

# Data Aggregators: A Study of Data Quality and Responsiveness

Deborah Pierce ([dsp@privacyactivism.org](mailto:dsp@privacyactivism.org))  
Linda Ackerman ([lga@privacyactivism.org](mailto:lga@privacyactivism.org))  
May 19, 2005

## Executive Summary

This study examined the quality of data provided by ChoicePoint and Acxiom, two of the largest consumer data brokers in the United States, as well as their responsiveness to consumer requests – and found significant areas of concern in both areas.

100% of the reports given out by ChoicePoint had at least one error in them. Error rates for basic biographical data (including information people had to submit in order to receive their reports) fared almost as badly: Acxiom had an error rate of 67% and ChoicePoint had an error rate of 73%. In other words, the majority of participants had at least one such significant error in their reported biographical data from each data broker.

**Table 1: Overall Accuracy**

	<b>Acxiom (biographical fields only) n = 6</b>	<b>ChoicePoint (biographical fields only) n = 11</b>	<b>ChoicePoint (all fields in study) n = 11</b>
Percentage of participants with at least one error in report	67%	73%	100%
Average errors per report	0.8	1.4	5.4
Average data items in report	6.3	5.6	16.6
Overall error rate (incorrect / total data items)	13% (5/37)	25% (15/61)	35% (58/178)

In terms of responsiveness, Acxiom only fulfilled 55% of the requests by consumers for reports in the 120-day period of the study, and the average response time for successful requests was 89 days. ChoicePoint, by contrast, responded quickly to such requests: all participants received their reports, with an average response time of nineteen days.

**Table 4: Responsiveness (n = 11)**

<b>Company</b>	<b>Acxiom</b>	<b>ChoicePoint</b>
<b>% requests fulfilled</b>	55%	100%
<b>Average time for response</b>	89 days	19 days

The small sample size of this study (eleven participants) means that care must be taken in projecting these results to the overall population. However, as pointed out in the 2004 *FTC Round-table on Accuracy and Completeness in Credit Reports*<sup>1</sup>, the difficulties and expense of performing large studies mean that it is appropriate to begin with investigative studies such as this.

The extremely high error rates for both companies and Acxiom's lack of responsiveness indicates the likelihood that these problems are likely to occur on a broader scale. The results clearly point out the need for a larger study on the critical issues of data broker accuracy and responsiveness.

---

<sup>1</sup> <http://www.ftc.gov/be/workshops/methodologiesaacc/summary.pdf>, p. 12

# 1. Background and Motivation

A consumer reporting agency (CRA) gathers and packages personal information into consumer reports which they sell to creditors, employers, insurers, and other businesses. The best-known CRAs are credit bureaus such as Equifax, Experian and TransUnion. The United States Fair Credit Reporting Act (FCRA), as amended by the Fair and Accurate Credit Transactions Act of 2003 (FACTA), regulates CRAs<sup>2</sup>.

Data brokers (also known as data aggregators) such as ChoicePoint and Acxiom collect information from public records, criminal databases, and other sources, and package it into reports that they sell to businesses as well as local, state, and federal government agencies (as well as aggregate reports useful for marketing purposes, which are outside the scope of this study). Some, but not all, of data brokers' activities fall under the FCRA.

Both ChoicePoint and Acxiom provide many different kinds of reports. Individuals can request their own *consumer reports* (for \$5 from Acxiom, \$20 from ChoicePoint)<sup>3</sup>, which contain basic biographical information such as name, date of birth, current address, and phone number. Eligible and qualified third parties can request *employee background check reports*, which contain far more detailed information such as past addresses and length of residence, business affiliations, professional licenses, and criminal history.

This data is used for many important purposes, including employee background checks, decisions about insurance coverage and pricing, and law enforcement. As a result, the accuracy of the information provided ("data quality") is an important issue. While several previous studies have examined the data quality in reports from credit bureaus<sup>4</sup>, these results do not necessarily apply: data brokers gather data from different sources than credit bureaus, and may have different handling practices. While no previous published studies have addressed this question, anecdotal evidence points to the possibility of significant inaccuracies in data broker reports<sup>5</sup>.

# 2. Methodology

This study uses a consumer-review methodology: participants review data broker reports about themselves for correctness. Most participants in the FTC Round-table agreed that consumers are in the best position to identify inaccuracies in their own reports. The Round-table also identified potential issues with this methodology as applied to credit information, but because of the different focus of this study, we were able to design the

---

<sup>2</sup> see the Federal Trade Commission's web site at <http://www.ftc.gov/privacy/privacyinitiatives/credit.html>

<sup>3</sup> As of 2005, ChoicePoint provides consumer reports free of charge to individuals

<sup>4</sup> For example, [\*Mistakes Do Happen: A Look at Errors in Consumer Credit Reports\*](#), National Association of State PIRGs, June 2004.

<sup>5</sup> "ChoicePoint files found riddled with errors", Bob Sullivan, MSNBC, March 8, 2005; available at <http://www.msnbc.msn.com/id/7118767/>

study to mitigate all of these potential problems except for the possibility of intentional or unintentional bias by the participants. [Potential issues and their mitigations are discussed in more detail in Appendix A.]

Participants were volunteers from a pool of employees at a California-based company that was performing ChoicePoint ScreenNow background checks on employees for purposes related to their business. Volunteers were asked to order their own consumer reports. The company ordered and furnished the volunteers with copies of their ChoicePoint ScreenNow employee background check reports. Participants then filled out questionnaires about the information in the reports from each company. Eleven participants completed the study by requesting a consumer report from both ChoicePoint and Acxiom, receiving a report from at least one company, and filling out at least one questionnaire; as discussed in more detail in the responsiveness section, six of the eleven received their consumer reports from both companies.

The questionnaire was developed from a previously obtained sample of a ChoicePoint employee background check. Participants checked “yes” or “no” to indicate whether the information about them in the various fields in the report was correct or incorrect. In some cases, additional questions were asked to determine missing data (e.g., “do you have any phone numbers not listed on the report?”). Participants also recorded whether they actually received their reports, and if so, how long it took from when they sent in the request.

The questionnaires were returned anonymously to the employer, with no personal information that could identify a participant to the person tabulating the results. The employer verified that the questionnaires had been filled out, and in turn provided them to the authors of this report.

Based on the questionnaires, the accuracy of data item on the report was determined for the following fields: name, social security number (SSN), date of birth, current and past addresses, length of residence at current and past addresses, phone number(s), real property owned, purchase/sale dates of real property.

For the data in this study, the definition of *error* is straightforward: a data item is in error if the value provided in the report is different from the actual value. Examples of errors include an incorrect phone number, or an address where the person never lived. [Appendix B provides a detailed definition on the classification used.]

Not all errors are equally important. Depending on the way in which the report is used, different data may or may not be important: for example, information about length of residence in a particular location may or may not be significant from the standpoint of a background check, but may be material when used to infer how deeply an individual is “rooted” in their community, as is being considered for anti-terrorism purposes. In some analyses in this study, we distinguish the most basic biographical information (name, date of birth, SSN, phone numbers, and current address), which is likely to be significant in virtually all cases.

It is important to note that due to the small size of the data set, and the non-random distribution of participants, care should be taken in projecting this data to the full population. As participants in the FTC Round-table noted, a detailed national study would be extremely difficult and expensive, and so it is worthwhile beginning with smaller studies such as this.

It should also be stressed that this study was not designed to allow any comparison between the relative accuracy of the reports from different companies, and it would be a severe misuse of the data to draw any such conclusions. Results on ChoicePoint are based on information both in the consumer reports and the ScreenNow employee background reports; results on Acxiom are based solely on information in consumer reports.

### **3. Findings**

#### **Accuracy**

The study looked at accuracy from two different perspectives. One is from the standpoint of the individual: what percentage of consumers have errors in their reports? The other is in terms of the data: what is the *error rate* of data items in the reports – that is, of the total number of data items in reports, what percentage are inaccurate?

The key findings in these areas are:

- A majority of participants (73% for ChoicePoint, 67% for Acxiom) found errors in their reported biographical data.
- The error rate for basic biographical data (name, date of birth, SSN, current address, phone number) was 25% for ChoicePoint and 13% for Acxiom.
- The overall error rate was 35% for ChoicePoint. [Since only basic biographical data was reported for Acxiom, the overall error rate is the same as the error rate for basic biographical data: 13%.]

**Table 1 (repeated): Overall Accuracy**

	<b>Axiom (biographical fields only) n = 6</b>	<b>ChoicePoint (biographical fields only) n = 11</b>	<b>ChoicePoint (all fields in study) n = 11</b>
Percentage of participants with at least one error in report	67%	73%	100%
Average errors per report	0.8	1.4	5.4
Average data items in report	6.3	5.6	16.6
Overall error rate (incorrect / total data items)	13% (5/37)	25% (15/61)	35% (58/178)

Accuracy varied widely for the different categories of data in the report. Unsurprisingly, accuracy was somewhat higher on the data such that was furnished as part of the request (name, social security number, and address); interestingly, though, even for these fields the error rate was approximately 10%. Property ownership information was completely accurate (an error rate of 0%); conversely, length of residence at current and past addresses (provided only in the ChoicePoint ScreenNow report) was usually wrong (an error rate of 80%).

**Table 2: Accuracy of biographical data**

	<b>Axiom n = 6</b>	<b>ChoicePoint n = 11</b>
Percentage of participants with at least one error in biographical data	67%	73%
Average errors per participant	0.8	1.4
Average data items per participant	6.3	5.8
Error rate	13% (5/37)	25% (15/61)

**Table 3: Error rate by category of data (incorrect / total data items)**

Category	Acxiom error rate	ChoicePoint error rate
Name *	17% (1/6)	9% (1/11)
Date of Birth *	17% (1/6)	0% (0/11)
SSN *	25% (1/4)	9% (1/11)
Most recent address	14% (1/7)	9% (1/11)
Previous addresses	N/A	16% (6/37)
Length of residence	N/A	83% (35/42)
Phone number(s)	9% (1/11)	67% (12/18)
Property owned	0% (0/3)	0% (0/20)
Purchase date of property	N/A	11% (2/18)

\* data was provided by requestor as part of the request for a report  
N/A: data does not appear in the report

Although outside of the specific data that was the focus of the study, several other inaccuracies reported by the participants are worth mentioning:

- 100% of participants had at least one phone number omitted in reports from ChoicePoint.
- Three different participants were incorrectly reported as “officers of corporations” in the ChoicePoint reports.
- One participant’s ChoicePoint report had several of her ex-husband’s addresses listed under her name
- One participant’s Acxiom report had an incorrect gender

## Responsiveness

Both companies in the survey offer consumer reports for a charge (\$5 for Acxiom, \$20 for ChoicePoint) comparable to accessing credit reports. The study looked at two questions related to the responsiveness of the two companies to these requests: what percentages requests are fulfilled, and how long is the elapsed time between when a consumer requests a report and the report arrives.

As the table shows, responsiveness varied dramatically between the two companies:

**Table 4 (repeated): Responsiveness (n = 11)**

Company	Acxiom	ChoicePoint
% requests fulfilled	55%	100%
Average time for response	89 days	19 days

More than one-third of the participants did not receive their report from Acxiom; and even for those who did receive a report, the average wait time was almost three months. These same participants were able to request and receive their report from ChoicePoint (in many cases resubmitting their request after ChoicePoint asked for additional

information), so it seems unlikely that this is primarily due to user error in requesting the reports.

There may, however, be a potential bias in the ChoicePoint responsiveness data for consumer reports due to the additional requests for ScreenNow employee background reports for the same individuals. As a result, as with other data in the study, this data should not be used to compare the two companies.

## **4. Conclusion**

The results from the study strongly imply a high rate of serious errors in the information provided by two of the largest data brokers in the United States, as well as a lack of responsiveness to consumers requesting their own information. While the small sample size means we do not consider these results definitive, the figures unequivocally point to a need for a much larger study.



## Appendix A: Mitigations for potential methodology issues

Discussions of studies of credit data<sup>6</sup> highlight several potential methodology issues with consumer-review methodologies. This appendix discusses these issues and how this study mitigates most of them:

1. Consumers may be confused by the way in which information is reported, and may not recognize items that belong to them. Mitigation: the study was restricted to focus on specific data items in which the presentation is simple, and which only contained information relating to the individual.
2. Consumers may mistakenly challenge items because they do not understand their credit obligations. Mitigation: the study was restricted to data items that are easy to understand. [The one possible exception to this was the property ownership data item, where the reported error rate was zero.]
3. Some accurate items may be perceived as inaccurate because of lapses in the consumer's memory or failure to maintain and consult the appropriate records. Mitigation: the data items in the study at risk for this (past addresses, length of residence) were only included in the "error rate for all fields in study", not the discussion of biographical errors.
4. Consumer answers to a survey may be biased, either intentionally or unintentionally. Mitigation: the specific example discussed in the FTC report ("consumers may do a better job of remembering and confirming favorable information than derogatory information") does not apply to the data items covered in this study, although the possibility of participants intentionally providing biased data remains.
5. Consumers may not generally understand credit reporting procedures. Mitigation: the study was restricted to data items not requiring any understanding of procedures.
6. Consumer-review methodologies have difficulty in distinguish between "material" and "immaterial" inaccuracies. Mitigation: to avoid the "materiality" issue in general, we instead separated the data into basic biographical data (likely to be material for any use of the report) and other data (whose materiality depends on the use of the report). Note that this separation was done by the designers of the study, not by individual consumers.

---

<sup>6</sup> <http://www.ftc.gov/be/workshops/methodologiesaacc/summary.pdf>.

## Appendix B: Definitions of Accuracy

The data in each field in the report is classified as follows

Code	Meaning	Definition
C	Correct	The data in the field exactly matched the true data
I	Incorrect	The data in the field differed from the true data
N	Not present	No data was provided for the field
P	Partial	Not all of the data was provided, but the data that was there was correct
O	Other	The data applied to some other person, not the person who was the subject of the report
D	Duplicate	The data duplicated other data in the report
X	Not applicable	No data should have been provided in this field

For the purposes of the statistics in this report, we treated Partial data conservatively; that is, equivalent to Correct.

Code	Treatment in error rate calculations
C	Not an error
I	Error
N	Ignored
P	Not an error
O	Error
D	Ignored
X	Ignored