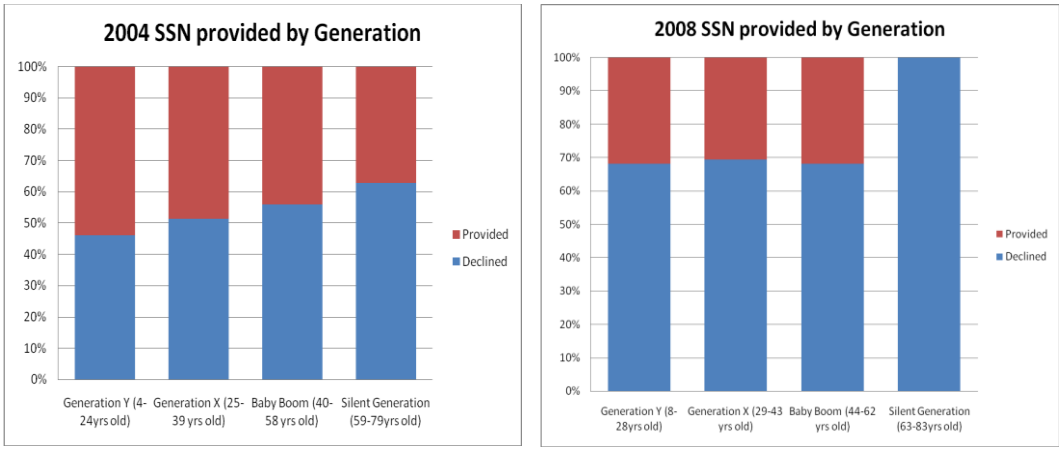


**Figure 5:** Highly sensitive responses by demographic groups (generation)

## 4.2 Social Security Number Provided by Gender

The second demographic variable that we assessed was gender. In 2004, 44% of the male participants and 50% of the female participants declined to provide their highly sensitive information. Once again, we observed an increase in the 2008 data collection period: 63.52% of male and 71.79% of female respondents declined to provide responses to highly sensitive questions (see Figure 6).

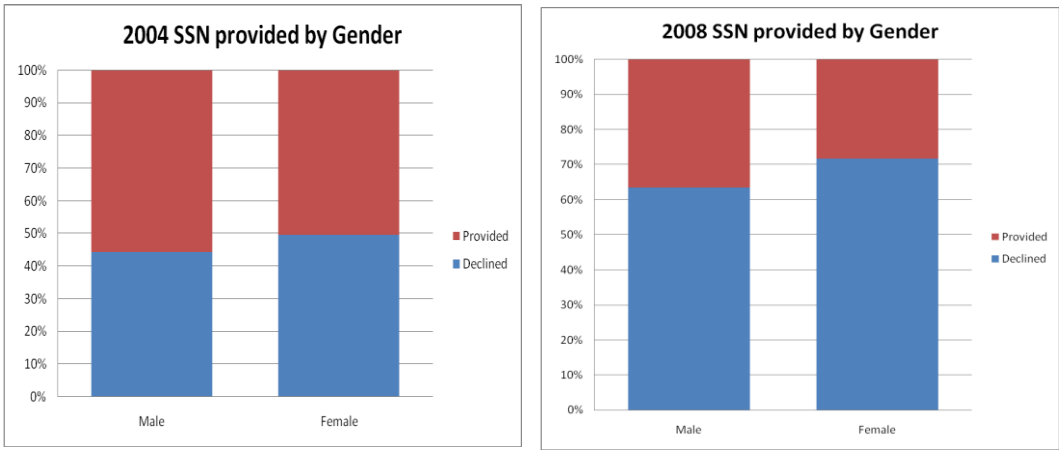


**Figure 6:** Highly sensitive responses by demographic groups (gender)

## 4.3 Social Security Number Provided by Race

Last, we explored whether race may have played a role in the declined responses to highly sensitive questions in 2004 and 2008. The racial categories of the participants during both data collection periods were Asian or Pacific Islander; Hispanic, regardless of race; White, not of Hispanic origin; Black, not of Hispanic origin; and American Indian or Alaskan Native. In 2004, 51% of the Asian or Pacific Islander, 46% of the Hispanic, regardless of race; 48% of the White, not of Hispanic origin; 47% of the Black, not of Hispanic origin; and 38% of the American Indian or Alaskan Native respondents declined to provide their SSNs. The percentage of refusal increased in the 2008 data collection period to 71% of the Asian or Pacific Islander; 65% of the Hispanic, regardless of race; 68% of the White, not of Hispanic origin; 71% of the Black, not of Hispanic origin; and 56% of the American Indian or Alaskan Native respondents (see Figure 7).

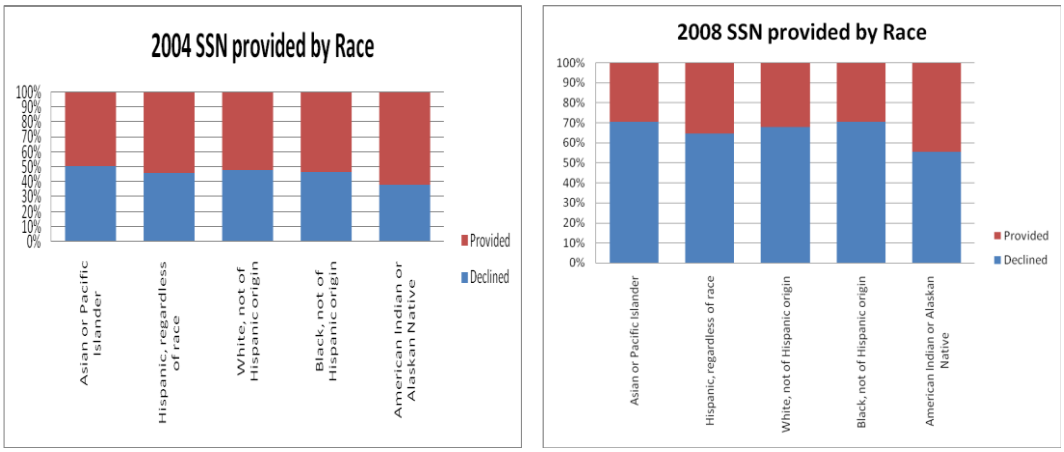


**Figure 7:** Highly sensitive responses by demographic groups (race)

### 4.4 Findings

On the basis of our analysis, we concluded that all demographics' percentages in declining to provide their SSNs increased from 2004 to 2008. Within data collection periods, as the participant age increases, the percentage of sample members declining to provide SSNs also increases. In both 2004 and 2008, females respondents had a higher percentage declining to provide their SSNs. Asians also displayed a higher percentage than other racial groups of participants declining to provide their SSNs for both data collection periods.

## 5. Mode Effects

We hope to continue our research by further examining how participants' concerns about data security and identity theft could be addressed by the modes of data collection to which they respond. Data from the 2004 and 2008 data collection periods suggests that participants are apprehensive about providing highly sensitive information regardless of data collection mode. In 2004, 51% of the self-administered Web questionnaire subjects, as well as 45% of the subjects who participated by telephone, declined to provide their SSNs. The percentage of subjects who declined to respond to highly sensitive questions increased in 2008 to 67% of Web and 73% of telephone respondents (see Figure 8).

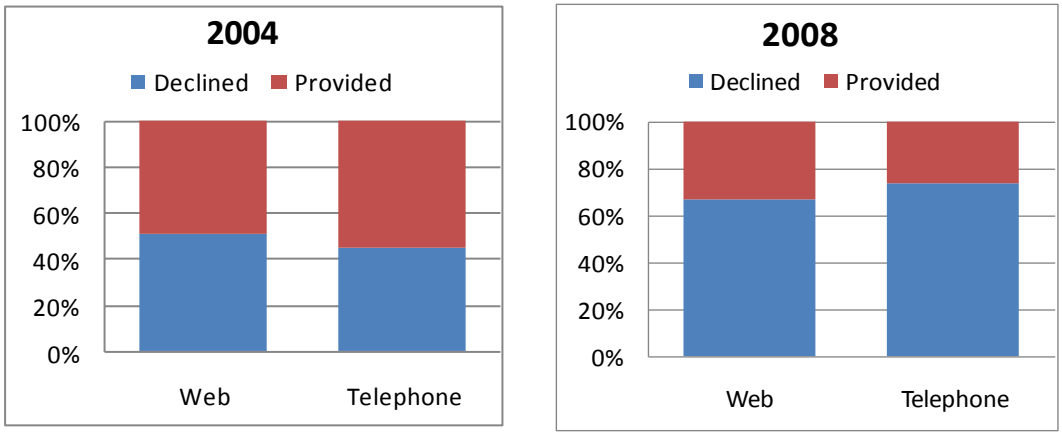


**Figure 8:** 2004 and 2008 mode effects

## 6. How RTI International Addresses Data Security

To address study participants' data security and identity theft concerns, RTI International has taken multiple proactive measures, including the following:

- data security compliance implementation, such as enhanced security network (ESN), IT security and compliance, information security officer and privacy officer (institutional review board)
- data handling guideline, IT security awareness training, project-specific data security training, and regular data security communications
- key moderate-risk security controls, which allow limited e-mail usage, encrypted printing, two-factor authentication, and limited Internet access

- physical security controls: restricted access, fire suppression, climate controls, uninterruptible power supplies, cameras on access doors, visitor escorts, visitor sign-in and sign-out, periodic inspections by security staff, and periodic audits of data center and visitor logs

- other security measures: continuous monitoring program, periodic and unannounced audits, regular agency testing, continuous process improvement Information Technology Infrastructure Library (ITIL), Plan of Action and Milestones (PoAM) tracking, information security incident reporting and tracking, document management, federal regulation and best practices monitoring

## 7. Conclusion

We postulated that increased awareness of data security breaches and identity theft would be highly correlated with sample members' willingness to provide personal information such as SSN, race, income, employment, age, and gender. Overall, our assumption appeared to be correct. Our findings show a significant decline in provided responses to questions collecting personal information in 2008 compared with responses provided in 2004. The security breach notification laws enacted in 37 states from 2004 to 2007 forced agencies to report data breaches involving PII to affected individuals. Tehan (2007) attributes the increased awareness in security breaches to these laws.

## Acknowledgement

Joe Simpson, a senior research programmer/analyst at RTI International, provided the raw data for our analysis.

## References

*Agency chief: Data on stolen VA laptop may have been erased.* (2006, June 8). Retrieved from http://www.cnn.com/2006/US/06/08/vets.data/index.html?iref=allsearch

Baum, K., & Langton, L. (2010, June 30; revised September 29, 2010). *National Crime Victimization Survey, 2007: Identity theft reported by households, 2007—Statistical tables* (NCJ 230742). Washington, DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. Available from http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=2294

MerriamWebster.com. (n.d.) *Identity theft*. Retrieved from http://www.merriam-webster.com/dictionary/identity+theft?show=0&t=1317759961

National Conference of State Legislatures. (2010, October 12). *State security breach notification laws.* Washington, DC: Author. Retrieved from http://www.ncsl.org/Default.aspx?TabId=13489

SearchSecurity.com. (2010, March). *Data breach*. Retrieved from http://searchsecurity.techtarget.com/definition/data-breach

Tehan, R. (2007, May). *Data security breaches: Context and incident summaries* (Congressional Research Service Report No. RL33199). Washington, DC: Congressional Research Service.