

# EXHIBIT D

Oracle USA, Inc., *et al.*

)

Plaintiffs,

)

v.

) Case No. 07-CV-01658 PJH (EDL)

)

SAP AG, *et al.*

)

Defendants.

)

---

**EXPERT REPORT OF DANIEL S. LEVY, Ph.D.**

November 16, 2009

Confidential

**Table of Contents**

1) **Executive Summary** .....1  
2) **Introduction and Assignment** .....7  
3) **Qualifications** .....7  
4) **Information Considered** .....8  
5) **Background** .....8  
6) **Questions of Interest** .....8  
7) **Sampling** .....9  
    a. **General Description** .....9  
    b. **Protocol** .....13  
8) **Measurement of Results** .....18  
    a. **Population Count Measures** .....18  
    b. **Extrapolation Methodology for Sample Measures** .....20  
9) **Results** .....24  
10) **Conclusion** .....36  
**Appendix 1 – Technical Appendix** .....37  
    a. **Mean Per Unit Estimator – Related Formulas** .....37  
    b. **Union Measures – Related Formulas** .....38  
    c. **Sum of Two Measures – Related Formulas** .....38  
    d. **Ratio Estimator – Related Formulas** .....38  
    e. **Estimates for Lower and Upper Bounds Based on Repeated Sampling** .....39  
**Appendix 2 – Additional Figures and Tables** .....41  
**Appendix 3 – Curriculum Vitae** .....43  
**Appendix 4 – Information Considered** .....50

## 1) Executive Summary

I have been retained by counsel for the Plaintiffs in the matter of Oracle International Corporation, *et al.* v. SAP AG, *et al.* (Case No. 07-CV-01658 PJH (EDL)) to design a statistically-valid sample of Defendant TomorrowNow Inc.'s ("SAP TN") Fixes to PeopleSoft HRMS that can be used to reliably estimate the number of Fixes delivered to customers by SAP TN that infringed Oracle copyrights or otherwise resulted from impermissible cross-use of Oracle's software.<sup>1</sup>

Oracle is interested in measuring the activity that SAP TN was engaged in to develop, test, and distribute the Fixes created through either SAP TN's Retrofit Support Model<sup>2</sup> or its Critical Support Model,<sup>3</sup> through the use of copies of Oracle software. Oracle's computer forensics expert, Mr. Mandia, has designed various measures to capture SAP TN's activity. Mr. Mandia has gathered the information for many of these measures for the entire population of Retrofit and Critical Support Fixes to demonstrate the extent of SAP TN's activity. However, because of the enormous amount of time and effort that would be required to gather data for many of the other measures of interest, Oracle has collected some information for Retrofit Fixes and Critical Support Fixes based on a scientific random sample

---

<sup>1</sup> Based on conversations with Oracle's computer forensics expert, Mr. Mandia, it is my understanding that a Fix is any software application patch, fix, code change, or Update. It is my understanding that an Update is a group of Fixes delivered together in a single deliverable, either by Oracle or SAP TN. Fixes typically address known or reported issues with the functionality of the software. For example if the minimum wage in California went up, there would be a Fix that applied to the incorporation of that regulatory change to the software.

<sup>2</sup> Based on conversations with Mr. Mandia, it is my understanding that, in the Retrofit Support Model, SAP TN used a published Oracle Update to generate an Update for earlier software releases. The Retrofit Support Model required SAP TN to make many copies of Oracle software to produce a Fix.

<sup>3</sup> Based on conversations with Mr. Mandia, it is my understanding that the Critical Support Model involved a similar process as the Retrofit Support Model, except that it did not typically involve using an Oracle Update as SAP TN's source. However, both models generally involved developing and testing a given Fix in multiple Environments and then providing the Fix to multiple customers. Both the Retrofit Support Model and the Critical Support Model required SAP TN to make many copies of Oracle software to produce a Fix.

of the Retrofit population and Critical Support population of Fixes. The Retrofit Fix sample size is 46 Fixes out of a total population of 212 Fixes. The Critical Support sample size is 238 Fixes out of a population of 973 Fixes. I use standard statistical formulas to extrapolate results from the sample to the entire population of Fixes. Later in this report, I will discuss the construction of the sample, the use of sampling to learn information about a wider population, and the extrapolation of results. I will also provide details about the measures for which Mr. Mandia gathered information. Oracle is interested in capturing SAP TN's activity as it related to the use of Objects<sup>4</sup> and Environments.<sup>5</sup> In this executive summary, I highlight some of the measures that address these particular topics.

One piece of information collected for the samples of Retrofit and Critical Support Fixes was the number of customers that received a Fix as a First or Identified Deliverable.<sup>6</sup> Using the

---

<sup>4</sup> Based on conversations with Mr. Mandia, I understand an Object to mean, with respect to the PeopleSoft product family, a File-based Object, a DAT file, or a DMS file. A File-based Object refers to PeopleSoft COBOL (Common Business-Oriented Language, a programming language with its primary domain in business, finance, and administrative systems for companies and governments) source code files, SQR (Structured Query Reporter, a programming language designed for generating reports from database management systems) files, and SQC (Structured Query Language Common Code) files. A DAT file is a data file or a file with a .dat extension. A DMS file is a Data Mover Script file. I understand there might be additional Object types not included in this analysis.

<sup>5</sup> Based on conversations with Mr. Mandia, it is my understanding that an Environment is the combination of an installation or copy on SAP TN systems of Oracle Enterprise Application Software and a corresponding database. I further understand from Mr. Mandia that SAP TN used Environments as a crucial tool in its creation of Fixes in both the Retrofit and Critical Support Models.

<sup>6</sup> Based on conversations with Mr. Mandia, it is my understanding that a First Deliverable is one or more Objects (typically, a compressed file) received by an SAP TN customer, as indicated on Delivered Updates and Fixes. A First Deliverable is a combination of a customer and a Fix because it is the first time a customer received a particular Fix. An Identified Deliverable is an occurrence in which an SAP TN customer received a Fix, as indicated in the SAS database.

Based on conversations with Mr. Mandia, it is my understanding that Delivered Updates and Fixes refers to the unique set of PeopleSoft HRMS Fixes that were delivered to SAP TN customers based on Mr. Mandia's analysis of three sources: TN Hard Drive 78, TN Disc 09, and TN Disc 186.

sample of Retrofit Fixes, I find that on average, 7.13 customers received each Retrofit Fix as a First or Identified Deliverable.<sup>7</sup> Using this average, I estimate that the total number of instances in which customers received a First or Identified Deliverable in the Retrofit population is 1,511.<sup>8</sup> Using the sample of Critical Support Fixes, I find that on average, 26.79 customers received a Critical Support Fix as a First or Identified Deliverable.<sup>9</sup> Using this average, I estimate that that total number of instances in which customers received a First or Identified deliverable in the Critical Support population is 26,070.<sup>10</sup>

For each Fix in the Retrofit and Critical Support samples, Mr. Mandia also counted the number of Objects comprising First Deliverables or Identified Deliverables, and the number of copies of those Objects.<sup>11</sup> Such Objects are termed Associated Files.<sup>12</sup> Using the sample of Retrofit Fixes, I find that on average, there are 237.02 Associated Files for a Retrofit

---

Based on conversations with Mr. Mandia, it is my understanding that SAP TN used the SAS database to manage its relationships with customers, for instance by tracking information such as the relevant software version for the client and the point of contact at the client.

<sup>7</sup> Based on my results the 90% confidence interval ranges from 5.79 to 8.46. This means that in repeated samples the true value in the population has a 90% chance of falling within the confidence intervals constructed in this way from the sample. Confidence intervals are discussed in the body of this report on page 10.

<sup>8</sup> The 90% confidence interval for this measure ranges from 1,228 to 1,794.

<sup>9</sup> Based on my results the 90% confidence interval ranges from 26.78 to 26.81.

<sup>10</sup> The 90% confidence interval for the Critical Support population ranges from 26,057 to 26,084.

<sup>11</sup> Based on conversations with Mr. Mandia, it is my understanding that the total number of copies of Objects here includes Objects delivered to customers as well as any additional copies of those Objects found or recorded on SAP TN's systems for that Fix. Objects included in compressed files are counted as copies.

<sup>12</sup> Based on conversations with Mr. Mandia, it is my understanding that Fixes were sometimes grouped together for delivery in "bundles." If two Fixes that affected the same Object were delivered to the same customer in the same bundle, only one Object would be delivered to that customer; however, the customer would have received two Associated Files, one for each Fix.

Fix.<sup>13,14</sup> Based on this sample the total number of copies of Associated Files for Retrofit Fixes is estimated to be 50,247.<sup>15</sup> Similarly, I find that on average, there are 655.10 copies of Objects associated with First or Identified Deliverables of a Critical Support Fix.<sup>16</sup> For the Critical Support population, I estimate that there were 637,412 copies of Associated Files for Critical Support Fixes.<sup>17</sup>

For each Fix in the Retrofit and Critical Support samples, Mr. Mandia also counted the number of Environments used in the development or testing of that Fix. Using the sample of Retrofit Fixes, I find that on average, SAP TN used 2.63 Environments in the development or testing of a Retrofit Fix.<sup>18</sup> Based on this sample the total number of Environments used by SAP TN for Retrofit Fixes is estimated to be 558.<sup>19</sup> Similarly, I find that on average, SAP TN used 6.35 Environments to develop or test Critical Support Fixes.<sup>20</sup> For the Critical Support population, I estimate that 6,177 Environments were used by SAP TN in the development and testing of Critical Support Fixes.<sup>21</sup>

---

<sup>13</sup> Based on conversations with Mr. Mandia, it is my understanding that an Associated File is a COBOL, SQR, SQC, DAT, or DMS file that is associated with the development, testing or delivery of a Fix.

<sup>14</sup> Based on my results the 90% confidence interval ranges from 232.62 to 241.41.

<sup>15</sup> The 90% confidence interval ranges from 49,315 to 51,180 for the population of Retrofit Fixes.

<sup>16</sup> The 90% confidence interval ranges from 655.06 to 655.14.

<sup>17</sup> The 90% confidence interval for the total for Critical Support Fixes ranges from 637,370 to 637,454.

<sup>18</sup> Based on my results the 90% confidence interval ranges from 2.30 to 2.96.

<sup>19</sup> The 90% confidence interval ranges from 488 to 628 in the population of Retrofit Fixes.

<sup>20</sup> The 90% confidence interval ranges from 5.92 to 6.77.

<sup>21</sup> The 90% confidence interval ranges from 5,763 to 6,592.

Mr. Mandia also gathered information to understand Contamination<sup>22</sup> of Objects and Environments. One measure to capture this information, which was counted for the entire population of Fixes for Retrofit and for Critical Support, is the percentage of instances in which customers received a First Deliverable that was contaminated based on Mr. Mandia's analysis of Objects. For Retrofit Fixes, 89.75% of the time that customers received a First Deliverable, that deliverable was contaminated based on Mr. Mandia's analysis of Objects found in Delivered Updates and Fixes. This same measure for Critical Support Fixes shows that 93.72% of the time, customers received a First Deliverable that was contaminated based on Mr. Mandia's analysis of Objects found in Delivered Updates and Fixes.

A second measure of Contamination is provided by the percentage of instances in which customers received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on either Object analysis or on analysis of SAP TN's development and testing documentation. Based on my analysis of the Retrofit sample, 83.92% of the instances in which customers received First or Identified Deliverables were contaminated, based on Mr. Mandia's Object and documentation analysis.<sup>23</sup> This same measure for Critical Support shows that 99.12% of the instances in which customers received First or Identified Deliverables were contaminated, based on Mr. Mandia's Object and documentation analysis.<sup>24</sup>

Another measure of Contamination is provided by the percentage of instances in which customers received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on Object and documentation analysis or based on the fact that the

---

<sup>22</sup> Based on conversations with Mr. Mandia, it is my understanding that a Fix is Contaminated if Cross-Use of any software occurred at any point in the development, testing, or production of any Object for that Fix. Cross-Use means a use of an Environment licensed to one customer to provide support to another customer.

<sup>23</sup> The 90% confidence interval ranges from 71.56% to 96.29% in the population of Retrofit Fixes.

<sup>24</sup> The 90% confidence interval ranges from 98.65% to 99.52% in the population of Critical Support Fixes.

customer's Environment was used to support other customers. My analysis of the Retrofit sample shows that 87.19% of the instances in which customers received First or Identified Deliverables were contaminated, based on Mr. Mandia's Object and documentation analysis and his analysis of the customer's Environment.<sup>25</sup> This same measure for Critical Support shows that 99.19% of the instances in which customers received First or Identified Deliverables were contaminated, based on Mr. Mandia's Object and documentation analysis and his analysis of the customer's Environment.<sup>26</sup>

A fourth measure of Contamination is captured by the percentage of hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable that were contaminated based on Object analysis. The components of this measure were counted for the full population of Retrofit and Critical Support Fixes. My analysis of the Retrofit population shows that 87.66% of hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable were contaminated based on Object analysis. This same measure for Critical Support shows 67.96% of hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable were contaminated based on Object analysis.

A final measure of Contamination is captured by the percentage of hashes for the set of DAT Associated Files in any First Deliverable that were contaminated based on Object analysis. The components of this measure were counted for the full population of Retrofit and Critical Support Fixes. My analysis of the Retrofit population shows that 89.29% of hashes for the set of DAT Associated Files in any First Deliverable were contaminated based on Object analysis. This same measure for Critical Support shows 82.59% of hashes for the set of DAT Associated Files in any First Deliverable were contaminated based on Object analysis

---

<sup>25</sup> The 90% confidence interval ranges from 74.64% to 99.74% in the population of Retrofit Fixes.

<sup>26</sup> The 90% confidence interval in the Critical Support Fix population ranges from 98.73% to 99.57%.

## 2) Introduction and Assignment

I have been retained by counsel for the Plaintiffs in the matter of Oracle USA, Inc., *et al.* v. SAP AG, *et al.* (Case No. 07-CV-01658 PJH(EDL)) to design a statistically valid sample of SAP TN's PeopleSoft HRMS payroll tax and regulatory Updates that can be used to scientifically estimate the number of Fixes delivered to customers by SAP TN that infringe Oracle copyrights or otherwise resulted from impermissible cross-use of Oracle's software. I understand that Plaintiffs will use the sample to estimate the percentage of instances in which the Fixes delivered to SAP TN's customers were contaminated, in the sense that they were handled or produced in a way that resulted from copyright infringement or breached other laws. My role in this engagement is to generate a random sample of Fixes to be reviewed and to calculate population and sample statistics for a number of measures, including measures of Contamination, based on information and data gathered by Plaintiffs' computer forensics expert, Mr. Mandia.

This report is organized as follows: In Section 3, I discuss my qualifications. In Section 4, I list the information considered in this case. In Section 5, I provide a brief background of the matter. In Section 6, I briefly describe the questions of interest. In Section 7, I discuss why sampling is statistically valid in this case. I further describe the sampling protocol used for this analysis. In Section 8, I discuss the extrapolation methods used for my analysis. In Section 9, I present my results. In Section 10, I conclude my report.

I reserve the right to update, supplement, and amend this report as additional information becomes available.

## 3) Qualifications

I am the National Managing Director and a founder of Advanced Analytical Consulting Group, Inc. ("AACG"). I have a Ph.D. in Economics from The University of Chicago. I have testified in a range of matters over a number of years. My curriculum vitae is attached in *Appendix 3*.

My billing rate for this case is \$627 per hour. The rates of my staff assigned to this project range from \$250 to \$507. Compensation for AACG is not contingent on the outcome of the proceedings.

#### 4) **Information Considered**

My opinions are based upon the review of the documents produced, data provided to me in this matter, academic references, and my education and experience in research and consulting. See *Appendix 4*.

#### 5) **Background**

Oracle is a leading global database and applications software development company. Oracle provides its licensed customers with robust customer support services that consist of telephone and email customer service access, Fixes and Updates,<sup>27</sup> and articles that can help customers address software issues. Oracle specifically provided tax and regulatory Updates for PeopleSoft HRMS to its customers on a regular basis, so that customers using PeopleSoft HRMS software could remain in compliance with all relevant laws and regulations. One of the ways in which licensed customers could access these support materials during the relevant time frame was by logging in to Oracle's password-protected support website, Customer Connection.

In December 2004, SAP TN was a small company in Bryan, Texas (formed by ex-PeopleSoft employees) that offered low-cost maintenance for PeopleSoft and JD Edwards software to its customers. SAP TN continued to offer this support after being purchased by SAP in January 2005. SAP TN regularly distributed Fixes, including regular deliveries of tax and regulatory Updates. SAP TN created its Fixes, including its tax and regulatory Updates, using copies of Oracle software.<sup>28</sup>

#### 6) **Questions of Interest**

Oracle is interested in measuring the activity that SAP TN was engaged in to develop, test, and distribute these Fixes. There are a number of ways to measure SAP TN's activity. The

---

<sup>27</sup> From my conversations with Mr. Mandia, I understand that an Update is a group of Fixes delivered together in a single deliverable, either by Oracle or SAP TN.

<sup>28</sup> I have been instructed by Oracle's counsel to assume that, internally, SAP TN distinguished between its Retrofit Support process and its Critical Support process. I have also been asked to assume that both processes involved the generation and delivery to customers of tax and regulatory Updates relevant to various releases and versions of PeopleSoft HRMS payroll software.

information for many of these measures can be gathered for the entire population of Retrofit and Critical Support Fixes to demonstrate the extent of SAP TN's activity. For example, the number of COBOL, SQR, and SQC files used in the delivery of Fixes can be calculated for all Fixes by a computer forensics expert. However, because of the lack of clear documentation, the lack of systematic record-keeping by SAP TN, the technical difficulty in gathering the information required for many of these measures, and therefore the enormous increase in the amount of time and effort that would be required to gather data for many of the other measures of interest, Mr. Mandia has collected information for some aspects of Retrofit Fixes and Critical Support Fixes based on a scientific, random sample of the Retrofit and Critical Support population of Fixes.<sup>29</sup> These data are used to estimate the value of measures of interest in the population of Retrofit Fixes and the population of Critical Support Fixes.

The statistical techniques and calculations that I use in this report to generate results have been chosen because they are well-tested, generally known and accepted, and well-documented in standard statistical textbooks. I have selected specific statistical methods for each type of measure investigated by Mr. Mandia to provide accurate estimates of measures that are of interest in detecting Contamination. The selection of statistical methods is not driven by any legal facts or conclusions. These are standard, well-known formulas that are used in sampling situations.

## 7) **Sampling**

### **a. General Description**

Sampling is used in many different scientific disciplines – biology, chemistry, economics and sociology to name a few – to make statements about a measure of interest for a population, when it may be too expensive, difficult, or time-consuming to collect information about that measure of interest for the entire population. Sampling offers some advantages, such as the ability to gather information at a lower cost and with greater speed. In many instances,

---

<sup>29</sup> In conversations with me, Mr. Mandia thought that it would require thousands of hours of time by highly-trained computer forensic staff to capture data for some groups of measures across the entire population of Fixes.

including those in which highly technical or trained professionals may be required to evaluate characteristics of the observations, sampling also permits information to be obtained about a population when review of the entire population is not feasible. While the estimates produced by sampling are measured with some variation, the extent of that variation can be measured scientifically to report the statistical precision of the estimate. Furthermore, one of the powerful advantages of scientific random samples is that as the size of the sample is increased, the accuracy of estimates based on those samples also tends to increase. So, scientific random samples can be designed to measure a characteristic of interest with the level of precision required for the specific purpose at hand. For example, if we wanted to know how many households in a town of 2000 households had a dog, we would need to sample about 40 households to estimate that figure within a range of plus or minus 15 percentage points.<sup>30</sup> But if we wanted to know that same figure plus or minus 2.5 percentage points, we would need to sample approximately 800 households. The level of effort put into the sampling depends on how precisely the characteristic of interest needs to be measured for the purposes at hand.

The required level of precision is an input given to the statistician. Given the required level of precision for the question at hand, the statistician can develop a sampling process and sample size that is likely to provide the precision required. Typically, a sample and associated statistical tests allow the statistician to make a statement of the following form: The sample of data show that we are 95% sure that the population average lies within a certain range, which statisticians define as the 95% confidence interval.<sup>31</sup> The number of observations sampled and the underlying variability of the data will determine the degree of precision that can be achieved.

---

<sup>30</sup> This example is based on a confidence interval somewhere between 90% and 95%.

<sup>31</sup> More precisely, Cochran defines a confidence interval as follows: "The '99% confidence' figure implies that if the same sampling plan were used many times in a population, a confidence statement being made from each sample, about 99% of these statements would be correct and 1% wrong." William G. Cochran, Sampling Techniques, Third Edition (New York: Wiley & Sons, 1977), 12.

Often prior to taking a sample, an analysis will be performed to estimate the number of sample observations likely to be needed to achieve the desired precision of the estimates. I list three elements to this process and discuss them further below. First, the desired confidence level at which results are to be reported needs to be established. The higher the desired confidence level, the larger the required sample size will be. Second, the desired precision with which the results are to be reported is chosen. Finally, some rough estimate of the standard deviation of the measure of interest in the population is often used to determine how varied the population being sampled might be. This is done based on either some prior information about the characteristics of the population or on an initial review of the data or even a small sample of the data, often called a probe sample.

Once the sample size has been determined, a scientific random sample is drawn from the population. If the sample is chosen so that every member of the population has a known (non-zero) probability of being selected in the sample, it is possible to draw conclusions about the broader population based on the sample. A classic sampling method begins with the enumeration of the entire population. The sampled items can then be drawn appropriately from this population using a random number generator, which selects a number from across the range of the enumerated population one after another until the desired sample size is collected. This type of classic random sampling is known as sampling with replacement, as an item selected in any round is replaced into the population so that all items in the population have an equal probability of being selected into the sample in each draw.

For purposes of illustration, suppose that a researcher is interested in determining the total number of computers in office buildings in two city blocks, but it is very time-consuming for the researcher to walk around each floor of each building to count the number of computers on the desks. Further suppose that she has some rough knowledge from a public source that on average there are about 40 computers on each floor of a high-rise office building, but some have more and some have fewer. She estimates that on average the actual number on each floor will vary from the 40 by about 18. She knows that there are a total of 600 floors in all of the buildings in these two city blocks. She assigns each floor in each building a number so that the floors are numbered from 1 to 600.

In order to design the sample size, she decides to use a 90% confidence level and a 20% precision range; that is, she wants to be able to say that if she sampled from this population repeatedly, 90% of the time the true number of computers would be within plus or minus 10% of her result. Based on the information she has about the average number of computers per floor and how this average might vary across floors, and her goals of a 90% confidence level and 20% precision range, she uses standard statistical formulas to determine that she is likely to need to sample 50 floors. The final estimate of how many computers actually exist on those floors will be calculated once the sample is taken, but for now the researcher is just trying to determine approximately how many floors need to be sampled to get a precise enough estimate of the average number of computers per floor.

To determine which 50 floors to sample from the population total of 600, she would need to randomly pick 50 numbers from 1 to 600, then go to those floors and count the total number of computers on each floor. Table 1 is an example of the data she would have once she has counted the number of computers on each of the 50 randomly selected floors:

**TABLE 1: EXAMPLE**

Floor #	Number of Computers								
461	30	34	13	90	10	159	39	285	11
325	50	143	25	223	45	315	57	32	65
314	27	102	47	250	62	77	44	397	27
100	49	76	53	374	33	22	32	76	53
189	45	48	33	369	29	46	15	205	49
351	24	574	31	272	42	136	7	599	13
342	56	9	42	260	90	205	49	401	56
441	20	540	34	532	23	512	32	141	43
200	45	144	45	198	15	508	22	511	77
150	57	186	44	194	77	56	11	397	27

Using these data, she can calculate the average number of computers per floor in her sample, which turns out to be 38.5; a bit lower than the 40 she used to determine the sample size, but that difference is of no concern. Her result of 38.5 is the estimate obtained from the sample and is the best estimate of the average number of computers per floor based on the sample. Further, she can extrapolate to her population of 600 floors and estimate that there are 23,100 computers in the buildings in these two blocks

If her random number generator would have produced a different set of 50 floors, her sample average and the extrapolated total could have been different. The likely difference would depend on the sample size she has chosen and the variation in the number of computers on each floor. Using the data she has, she can estimate that likely variation, with a calculation known as the standard error, which measures the average variation in her sample. If the standard error in the sample is large, her estimate of the number of computers on each floor will be less precise than if the standard error in her sample were smaller.<sup>32</sup> Once that standard error of the sample is determined, it can be used to calculate and report an upper and lower bound for the estimated average number of computers on each floor and the total number of computers on all 600 floors of the office buildings in the population. At the 90% confidence level, the upper bound for this total is 25,711 and the lower bound is 20,489. This is equivalent to saying that if she were to take repeated scientific random samples of 50 floors, there is a 90% chance that the true value of the total number of computers on those 600 floors is between the confidence bounds calculated for each of those samples, in this sample 20,489 and 25,711. If the researcher required an estimate of the number of computers in the buildings plus or minus 1000 or even 200, she could achieve that by increasing the sample size.

**b. Protocol**

The issues in this matter provide a compelling reason to use sampling because the kind of information necessary for determining the extent of infringement by SAP TN is extremely difficult to gather for some measures of interest, in terms of both time and cost of collecting the data.<sup>33</sup> The required degree of precision for estimating these measures can be achieved through a sample. The unit of measure is a Fix, as defined by Oracle's computer forensics expert, Mr. Mandia. Internally, SAP TN distinguished between its Retrofit Support process

---

<sup>32</sup> The statistician has no influence over the standard error found in a sample of a given size. The standard error of the sample is simply a characteristic of the sample she has drawn that reflects the variability in the sample of a given size.

<sup>33</sup> In conversations with me, Mr. Mandia thought that it would require thousands of hours of time by highly trained computer forensic staff to capture data for some groups of measures across the entire population of Fixes.

and its Critical Support process. Therefore, these two types of Fixes are treated as two distinct populations. Mr. Mandia generated a complete listing of Retrofit Fixes and a complete listing of Critical Support Fixes.

There are many measures for which I have been asked to count or estimate the value for the broader population of Fixes. The measures of interest and their measure number are listed in Table 2, below, which I received from Mr. Mandia as ORCLX-MAN-000205. In addition to detailing the description of the measure, Table 2 indicates whether a measure is counted for the total population or sampled. It also lists the source from which the data were collected.<sup>34</sup> The information for some of the measures was stored by SAP TN in such a way that it could be readily captured electronically for the entire population of Fixes in both the Retrofit and the Critical Support populations. These are the measures that are indicated as having a full "Population Count." The information for the remaining measures required more resources per collected observation because it was not easily captured electronically, required time-consuming investigation per individual Fix, or was extremely costly to gather. These measures are described as having been "Sampled." Some of the measures listed below were calculated by Mr. Mandia as the sum of other measures in the table or as the union of two or

---

<sup>34</sup> The information in Table 2 was provided by Mr. Mandia. Based on conversations with Mr. Mandia, it is my understanding that:

- The Data Warehouse referenced in Table 2 consists of approximately fifty computer systems inside the SAP TN infrastructure; these systems contain file servers, libraries of software and support materials downloaded by SAP TN, and records of when files were created, last edited, or last accessed.
- BakTrak was used by SAP TN to track the creation and restoration of Environment backups.
- Consultant Docs and Templates is a directory on server TN-FS01.
- Environment Backups refers to compressed files containing copies of all or part of a PeopleSoft application.
- Analysis refers to Mr. Mandia's analysis of the Fix, Objects, documentation, or Environments, as referenced in the measure definition.

more measures in the table. In addition, I calculate some measures as the ratio of two measures that I received in the data. I will discuss all of these in further detail below.

The sample for each population was selected using sampling with replacement as follows:

1. There were 223 Fixes in the Retrofit population and 1,386 Fixes in the Critical Support population. Within each population, each Fix was assigned an item number sequentially.
2. Using a random number generator, I selected numbers ranging from 1 to 223 for Retrofit and from 1 to 1,386 for Critical Support, and I generated the random sample for each population. As mentioned above, this method is known as sampling with replacement so that each item in the population has an equal probability of being selected in each draw of the random number. Sampling with replacement allows for use of classic statistical formulas to estimate precision of the findings from the sample. This well-known technique is discussed extensively in statistical texts.<sup>35</sup>
3. The sampling numbers were then given to Mr. Mandia to collect the relevant data for each Fix in the sample.<sup>36</sup>
4. The sample size was determined based on examination of the data available for two measures of interest for which Mr. Mandia was able to collect data for the entire population of Fixes. Mr. Mandia provided this data to me as ORCLX-MAN-000060, the Excel workbook containing the results of his findings for the two measures of interest across the entire population of Fixes in the Retrofit and Critical Support

---

<sup>35</sup> William G. Cochran, Sampling Techniques, Third Edition (New York: Wiley & Sons, 1977) 18, 29-30.

<sup>36</sup> The list of Fixes for each of the populations included Fixes with a status of “Cancelled,” “Research Only,” or “0.” However, these three types of status were determined to be uninformative for the purposes of the measures of interest. Therefore, if the random number generator selected one of these Fixes, it was replaced (in the sampling order) with another Fix. Accordingly, the population to which the sample results are extrapolated does not include Fixes with a status of “Cancelled,” “Research Only,” or “0.” The distribution of the status of Fixes for each population and sample is shown in *Appendix 2*. See Figure 1.

populations. Additionally, there was discussion of the characteristics of one of the measures for which it was extremely costly to gather data; this measure was *the number of Environments used in the development or testing of the Fix, as identified in the development, test or other documentation*. I was asked by counsel to calculate samples sizes based on a 90% confidence level and 50% precision range for this measure, which yielded a sample size of 46 for Retrofit and 238 for Critical Support Fixes.<sup>37</sup>

5. For the samples, Mr. Mandia and his colleagues reviewed information related to the Fixes in the order that I provided to them from the random number generator. The results of his finding for each individual Fix in the two samples were entered into an Excel workbook for each of the 44 measures Mr. Mandia reviewed for the HRMS Fix analysis. For some measures, Mr. Mandia reviewed the entire population of Fixes in the Retrofit and Critical Support populations. Mr. Mandia provided me with ORCLX-MAN-000061, the Excel workbook containing the results of his findings for the samples and the populations of Fixes. I use the reported information in that Excel workbook to generate my results and conclusions.
6. I provided my results to Mr. Mandia as ORCLX-AACG-000001, a set of tables for the population of Critical Support Fixes, and ORCLX-AACG-000002, a set of tables for the population of Retrofit Fixes.

---

<sup>37</sup> Table 2 lists all the measures of interest. Measure 116, the number of Environments used in the development or testing of the Fix, as identified in development, test, and other documentation, was the basis for determining the sample size. The assumed averages and standard deviations used to calculate the sample size were based on simulated data for measure 116, where it was assumed that measure 116 should be zero whenever measure 104 is zero and that the distribution of the non-zero values measure 116 would be similar to that of measure 115.

**TABLE 2: LIST OF MEASURES**

Measure Number	Description or Reportable Information for Each Fix	Population Count or Sampled	Data Source
101	The versions to which the fix applied, according to the Application Release field in SAS	Population Count	SAS
102	The versions to which the fix applied, based on the versions supported by SAP TN for customers	Sampled	SAS, Delivered Updates and Fixes
103	The reported status of the fix, according to the Status field in SAS	Population Count	SAS
104	The unique hashes for the set of COBOL, SQR or SQC Associated Files in any First Deliverable	Population Count	Delivered Updates and Fixes
105	The unique hashes for the set of DAT or DMS Associated Files in any First Deliverable (142*143)	Population Count	Delivered Updates and Fixes
106	The unique hashes for the set of COBOL, SQR, SQC, DAT or DMS Associated Files in any First Deliverable (104*105)	Population Count	Delivered Updates and Fixes
107	The unique hashes for the set of COBOL, SQR, SQC, DAT or DMS Associated Files in any First Deliverable that were not in any Subsequent Deliverable	Population Count	Delivered Updates and Fixes
108	The COBOL, SQR, SQC, DAT or DMS Associated Files in each customer's First Deliverable	Population Count	Delivered Updates and Fixes
109	The COBOL, SQR, SQC, DAT or DMS Associated Files in duplicate copies of each customer's First Deliverable found on Delivered Updates and Fixes	Population Count	Delivered Updates and Fixes
110	Duplicate copies of the COBOL, SQR, SQC, DAT or DMS Associated Files in a First Deliverable that are never Associated Files for a Subsequent Delivery found in the Data Warehouse (other than in environment backups)	Population Count	Data Warehouse
111	The COBOL, SQR, SQC, DAT or DMS Associated Files in duplicate copies of each customer's First Deliverable, found in the Data Warehouse (other than in environment backups)	Population Count	Data Warehouse
112	Duplicate copies of the COBOL, SQR, SQC, DAT or DMS Associated Files in a First Deliverable that are never Associated Files for a Subsequent Delivery, found in environment backups	Population Count	Environment Backups
113	Duplicate copies of the COBOL, SQR and SQC Associated Files recorded by a BakTrak entry where the associated backup is not present in the Data Warehouse	Population Count	BakTrak
114	The COBOL, SQR, SQC, DAT or DMS Associated Files attached to the SAS record	Sampled	SAS
115	Environments used in development or testing of the fix, as identified in DAT Associated Files in a First Deliverable	Population Count	Delivered Updates and Fixes
116	Environments used in development or testing of the fix, as identified in development, test and other documentation	Sampled	SAS, Consultant Docs and Templates
117	Environments used in development or testing of the fix (Union 115,116)	Sampled	SAS, Delivered Updates and Fixes, Consultant Docs and Templates
118	The customers receiving a First Deliverable	Population Count	Delivered Updates and Fixes
119	Additional customers receiving an Identified Deliverable	Sampled	SAS
120	All customers receiving a First or Identified Deliverable (118*119)	Sampled	Delivered Updates and Fixes, SAS
121	All First Deliverables containing documentation referencing the fix ID	Population Count	Delivered Updates and Fixes
122	The customers that received a First Deliverable contaminated through receipt of a DAT file not specifically for that customer	Population Count	Analysis
123	The customers that received a First or Identified Deliverable contaminated by use of a generic environment, of some other customer's environment, or of an environment built from another customer's software	Sampled	Analysis
124	The customers that received a First or Identified Deliverable contaminated by development using the source group model	Sampled	Analysis
125	The customers that received a First Deliverable contaminated because another customer received at least one object with an identical hash value	Population Count	Analysis
126	The customers that received a First or Identified Deliverable other than from an environment built solely from that customer's software and used solely for that customer (with respect to the analyzed fix)	Sampled	Analysis
127	The customers that received a contaminated First Deliverable based on analysis of delivered objects (Union 122,125)	Population Count	Analysis
128	The customers that received a contaminated First or Identified Deliverable based on analysis of objects and of development and testing documentation (Union 122,123,124,125)	Sampled	Analysis
129	The customers that received a contaminated fix based on object and documentation analysis, when cross-use of a customer-specific environment renders activities on behalf of the customer infringing (Union 122, 123, 124, 125, 126)	Sampled	Analysis

**TABLE 2: LIST OF MEASURES (Continued)**

Measure Number	Description or Reportable Information for Each Fix	Population Count or Sampled	Data Source
130	Percentage of customers that received a First Deliverable where that First Deliverable was contaminated based on analysis of objects found in Delivered Updates and Fixes (127/118)	Population Count	Analysis
131	Percentage of customers that received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on object and documentation analysis (128/120)	Sampled	Analysis
132	Percentage of customers that received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on object or documentation analysis or based on use of the customer's environment to support other customers (129/120)	Sampled	Analysis
133	Total copies of objects comprising First Deliverables (including copies located in compressed files) (108+109+110+111+112+113)	Population Count	Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak
134	Total copies of objects comprising First or Identified Deliverables (including copies located in compressed files) (133+114)	Sampled	Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak, SAS
135	The unique hashes for the set of COBOL, SQR, SQC, or DAT Associated Files in any First Deliverable contaminated because they were delivered to more than one customer	Population Count	Delivered Updates and Fixes -- Analysis
136	The unique hashes for the set of DAT Associated Files in any First Deliverable contaminated because they were created using a generic environment or created with one customer's environment but delivered to another customer	Population Count	Delivered Updates and Fixes -- Analysis
137	The contaminated, unique hashes for the set of COBOL, SQR, SQC, or DAT Associated Files in any First Deliverable (Union 135,136)	Population Count	Delivered Updates and Fixes -- Analysis
138	Percentage of hashes for the set of COBOL, SQR, SQC, or DAT Associated Files in any First Deliverable that are contaminated based on object analysis (137/(104+142))	Population Count	Delivered Updates and Fixes -- Analysis
139	The unique hashes for the set of COBOL, SQR or SQC Associated Files in an Identified Deliverable	Sampled	SAS
140	The unique hashes for the set of DAT or DMS Associated Files in an Identified Deliverable	Sampled	SAS
141	The unique hashes for the set of COBOL, SQR, SQC, DAT or DMS Associated Files in an Identified Deliverable (139+140)	Sampled	SAS
142	Total number of hashes for DAT Associated Files that were modified for or included in First Deliverables	Population Count	Delivered Updates and Fixes
143	Total number of hashes for DMS Associated Files that were modified for or included in First Deliverables	Population Count	Delivered Updates and Fixes
144	Percentage of hashes for the set of DAT Associated Files in any First Deliverable contaminated because they were created using a generic environment or created with one customer's environment but delivered to another customer (136/142)	Population Count	Delivered Updates and Fixes

## 8) Measurement of Results

### a. Population Count Measures

Table 3 below lists the measures that were counted for the entire population of Fixes, in each of the Retrofit and Critical Support populations.<sup>38</sup> For these measures, I calculate population statistics that are presented in the next section.

<sup>38</sup> Measure 103 is the status of the Fix. I do not calculate any statistical results for this measure and therefore it is not listed in Table 3. However, the distribution of the status of Fixes for each population and sample is shown in *Appendix 2*. See Figure 1.

**TABLE 3: LIST OF MEASURES COUNTED FOR ENTIRE POPULATION**

Measure	
101	Total number of fixes for distinct versions, according to the Application Release field in SAS (SAS)
104	Total number of hashes for COBOL, SQR, or SQC Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)
105	Total number of hashes for DAT or DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)
106	(142+143)
107	Total number of hashes for COBOL, SQR, SQC, DAT, or DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)
108	(104+105)
109	Total number of hashes for COBOL, SQR, SQC, DAT or DMS Associated Files in any First Deliverable that were not in any Subsequent Deliverable (Delivered Updates and Fixes)
110	Total number of COBOL, SQR, SQC, DAT or DMS Associated Files found in First Deliverables for customers (Delivered Updates and Fixes)
111	Total number of COBOL, SQR, SQC, DAT or DMS Associated Files in duplicate copies of each customer's First Deliverable found on Delivered Updates and Fixes (Delivered Updates and Fixes)
112	Total number of duplicate copies of the COBOL, SQR, SQC, DAT or DMS Associated Files, enumerated in measure 107, found in the Data Warehouse (other than in environment backups) (Data Warehouse)
113	Total number of COBOL, SQR, SQC, DAT or DMS Associated Files in duplicate copies of each customer's First Deliverable, found in the Data Warehouse (other than in environment backups) (Data Warehouse)
115	Total number of duplicate copies of the COBOL, SQR, SQC, DAT or DMS Associated Files, enumerated in measure 107, found in environment backups (Environment Backups)
118	Total number of duplicate copies of the COBOL, SQR and SQC Associated Files recorded by a Baktrak entry where the associated backup is not present in the Data Warehouse (BakTrak)
121	Total number of environments used in development or testing of fixes, as identified in DAT Associated Files in a First Deliverable (Delivered Updates and Fixes)
122	Total number of instances in which customers received a First Deliverable (Delivered Updates and Fixes)
125	Total number of First Deliverables containing documentation referencing fix IDs (Delivered Updates and Fixes)
127	Total number of instances in which customers received a First Deliverable contaminated through receipt of a DAT file not specifically for those customers (Analysis)
130	Total number of instances in which customers received a First Deliverable contaminated because another customer received at least one object with an identical hash value (Analysis)
133	Total number of instances in which customers received a contaminated First Deliverable based on analysis of delivered objects (Analysis)
135	(Union 122,125)
136	Percentage of instances in which customers received a First Deliverable where that First Deliverable was contaminated based on analysis of objects found in Delivered Updates and Fixes* (Analysis)
137	(127/118)
138	Total number of copies of objects comprising First Deliverables (including copies located in compressed files) (Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak)
142	(108+109+110+111+112+113)
143	Total number of unique hashes for the set of COBOL, SQR, SQC, or DAT Associated Files in any First Deliverable contaminated because they were delivered to more than one customer (Delivered Updates and Fixes--Analysis)
144	Total number of unique hashes for the set of DAT Associated Files in any First Deliverable contaminated because they were created using a generic environment or created with one customer's environment but delivered to another customer (Delivered Updates and Fixes--Analysis)
	Total number of contaminated, unique hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable (Delivered Updates and Fixes--Analysis)
	(Union 135,136)
	Percentage of hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable that are contaminated based on object analysis* (Delivered Updates and Fixes--Analysis)
	(137/(104+142))
	Total number of hashes for DAT Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)
	Total number of hashes for DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)
	Percentage of hashes for the set of DAT Associated Files in any First Deliverable contaminated because they were created using a generic environment or created with one customer's environment but delivered to another customer* (Delivered Updates and Fixes--Analysis)
	(136/142)

\* The result for this measure is a weighted average of the component measures: Measure 130=Measure 127/Measure 118 Measure 138=Measure 137/(Measure 104+Measure142); Measure 144=Measure 136/Measure 142.

## **b. Extrapolation Methodology for Sample Measures**

In this section, I describe the techniques used to estimate the values for measures that were reported only for Fixes that were sampled from the populations of Retrofit and Critical Support Fixes.

For the measures that are sampled, I use standard statistical formulas to calculate the estimated frequency of occurrence in the population and the precision with which that estimated occurrence is calculated. I use the information provided by Mr. Mandia in the sampled data to estimate population totals and averages. The methods used for these estimations are standard statistical techniques that are found in commonly used statistics books and articles. These texts and articles discuss which methods to use for different types of data. For example, when estimating the total or average number of computers in a sample of floors in a set of office buildings, a mean (average) estimator may be appropriate. However, when estimating the proportion of movie patrons that liked a particular movie, a different estimator tailored to a “yes” or “no” answer would be used. Similarly, in this report, the specific statistical method used for each measure depends on the nature of the data for in each measure. I use the mean per unit estimator for most measures (102, 114, 116, 119, 123, 124, 126, 139, 140 and 141), and the ratio estimator for two of the measures (131 and 132). For the measures that are the union or sum of other measures (117, 120, 128, 129, and 134), I estimate the population total and its variance by taking into account the fact that some of the underlying measures have been counted for the whole population. I describe this method in detail below.

1. The step-by-step process I use to project the averages and totals from the observed samples to the total population for measures 102, 114, 116, 119, 123, 124, 126, 139, 140 and 141 are described below. The related formulas for the mean estimator I use for these measures are presented in *Appendix I*.
  - The first piece of information needed to calculate the population total is the average occurrence of the measure per observation in the sample (sample average). Each occurrence in the sample is added and then divided by the total number of observations in the sample.

- The occurrence of the measure in the total population is estimated by multiplying the sample average, calculated above, by the total number of observations in the population.
- Next, the standard error of the sample is calculated by taking the difference of each observation from the sample average, squaring it, adding up that squared difference across all observations in the sample, dividing the resulting number by the number of observations in the sample minus 1, and taking the square root of the result.
- The standard error of the sample mean is calculated by dividing the standard error of the sample by the square root of the number of observations in the sample.
- The standard error for the population total is the number of observations in the population multiplied by the standard error of the sample mean.
- The estimates of the population total, the population mean, and their respective standard errors are scientifically valid estimates of the true values in the population when the sample is generated randomly.<sup>39</sup>

Table 4 lists the measures for which I use the mean per unit estimator.

---

<sup>39</sup> Paul S. Levy and Stanley Lemeshow, Sampling of Populations, Fourth Edition (New York: Wiley & Sons, 2008), 55-58.

**TABLE 4: MEASURES ESTIMATED WITH MEAN PER UNIT ESTIMATOR**

Measure	
102	Total number of fixes for distinct versions, based on the versions supported by SAP TN for customers (SAS, Delivered Updates and Fixes)
114	Total number of COBOL, SQR, SQC, DAT or DMS Associated Files attached to SAS records (SAS)
116	Total number of environments used in development or testing of fixes, as identified in development, test and other documentation (SAS, Consultant Docs and Templates)
119	Total number of instances in which customers who did not receive a First Deliverable received an Identified Deliverable (SAS)
123	Total number of instances in which customers received a First or Identified Deliverable contaminated by use of a generic environment, of some other customer's environment, or of an environment built from another customer's software (Analysis)
124	Total number of instances in which customers received a First or Identified Deliverable contaminated by development using the source group model (Analysis)
126	Total number of instances in which customers received a First or Identified Deliverable other than from an environment built solely from their software and used solely for them (Analysis)
139	Total number of unique hashes for the set of COBOL, SQR or SQC Associated Files in Identified Deliverables (SAS)
140	Total number of unique hashes for the set of DAT or DMS Associated Files in Identified Deliverables (SAS)
141	Total number of unique hashes for the set of COBOL, SQR, SQC, DAT or DMS Associated Files in Identified Deliverables (SAS) (139+140)

2. Measures 117, 128, 129 are composed of the union of two or more measures, and therefore I call them composite measures. A union may be defined as a sum with overlapping duplicates removed. The total and variance for these measures are estimated by taking into account the fact that I have full information on the component measures taken from the analysis of the full population. I demonstrate by example for measure 117, which is the union of measures 115 and 116. Measure 115 is counted for the entire population, but measure 116 is counted only in the sample. In addition, some of the occurrences counted in 115 are also counted in 116. In this sense there is an overlap in the count of occurrences in these measures, which Mr. Mandia recorded. I calculate the amount of this overlap, designated as  $Overlap_{115, 116}$ , for each observation in the sample by subtracting the sum of measures 115 and 116 from measure 117. Similar calculations can be performed for measures 128 and 129. Table 5 lists these measures and I discuss the relevant relationships in *Appendix 1*.

**TABLE 5: UNION MEASURES**

Measure	
	Total number of environments used in development or testing of fixes (SAS, Delivered Updates and Fixes, Consultant Docs and Templates)
117	(Union 115,116)
	Total number of instances in which customers received a contaminated First or Identified Deliverable based on analysis of objects and of development and testing documentation
128	(Union 122,123,124,125)
	Total number of instances in which customers received a contaminated fix based on object and documentation analysis, when cross-use of a customer-specific environment renders activities on behalf of the customers infringing (Analysis)
129	(Union 122, 123, 124, 125, 126)

3. Measures 120 and 134 are each composed of two measures, and therefore I call them composite measures. Measure 120 is composed of component measures 118 and 119. Measure 134 is composed of component measures 114 and 133. For both measures 120 and 134, one component of the measure is recorded for the entire population and the other component is reported for the sample. For measures 120 and 134, since one of the components is measured for the full population, the only source of variance is the component that is measured only in the sample. Therefore, for measure 120, the only source of variance is measure 119, and for measure 134, the only source of variance is measure 114. Table 6 lists these measures and I present the related formulas in *Appendix 1*.

**TABLE 6: MEASURES THAT ARE SUMS OF OTHER MEASURES**

Measure	
	Total number of instances in which customers received a First Deliverable or Identified Deliverable (Delivered Updates and Fixes, SAS)
120	(118+119)
	Total number of copies of objects comprising First or Identified Deliverables (including copies located in compressed files) (Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak, SAS)
134	(133+114)

4. For two measures, measures 131 and 132, I use a ratio estimator to estimate their population means and population variances. The ratio estimator can be used to report

population totals, means or ratios, such as in measures 131 and 132.<sup>40</sup> Table 7 lists these two measures and I discuss the related formulas in *Appendix 1*.

**TABLE 7: MEASURES ESTIMATED WITH THE RATIO ESTIMATOR**

Measure	
131	Percentage of instances in which customers received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on object and documentation analysis (Analysis) (128/120)
132	Percentage of instances in which customers received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on object or documentation analysis or on the fact that the customer's environment was used to support other customers (Analysis) (129/120)

## 9) Results

In this section, I discuss the results of the population counts as well as the various extrapolation methods described in the section above. The tables below are presented in groups of measures that conceptually belong together. This grouping has been provided by Mr. Mandia.

Tables 8A and 8B show the results for measures related to numbers of versions affected by SAP TN's activities. Measure 101 is measured for the full population and therefore does not have any reported standard deviation or confidence interval. Measure 102 is measured only for the samples for each type of Fix. Table 8A illustrates that based on this sample the 90% confidence interval for the true value for the total number of Retrofit Fixes for distinct versions ranges from 330 to 463. This means that in repeated samples the true value in the population has a 90% chance of falling within the confidence intervals constructed in this way from the sample.

---

<sup>40</sup> The ratio estimator is often used for estimating ratios in the population. See William G. Cochran, *Sampling Techniques*, Third Edition (New York: Wiley & Sons, 1977) 30-31.

**TABLE 8A: VERSION MEASURES FOR RETROFIT**

Retrofit								
Sample Size: 46								
Population Size: 212								
Confidence Level: 90%								
Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
101 Total number of fixes for distinct versions, according to the Application Release field in SAS (SAS)		245				1.16		
102 Total number of fixes for distinct versions, based on the versions supported by SAP TN for customers (SAS, Delivered Updates and Fixes)	330	396	463	40	1.56	1.87	2.18	0.19

**TABLE 8B: VERSION MEASURES FOR CRITICAL SUPPORT**

Critical Support								
Sample Size: 238								
Population Size: 973								
Confidence Level: 90%								
Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
101 Total number of fixes for distinct versions, according to the Application Release field in SAS (SAS)		712				0.73		
102 Total number of fixes for distinct versions, based on the versions supported by SAP TN for customers (SAS, Delivered Updates and Fixes)	4,817	5,213	5,608	240	4.95	5.36	5.76	0.25

Tables 9A and 9B show the results for measures that record information based on analysis of hashes for COBOL, SQR, SQC, DAT or DMS files. Measures 104, 105, 106, 142, and 143 are measured for the full population and therefore their respective results do not need confidence intervals because they are reported with full certainty. The results for measures 139, 140, and 141, on the other hand, are reported based on extrapolation from the samples and therefore are presented with confidence intervals. Table 9A shows that there were 2,228 total numbers of hashes for COBOL, SQR, SQC, DAT or DMS Associated Files that were modified for, or included in, First Deliverables for the population of Retrofit Fixes. Table 9B shows that there were 10,245 total number of hashes for COBOL, SQR, SQC, DAT or DMS Associated Files that were modified for or included in First Deliverables for the population of Retrofit Fixes.

**TABLE 9A: HASH-RELATED MEASURES FOR RETROFIT**

Retrofit  
 Sample Size: 46  
 Population Size: 212  
 Confidence Level: 90%

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
104 Total number of hashes for COBOL, SQR, or SQC Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)		1,746				8.24		
105 Total number of hashes for DAT or DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes) (142+143)		482				2.27		
106 Total number of hashes for COBOL, SQR, SQC, DAT, or DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes) (104+105)		2,228				10.51		
139 Total number of unique hashes for the set of COBOL, SQR or SQC Associated Files in Identified Deliverables (SAS)	368	903	1,439	325	1.74	4.26	6.79	1.53
140 Total number of unique hashes for the set of DAT or DMS Associated Files in Identified Deliverables (SAS)	203	396	590	118	0.96	1.87	2.78	0.56
141 Total number of unique hashes for the set of COBOL, SQR, SQC, DAT or DMS Associated Files in Identified Deliverables (SAS) (139+140)	586	1,300	2,013	434	2.77	6.13	9.49	2.05
142 Total number of hashes for DAT Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)		224				1.06		
143 Total number of hashes for DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)		238				1.22		

**TABLE 9B: HASH-RELATED MEASURES FOR CRITICAL SUPPORT**

Critical Support  
 Sample Size: 238  
 Population Size: 973  
 Confidence Level: 90%

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
104 Total number of hashes for COBOL, SQR, or SQC Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)		3,904				4.01		
105 Total number of hashes for DAT or DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes) (142+143)		6,341				6.52		
106 Total number of hashes for COBOL, SQR, SQC, DAT, or DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes) (104+105)		10,245				10.53		
139 Total number of unique hashes for the set of COBOL, SQR or SQC Associated Files in Identified Deliverables (SAS)	11	37	63	16	0.01	0.04	0.06	0.02
140 Total number of unique hashes for the set of DAT or DMS Associated Files in Identified Deliverables* (SAS)	0	8	25		0.00	0.01	0.03	
141 Total number of unique hashes for the set of COBOL, SQR, SQC, DAT or DMS Associated Files in Identified Deliverables (SAS) (139+140)	16	45	74	18	0.02	0.05	0.08	0.02
142 Total number of hashes for DAT Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)		3,618				3.72		
143 Total number of hashes for DMS Associated Files that were modified for or included in First Deliverables (Delivered Updates and Fixes)		2,723				2.80		

\* The lower and upper bounds are calculated using 10,000 iterations of repeated sampling. The confidence interval is directly computed. Therefore the standard error is not computed. See B. Efron & R.J. Tibshirani, An Introduction to the Bootstrap, (Chapman & Hall, 1993), 158-176.

Tables 10A and 10B show the results for measures that record information for Objects impacted by SAP TN's activities. Many of these measures are counted for the full population and therefore their results do not need confidence intervals because they are reported with full certainty. Measure 134 is reported for the samples of Retrofit and Critical Support Fixes and, as described in the preceding section, is a composite measure. Additionally, the results presented in Tables 10A and 10B take into account the fact that measure 133 (component measure for measure 134) is counted for the entire population of Fixes.<sup>41</sup> Table 10A shows that the estimated total number of copies of Objects comprising First or Identified Deliverables (including copies located in compressed files) is 50,247 for Retrofit Fixes. The 90% confidence interval ranges from 49,315 to 51,180. This means that in repeated samples the true value in the population has a 90% chance of falling within the confidence intervals constructed in this way from the sample. Similarly, Table 10B shows that the estimated total number of copies of Objects comprising First or Identified Deliverables (including copies located in compressed files) is 637,412 for Critical Support Fixes. The 90% confidence interval ranges from 637,370 to 637,454. This means that in repeated samples, the true value in the population has a 90% chance of falling within the confidence intervals constructed in this way from the sample.

---

<sup>41</sup> In *Appendix 2*, I present the results for these measures based on extrapolating them from the sample. See Tables A.2 and A.3.

**TABLE 10A: OBJECT-RELATED MEASURES FOR RETROFIT**

Retrofit  
 Sample Size: 46  
 Population Size: 212  
 Confidence Level: 90%

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
107 Total number of hashes for COBOL, SQR, SQC, DAT or DMS Associated Files in any First Deliverable that were not in any Subsequent Deliverable (Delivered Updates and Fixes)		517				2.44		
108 Total number of COBOL, SQR, SQC, DAT or DMS Associated Files found in First Deliverables for customers (Delivered Updates and Fixes)		5,128				24.19		
109 Total number of COBOL, SQR, SQC, DAT or DMS Associated Files in duplicate copies of each customer's First Deliverable found on Delivered Updates and Fixes (Delivered Updates and Fixes)		12,628				59.57		
110 Total number of duplicate copies of the COBOL, SQR, SQC, DAT or DMS Associated Files, enumerated in measure 107, found in the Data Warehouse (other than in environment backups) (Data Warehouse)		9,137				43.10		
111 Total number of COBOL, SQR, SQC, DAT or DMS Associated Files in duplicate copies of each customer's First Deliverable, found in the Data Warehouse (other than in environment backups) (Data Warehouse)		5,644				26.62		
112 Total number of duplicate copies of the COBOL, SQR, SQC, DAT or DMS Associated Files, enumerated in measure 107, found in environment backups (Environment Backups)		16,941				75.67		
113 Total number of duplicate copies of the COBOL, SQR and SQC Associated Files recorded by a Baktrak entry where the associated backup is not present in the Data Warehouse (BakTrak)		38				0.18		
114 Total number of COBOL, SQR, SQC, DAT or DMS Associated Files attached to SAS records (SAS)	699	1,631	2,564	567	3.30	7.70	12.09	2.67
133 Total number of copies of objects comprising First Deliverables (including copies located in compressed files) (Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak) (108+109+110+111+112+113)		48,616				229.32		
134 Total number of copies of objects comprising First or identified Deliverables (including copies located in compressed files)* (Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak, SAS) (133+114)	49,315	50,247	51,180	567	232.62	237.02	241.41	2.67

\* This measure is reported by taking into account that one of its components is known in the full population

**TABLE 10B: OBJECT-RELATED MEASURES FOR CRITICAL SUPPORT**

Critical Support  
 Sample Size: 238  
 Population Size: 973  
 Confidence Level: 90%

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
107 Total number of hashes for COBOL, SQR, SQC, DAT or DMS Associated Files in any First Deliverable that were not in any Subsequent Deliverable (Delivered Updates and Fixes)		9,433				9.69		
108 Total number of COBOL, SQR, SQC, DAT or DMS Associated Files found in First Deliverables for customers (Delivered Updates and Fixes)		54,548				56.06		
109 Total number of COBOL, SQR, SQC, DAT or DMS Associated Files in duplicate copies of each customer's First Deliverable found on Delivered Updates and Fixes (Delivered Updates and Fixes)		69,986				71.93		
110 Total number of duplicate copies of the COBOL, SQR, SQC, DAT or DMS Associated Files, enumerated in measure 107, found in the Data Warehouse (other than in environment backups) (Data Warehouse)		149,655				153.81		
111 Total number of COBOL, SQR, SQC, DAT or DMS Associated Files in duplicate copies of each customer's First Deliverable, found in the Data Warehouse (other than in environment backups) (Data Warehouse)		71,785				73.78		
112 Total number of duplicate copies of the COBOL, SQR, SQC, DAT or DMS Associated Files, enumerated in measure 107, found in environment backups (Environment Backups)		274,349				281.96		
113 Total number of duplicate copies of the COBOL, SQR and SQC Associated Files recorded by a Baktrak entry where the associated backup is not present in the Data Warehouse (BakTrak)		17,024				17.50		
114 Total number of COBOL, SQR, SQC, DAT or DMS Associated Files attached to SAS records (SAS)	23	65	107	26	0.02	0.07	0.11	0.03
133 Total number of copies of objects comprising First Deliverables (including copies located in compressed files) (Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak)		637,347				655.03		
134 Total number of copies of objects comprising First or Identified Deliverables (including copies located in compressed files)* (Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak, SAS)	637,370	637,412	637,454	26	655.06	655.10	655.14	0.03

\* This measure is reported by taking into account that one of its components is known in the full population

Tables 11A and 11B present results for measures that record information for Environments affected by the Fixes distributed by SAP TN. Measure 115 is reported for the full population, and therefore its results are reported with full certainty and do not require confidence intervals. Measure 117 is reported for the samples of Retrofit and Critical Support Fixes and, as described in the preceding section, is a composite measure. The results

presented in Tables 11A and 11B take into account the fact that measure 115 (component measure for measure 117) is counted for the entire population of Fixes.<sup>42</sup>

**TABLE 11A: ENVIRONMENT-RELATED MEASURES FOR RETROFIT**

**Retrofit**  
**Sample Size: 46**  
**Population Size: 212**  
**Confidence Level: 90%**

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
115 Total number of environments used in development or testing of fixes, as identified in DAT Associated Files in a First Deliverable (Delivered Updates and Fixes)		97				0.46		
116 Total number of environments used in development or testing of fixes, as identified in development, test and other documentation (SAS, Consultant Docs and Templates)	432	512	591	48	2.04	2.41	2.79	0.23
117 Total number of environments used in development or testing of fixes* (SAS, Delivered Updates and Fixes, Consultant Docs and Templates) (Union 115,116)	488	558	628	43	2.30	2.63	2.96	0.20

\* This measure is reported by taking into account that one of its components is known in the full population.

**TABLE 11B: ENVIRONMENT-RELATED MEASURES FOR CRITICAL SUPPORT**

**Critical Support**  
**Sample Size: 238**  
**Population Size: 973**  
**Confidence Level: 90%**

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
115 Total number of environments used in development or testing of fixes, as identified in DAT Associated Files in a First Deliverable (Delivered Updates and Fixes)		2,412				2.48		
116 Total number of environments used in development or testing of fixes, as identified in development, test and other documentation (SAS, Consultant Docs and Templates)	3,447	3,867	4,288	255	3.54	3.97	4.41	0.26
117 Total number of environments used in development or testing of fixes* (SAS, Delivered Updates and Fixes, Consultant Docs and Templates) (Union 115,116)	5,763	6,177	6,592	252	5.92	6.35	6.77	0.26

\* This measure is reported by taking into account that one of its components is known in the full population.

Tables 12A and 12B present results for measures that record the number of instances in which customers received First or Identified Deliverables. Measures 118 and 121 are reported for the full population. Therefore, results for measures 118 and 121 are reported

<sup>42</sup> In Appendix 2, I present the results for this measure based on extrapolating them from the sample. See Tables A.2 and A.3.

with full certainty and do not require confidence intervals. Measures 119 and 120 are reported for the samples of Retrofit and Critical Support Fixes. Measure 119 is extrapolated using a mean per unit estimator as described in the previous section. Measure 120 is a composite measure because it is the sum of two measures. Additionally, one of its components (118) is measured for the full population.

**TABLE 12A: CUSTOMER-RELATED MEASURES FOR RETROFIT**

**Retrofit**  
**Sample Size: 46**  
**Population Size: 212**  
**Confidence Level: 90%**

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
118 Total number of instances in which customers received a First Deliverable (Delivered Updates and Fixes)		907				4.28		
119 Total number of instances in which customers who did not receive a First Deliverable received an Identified Deliverable (SAS)	321	604	887	172	1.51	2.85	4.18	0.81
120 (118+119) Total number of instances in which customers received a First Deliverable or Identified Deliverable* (Delivered Updates and Fixes, SAS)	1,228	1,511	1,794	172	5.79	7.13	8.46	0.81
121 Total number of First Deliverables containing documentation referencing fix IDs (Delivered Updates and Fixes)		583				2.75		

\* This measure is reported by taking into account that one of its components is known in the full population.

**TABLE 12B: CUSTOMER-RELATED MEASURES FOR CRITICAL SUPPORT**

**Critical Support**  
**Sample Size: 238**  
**Population Size: 973**  
**Confidence Level: 90%**

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
118 Total number of instances in which customers received a First Deliverable (Delivered Updates and Fixes)		26,054				26.78		
119 Total number of instances in which customers who did not receive a First Deliverable received an Identified Deliverable (SAS)	3	16	30	8	0.00	0.02	0.03	0.01
120 (118+119) Total number of instances in which customers received a First Deliverable or Identified Deliverable* (Delivered Updates and Fixes, SAS)	26,057	26,070	26,084	8	26.78	26.79	26.81	0.01
121 Total number of First Deliverables containing documentation referencing fix IDs (Delivered Updates and Fixes)		26,931				27.68		

\* This measure is reported by taking into account that one of its components is known in the full population. Rounding in the average

Tables 13A and 13B show some measures that address the issue of Contamination. Measures 122 and 125 are counted for the full population, and therefore are reported with full

certainty. Measures 123, 124, and 126 are recorded for the sample. They are reported with 90% confidence intervals.

**TABLE 13A: CONTAMINATION-RELATED MEASURES FOR RETROFIT**

Retrofit  
 Sample Size: 46  
 Population Size: 212  
 Confidence Level: 90%

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
122 Total number of instances in which customers received a First Deliverable contaminated through receipt of a DAT file not specifically generated for those customers (Analysis)		256				1.21		
123 Total number of instances in which customers received a First or Identified Deliverable contaminated by use of a generic environment, of some other customer's environment, or of an environment built from another customer's software (Analysis)	947	1,350	1,754	245	4.47	6.37	8.27	1.16
124 Total number of instances in which customers received a First or Identified Deliverable contaminated by development using the source group model* (Analysis)	0	0	0	0	0.00	0.00	0.00	0.00
125 Total number of instances in which customers received a First Deliverable contaminated because another customer received at least one COBOL, SQR, SQC or DAT Associated File with an identical hash value (Analysis)		723				3.41		
126 Total number of instances in which customers received a First or Identified Deliverable other than from an environment built solely from their software and used solely for them (Analysis)	982	1,387	1,793	247	4.63	6.54	8.46	1.16

\*Measure 124 was zero in all 46 observations of the Retrofit sample files, which makes the calculation of an upper bound for this measure less pertinent. A conservative estimate of the upper bound, using formulas from classical sampling theory, is zero. There are further refinements available, such as an exact binomial test, that allow for the calculation of the probability of a non-zero value in the population. Such a calculation implies that there is a 4.8% chance of observing a non-zero value, which translates to no more than 10 non-zero occurrences in the population.

**TABLE 13B: CONTAMINATION-RELATED MEASURES FOR CRITICAL SUPPORT**

Critical Support  
 Sample Size: 238  
 Population Size: 973  
 Confidence Level: 90%

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
122 Total number of instances in which customers received a First Deliverable contaminated through receipt of a DAT file not specifically generated for those customers (Analysis)		19,899				20.45		
123 Total number of instances in which customers received a First or Identified Deliverable contaminated by use of a generic environment, of some other customer's environment, or of an environment built from another customer's software (Analysis)	20,884	23,552	26,220	1,622	21.46	24.21	26.95	1.67
124 Total number of instances in which customers received a First or Identified Deliverable contaminated by development using the source group model (Analysis)*	49	319	683		0.05	0.33	0.70	
125 Total number of instances in which customers received a First Deliverable contaminated because another customer received at least one COBOL, SQR, SQC or DAT Associated File with an identical hash value (Analysis)		23,429				24.08		
126 Total number of instances in which customers received a First or Identified Deliverable other than from an environment built solely from their software and used solely for them (Analysis)	21,901	24,660	27,420	1,678	22.51	25.34	28.18	1.72

The lower and upper bounds are calculated using 10,000 iterations of repeated sampling. The confidence interval is directly computed. Therefore the standard error is not computed. See B. Efron & R.J. Tibshirani, An Introduction to the Bootstrap, (Chapman & Hall: 1993), 168-176.

Tables 14A and 14B show additional measures that address the issue of Contamination. Measures 127 and 130 are all recorded for the full population, and therefore are reported with full certainty. Measures 128 and 129 are composite measures, as defined in the preceding section. Measures 130 and 131 are reported for the sample. Measures 128, 129, 131 and 132 are reported with a 90% confidence interval. Measure 131 shows that in instances in which customers received a first deliverable Retrofit Fix, 83.92% of the First Deliverables were contaminated based on Object analysis. This same measure for the Critical Support Fix population is 99.12%.

**TABLE 14A: CONTAMINATION-RELATED COMPOSITE AND RATIO MEASURES FOR RETROFIT**

Retrofit  
 Sample Size: 46  
 Population Size: 212  
 Confidence Level: 90%

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
127 Total number of instances in which customers received a contaminated First Deliverable based on analysis of delivered objects (Analysis) (Union 122,125)		814				3.84		
128 Total number of instances in which customers received a contaminated First or Identified Deliverable based on analysis of objects and of development and testing documentation* (Union 122,123,124,125)	1,019	1,224	1,430	125	4.80	5.77	6.74	0.59
129 Total number of instances in which customers received a contaminated fix based on object and documentation analysis, when cross-use of a customer-specific environment renders activities on behalf of the customers infringing* (Analysis) (Union 122, 123, 124, 125, 126)	1,028	1,279	1,531	153	4.85	6.04	7.22	0.72
130 Percentage of instances in which customers received a First Deliverable where that First Deliverable was contaminated based on analysis of objects found in Delivered Updates and Fixes (Analysis) (127/118)		89.75%						
131 Percentage of instances in which customers received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on object and documentation analysis** (Analysis) (128/120)	71.56%	83.92%	96.29%	7.52%				
132 Percentage of instances in which customers received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on object or documentation analysis or on the fact that the customer's environment was used to support other customers** (Analysis) (129/120)	74.64%	87.19%	99.74%	7.63%				

\* These measures are reported by taking into account that some of their components are known in the full population.

\*\* Based on the values for the numerator and denominator in the sample of 46 fixes, as reported in Table A 2 in Appendix 3. These values are used to provide classic estimates of the standard errors. Values in Tables 12A and 14A produce a value of 81.03% for measure 131 and 84.69% for measure 132.

**TABLE 14B: CONTAMINATION-RELATED COMPOSITE AND RATIO MEASURES FOR CRITICAL SUPPORT**

Critical Support  
 Sample Size: 238  
 Population Size: 973  
 Confidence Level: 90%

Measure	Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
127 Total number of instances in which customers received a contaminated First Deliverable based on analysis of delivered objects (Analysis) (Union 122,125)		24,417				25.09		
128 Total number of instances in which customers received a contaminated First or Identified Deliverable based on analysis of objects and of development and testing documentation* (Union 122,123,124,125)	24,895	25,390	25,885	301	25.59	26.09	26.60	0.31
129 Total number of instances in which customers received a contaminated fix based on object and documentation analysis, when cross-use of a customer-specific environment renders activities on behalf of the customers infringing* (Analysis) (Union 122, 123, 124, 125, 126)	24,796	25,410	26,025	374	25.48	26.12	26.75	0.38
130 Percentage of instances in which customers received a First Deliverable where that First Deliverable was contaminated based on analysis of objects found in Delivered Updates and Fixes (Analysis) (127/118)		93.72%						
131 Percentage of instances in which customers received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on object and documentation analysis*** (Analysis) (128/120)	98.65%	99.12%	99.52%					
132 Percentage of instances in which customers received a First or Identified Deliverable where that First or Identified Deliverable was contaminated based on object or documentation analysis or on the fact that the customer's environment was used to support other customers*** (Analysis) (129/120)	98.73%	99.19%	99.57%					

\* These measures are reported by taking into account that some of their components are known in the full population

\*\* Based on the values for the numerator and denominator in the sample of 238 fixes, as reported in Table A.3 in Appendix 3. These values are used to provide classic estimates of the standard errors. Values in Tables 12B and 14B produce a value of 97.39% for measure 131 and 97.47% for measure 132.

\*\*\* The lower and upper bounds are calculated using 10,000 iterations of repeated sampling. The confidence interval is directly computed. Therefore the standard error is not computed. See B. Efron & R.J. Tibshirani, An Introduction to the Bootstrap, (Chapman & Hall: 1993), 168-176

Tables 15A and 15B present results for measures that address Object Contamination. All these measures are counted for the entire population of Fixes. Measure 138 demonstrates that 87.66% of hashes that were associated with a First Deliverable were contaminated in the Retrofit Fix population. This same measure for the Critical Support Fix population is 67.96%.

**TABLE 15A: CONTAMINATION-RELATED MEASURES FOR RETROFIT BASED ON OBJECT ANALYSIS**

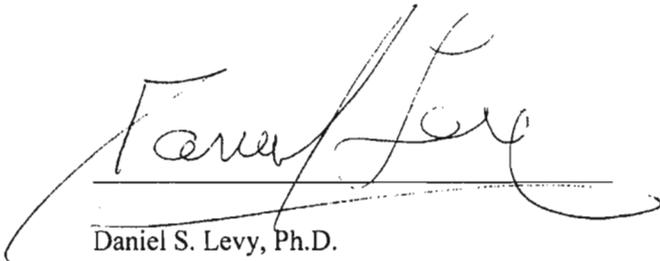
Retrofit		
Population Size: 212		
Measure	Total	Average
135 Total number of unique hashes for the set of COBOL, SQR, SQC, or DAT Associated Files in any First Deliverable contaminated because they were delivered to more than one customer (Delivered Updates and Fixes--Analysis)	1,628	7.68
136 Total number of unique hashes for the set of DAT Associated Files in any First Deliverable contaminated because they were created using a generic environment or created with one customer's environment but delivered to another customer (Delivered Updates and Fixes)	200	0.94
137 Total number of contaminated, unique hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable (Delivered Updates and Fixes--Analysis) (Union 135,136)	1,727	8.15
138 Percentage of hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable that are contaminated based on object analysis (Delivered Updates and Fixes--Analysis) (137/(104+142))	87.66%	
144 Percentage of hashes for the set of DAT Associated Files in any First Deliverable contaminated because they were created using a generic environment or created with one customer's environment but delivered to another customer (Delivered Updates and Fixes--Analysis) (136/142)	89.29%	

**TABLE 15B: CONTAMINATION-RELATED MEASURES FOR CRITICAL SUPPORT BASED ON OBJECT ANALYSIS**

Critical Support		
Population Size: 973		
Measure	Total	Average
135 Total number of unique hashes for the set of COBOL, SQR, SQC, or DAT Associated Files in any First Deliverable contaminated because they were delivered to more than one customer (Delivered Updates and Fixes--Analysis)	4,073	4.19
136 Total number of unique hashes for the set of DAT Associated Files in any First Deliverable contaminated because they were created using a generic environment or created with one customer's environment but delivered to another customer (Delivered Updates and Fixes)	2,988	3.07
137 Total number of contaminated, unique hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable (Delivered Updates and Fixes--Analysis) (Union 135,136)	5,112	5.25
138 Percentage of hashes for the set of COBOL, SQR, SQC or DAT Associated Files in any First Deliverable that are contaminated based on object analysis (Delivered Updates and Fixes--Analysis) (137/(104+142))	67.96%	
144 Percentage of hashes for the set of DAT Associated Files in any First Deliverable contaminated because they were created using a generic environment or created with one customer's environment but delivered to another customer (Delivered Updates and Fixes--Analysis) (136/142)	82.59%	

## 10) Conclusion

In this report, I have applied standard statistical theory to the question at hand. I have discussed the reasons that sampling is appropriate in this particular setting. I have presented my results above. These results are based on standard statistical formulas that are used in sampling situations.

A handwritten signature in cursive script, appearing to read "Daniel S. Levy". The signature is written in black ink and is positioned above a horizontal line.

Daniel S. Levy, Ph.D.

November 16, 2009

## Appendix 1 – Technical Appendix

Table A.1 lists the variable names and definitions that are used in the calculations.

**TABLE A.1: VARIABLE DEFINITIONS**

Variable	
Name	Variable Description
$N$	Population Size
$n$	Sample Size
$y_i$	Measure of Interest (information collected for each measure) for fix $i$ ; The numerator for Ratio Measures.
$\bar{y}$	Sample Mean
$\hat{Y}$	Estimated Population Total
$s_y$	Standard Error of Sample
$s_{\bar{y}}$	Standard Error of Population Mean
$s_{\hat{Y}}$	Standard Error of Population Total
$\bar{x}$	Sample Mean for Denominator for Ratio Measures
$\hat{R}$	Estimated Population Ratio
$s_{\hat{y}}$	Sample Variance of Numerator of Ratio Measures
$s_{\hat{x}}$	Sample Variance of Denominator of Ratio Measures
$s_{yx}$	Sample Covariance of Numerator and Denominator for Ratio Measures
Var	Variance
Cov	Covariance

### a. Mean Per Unit Estimator – Related Formulas

I use the following equations to calculate the various results for measures 102, 114, 116, 119, 123, 124, 126, 139, 140 and 141:

- 1) Sample Mean:  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
- 2) Estimated Population Total:  $\hat{Y} = N\bar{y}$
- 3) Standard Error of the Sample:  $s_y = \sqrt{\sum_i (y_i - \bar{y})^2 / (n - 1)}$
- 4) Standard Error of the Sample Mean:  $s_{\bar{y}} = \frac{s_y}{\sqrt{n}}$
- 5) Standard Error of the Population Total:  $s_{\hat{Y}} = N s_{\bar{y}}$ .

**b. Union Measures – Related Formulas**

Measures 117, 128 and 129 are the unions of two or more measures where there may be some overlap in the occurrences counted in each measure. The calculated population totals and variance of these types of measures take into account this overlap.

For measure 117, the following relationships are applicable:

- 1)  $\hat{Y}_{117} = Y_{115} + \hat{Y}_{116} - \hat{Y}_{115 \cap 116}$ , where  $\hat{Y}_{117}$  is the estimated population total for measure 117,  $Y_{115}$  is the fully measured population total for measure 115,  $\hat{Y}_{116}$  is the estimated population total for measure 116 and  $\hat{Y}_{115 \cap 116}$  is the estimated population total for the overlap between measures 115 and 116.  $y_{115 \cap 116}$  for each fix in the sample is calculated as  $y_{115} + y_{116} - y_{117}$ .
- 2) Because measure 115 is fully measured, it introduces no variance into measure 117. Therefore,  $Var(\hat{Y}_{117}) = Var(\hat{Y}_{116}) + Var(\hat{Y}_{115 \cap 116}) - 2Cov(\hat{Y}_{116}, \hat{Y}_{115 \cap 116})$ .
- 3) Similarly, results can be derived for measures 128 and 129.

**c. Sum of Two Measures – Related Formulas**

- 1)  $\hat{Y}_{120} = Y_{118} + \hat{Y}_{119}$ , where  $\hat{Y}_{120}$  is the estimated population total for measure 120,  $Y_{118}$  is the fully measured population total for measure 118,  $\hat{Y}_{119}$  is the estimated population total for measure 119.
- 2) Because measure 118 is fully measured, it introduces no variance into measure 120. Therefore,  $Var(\hat{Y}_{120}) = Var(\hat{Y}_{119})$
- 3) Similarly, results can be derived for measure 134.

**d. Ratio Estimator – Related Formulas**

If the numerator is y and the denominator is x, then the ratio,  $R$ , is estimated as the ratio of the sample means:  $\hat{R} = \bar{y}/\bar{x}$ . Unlike the mean per unit estimator, the ratio estimator

may be biased but the bias is considered negligible in most cases.<sup>43</sup> The upper bound of this bias can be estimated.<sup>44</sup> Furthermore, because a ratio estimator is the ratio of two random variables, the variance of this estimator is not well defined. However, it can be derived using the Delta Method.<sup>45</sup>

In the sample, values for the numerator and denominator were reported for measures 131 and 132. I calculate the ratio estimate,  $\hat{R}$ , by first calculating the sample mean of the numerator as shown in Step 1 below. I then calculate the sample mean of the denominator as shown in Step 2 below. Step 3 illustrates the calculation for the ratio estimate,  $\hat{R}$ , as the ratio of these two sample means. Finally, Step 4 shows the approximation for the standard error of  $\hat{R}$ .

- 1) Sample Mean of Numerator:  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$
- 2) Sample Mean of Denominator:  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- 3) Estimate of the Population Ratio:  $\hat{R} = \bar{y}/\bar{x}$
- 4)  $Var(\hat{R}) = \frac{1}{n\bar{x}^2} [s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{yx}]$

#### e. Estimates for Lower and Upper Bounds Based on Repeated Sampling

For four measures for the Critical Support sample, I use repeated sampling to estimate the upper and lower bounds because these estimates are close to the boundaries. For measure

---

<sup>43</sup> William G. Cochran, Sampling Techniques, Third Edition (New York: Wiley & Sons, 1977), 153. Paul S. Levy and Stanley Lemeshow, Sampling of Populations, Fourth Edition (New York: Wiley & Sons, 2008), 191.

<sup>44</sup> William G. Cochran, Sampling Techniques, Third Edition (New York: Wiley & Sons, 1977), 162.

<sup>45</sup> George Casella and Roger L. Berger, Statistical Inference, Second Edition (California: The Wadsworth Group/Duxbury, 2002), 243-245. William G. Cochran, Sampling Techniques, Third Edition (New York: Wiley & Sons, 1977), 155. John A. Rice, Mathematical Statistics and Data Analysis, Third Edition (California: Brooks/Cole, 2007), 165-166.

140 for the Critical Support, of the 238 observations in the sample, 237 have a value of zero and one has a value of 2. The formula for the confidence interval based on the normal approximation would yield a lower bound of less than zero. The point estimate for measure 131, which is the ratio of measure 128 to measure 120, is very close to 100%. For measure 132, which is the ratio of measure 129 to measure 120, the standard formula for the confidence interval would yield an upper bound of greater than 100%. For measure 124 for Critical Support, the formula for the confidence interval based on the normal approximation yields a lower bound of less than zero. To address these issues that can occur at the boundaries (i.e. close to 0 in the case of a count variable such as measure 140 or beyond 0 or 1 for a proportion measure such as measure 132), I sample repeatedly from the data to estimate the upper and lower bounds. This method involves drawing repeatedly and with replacement from the sample to calculate a statistic of interest in each of the samples. The resulting distribution of values of that statistic allows the estimation of the upper and lower bounds. My methodology involves 10,000 repeated draws from the sample for each of these measures.<sup>46</sup>

---

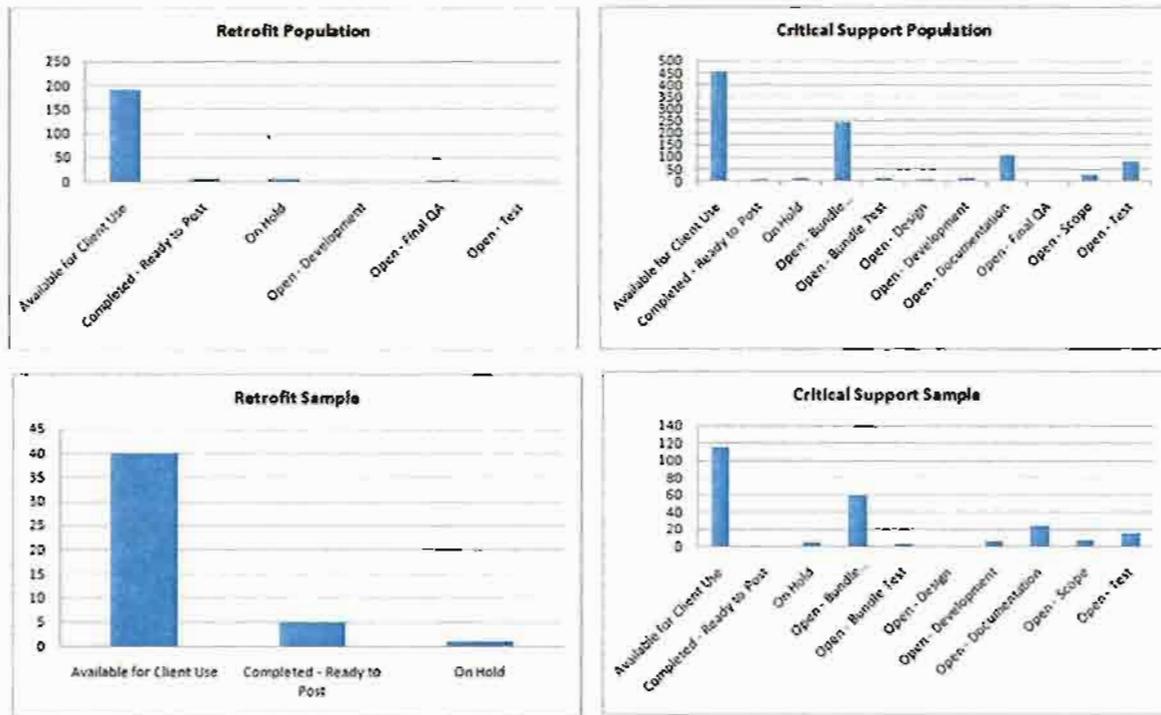
<sup>46</sup> See B. Efron & R.J. Tibshirani, An Introduction to the Bootstrap, (Florida: Chapman & Hall/CRC, 1993), 168-176.

## Appendix 2 – Additional Figures and Tables

### A. Distribution of Status of Fixes: Measure 103

The figure below shows the distribution of the status of Fixes in the populations and the samples of Retrofit and Critical Support Fixes.

FIGURE 1: DISTRIBUTION OF STATUS OF FIXES



### B. Composite Measures – Additional Tables

The results for measures 117, 120, 128, 129 and 134 are also reported in Section 8 of this report. These are composite measures because they are either the sum or the union of two or more measures. In Section 8, I report the results for these measures taking into account the fact that information was collected for one of the component measures for the entire population. The use of a full count of the entire population for one component eliminates the variance of that component in the composite measure. For the sake of completeness, however, in this appendix I present the results for these measures based on the samples.

TABLE A.2: COMPOSITE MEASURES FOR RETROFIT

Retrofit  
 Sample Size: 46  
 Population Size: 212  
 Confidence Level: 90%

Measure - EXTRAPOLATED FROM SAMPLE		Total Lower Bound	Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
117	Total number of environments used in development or testing of fixes (SAS, Delivered Updates and Fixes, Consultant Docs and Templates) (Union 115,116)	476	571	667	58	2.24	2.70	3.15	0.27
120	Total number of instances in which customers received a First Deliverable or Identified Deliverable (Delivered Updates and Fixes, SAS) (118+119)	1,270	1,691	2,113	256	5.99	7.98	9.97	1.21
128	Total number of instances in which Recipients received a contaminated First or Identified Deliverable based on object and documentation analysis (Analysis) (Union 122,123,124,125)	1,017	1,419	1,822	245	4.80	6.70	8.59	1.15
129	Total number of instances in which customers received a contaminated fix based on object and documentation analysis, when cross-use of a customer-specific environment renders activities on behalf of the customers infringing (Analysis) (Union 122, 123, 124, 125, 126)	1,066	1,475	1,883	248	5.03	6.96	8.88	1.17
134	Total number of copies of objects comprising First or Identified Deliverables (including copies located in compressed files)(Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak, SAS) (133+114)	13,057	34,505	55,944	13,034	61.63	162.76	263.89	61.48

TABLE A.3: COMPOSITE MEASURES FOR CRITICAL SUPPORT

Critical Support  
 Sample Size: 238  
 Population Size: 973  
 Confidence Level: 90%

Measure - EXTRAPOLATED FROM SAMPLE		Total Lower Bound	Estimated Total	Total Upper Bound	Standard Error of Total	Average Lower Bound	Average	Average Upper Bound	Standard Error of Average
117	Total number of environments used in development or testing of fixes (SAS, Delivered Updates and Fixes, Consultant Docs and Templates) (Union 115,116)	5,507	6,034	6,562	321	5.66	6.20	6.74	0.33
120	Total number of instances in which customers received a First Deliverable or Identified Deliverable (Delivered Updates and Fixes, SAS) (118+119)	26,569	29,345	32,122	1,688	27.31	30.16	33.01	1.73
128	Total number of instances in which Recipients received a contaminated First or Identified Deliverable based on object and documentation analysis (Analysis) (Union 122,123,124,125)	26,312	29,088	31,864	1,688	27.04	29.89	32.75	1.73
129	Total number of instances in which customers received a contaminated fix based on object and documentation analysis, when cross-use of a customer-specific environment renders activities on behalf of the customers infringing (Analysis) (Union 122, 123, 124, 125, 126)	26,332	29,108	31,884	1,688	27.06	29.92	32.77	1.73
134	Total number of copies of objects comprising First or Identified Deliverables (including copies located in compressed files)(Delivered Updates and Fixes, Data Warehouse, Environment Backup, BakTrak, SAS) (133+114)	590,170	663,978	737,787	44,872	606.55	682.40	758.26	46.12

## Appendix 3 – Curriculum Vitae

### ADVANCED ANALYTICAL CONSULTING GROUP

**DANIEL S. LEVY**

**ECONOMIST**

#### **EDUCATION**

Ph.D., Economics, The University of Chicago

A.B., Economics, The University of Chicago (With Special Honors in Economics)

Daniel S. Levy specializes in applications of economics and statistics in the study of corporate structures related to industrial organization/antitrust, damages issues, and corporate performance. He has studied economic, statistical and computing issues in a number of industries, including airlines, environment, oil and gasoline, healthcare, labor markets, pharmaceuticals, telecommunications, and high technology markets among others. His work includes detailed analyses and valuations of corporate functions, risks, and assets for international corporations for use in business decision and litigation. He has testified in State and Federal court and for Government Agencies.

Antitrust and Industrial Organization: For antitrust matters, Dr. Levy has studied the ability of collusive agreements to increase prices and has investigated the extent to which primary customers can pass along price increases to secondary customers. In addition, his work has focused on basic chemicals and commodities. He has studied international and national market prices for commodities such as citric acid, nickel, vitamins oil and gasoline.

Statistics and Sampling: He has also testified about sampling and statistical issues in Federal Court, presented statistical issues to the Department of Justice, the Securities Exchange Commission, other Government Agencies, and served as a court appointed Expert Arbitrator for the United States Internal Revenue Service.

Dr. Levy has developed and implemented advanced analytical methods for quality control tests for major corporations. For more than a decade, Dr. Levy led a team of economists and statisticians in monthly testing of quality of service for multiple telecommunications companies. He also has performed economic and statistical work in telecommunications, transportation, manufacturing, financial services, mining, oil and gas, consumer durables, healthcare, pharmaceuticals and medical devices industries. He has extensive experience in developing statistical methods for practical business applications.

Finance and Financial Services: Dr. Levy has worked on a number of cases involving late trading and rapid trading issues and Market Maker trading behavior brought against financial institutions by the Securities and Exchange Commission and the New York Attorney General. He has performed a wide range of event studies for securities valuations for 10B-5 disputes and for general damages. He has estimated damages associated with late and rapid trading. He has studied the appropriate use of interest rates for use in damages models. He has presented his research in Federal Court, and before Government Agencies including DOJ, NY-OAG, FCC, and SEC among others.

He is expert in numerous statistical and modeling applications, and has modeled complex economic and social factors affecting, labor, demographic and market behavior.

Prior to founding Advanced Analytical Consulting Group, Dr. Levy was the National Market Leader for Economic and Statistical Consulting for Deloitte Financial Advisory Services and had served as the Global Leader of Economic Consulting for Arthur Andersen. Prior to that he held research and consulting positions at Charles River Associates, The RAND Corporation, Needham-Harper Worldwide Advertising, SPSS Inc. and The University of Chicago Computation Center.

#### **EXPERT TESTIMONY/AFFIDAVITS**

- Drs Newco III, Inc v. Night Vision Equipment Company Holding, Inc, 2008, Expert Report, Testimony, *Damages in High Technology Market.*
- Invesco Institutional (N.A.) Inc v Deutsche Investment Management Americas, Inc, 2008, Expert Report, *Damages in Financial Services Industry.*
- Securities and Exchange Commission v. Kenneth D. Pasternak and John P. Leighton, 2007, 2008. Expert Report, Testimony, *Securities Trading and Market Making Damages.*
- Cytologix v. Ventana, 2002, 2007 Expert Report, Depositions, Testimony, *Antitrust in High Technology Medical Market.*
- Rubin Squared Inc. v. Cambrex Corporation, 2006 Case No. 03-CIV. 10138(PAC) Expert Report.

- Polaris Industries Inc. v. Commission of Revenue, 2005, Expert Report, Minnesota Tax Court, Docket No. 7694-R
- Before the New Mexico Department of Insurance, 2004, Expert Report, *Health Insurance Merger*.
- Carolyn Fears, et al. v. Wilhelmina Model Agency, Inc., et al., 2003, Expert Report and Deposition, *Antitrust Price-Fixing*.
- Shoshone and Arapaho Indian Tribes v. the United States of America, 2003, Expert Report and Deposition, *Statistical Sampling*.
- Pechiney Plastic Packaging Inc. v. Continental PET Technologies Inc. 2002, Expert Report and Deposition, 2002, *Statistical Sampling/Patent Infringement*.
- Before the Illinois Commerce Commission, 2001, Expert Report and Testimony, *Statistical Methods*.
- IRS Expert Arbitrator, 2000, James Schilling Inc., v. Internal Revenue Service, *Expert Arbitrator Report and Decision*.
- Before the Wisconsin Public Utilities Commission, 2000, Expert Report and testimony *Statistical Methods*.
- Before the New Mexico Insurance Commissioner, 1999, Expert Witness, *Hospital Merge*.
- Before the Michigan Public Service Commission, 1998, Expert Affidavit, *Statistical Analysis*.
- Statistical Methods for Parity Tests of Telecommunications Resale and Retail Markets, Before the Indiana Public Service Commission, 1998, Expert Affidavit, *Statistical Analysis*.
- Before the FCC, CC Docket No. 98-56, RM-9101, 1998, Expert Affidavit, *Statistical Analysis*.

- Graber, A. et al. v. Giuliani, United States District Court Southern District of New York, 1998, Expert Affidavit and Deposition, *Statistical Sampling and Survey Research*.
- Marisol, A. et al. v. Giuliani, United States District Court Southern District of New York, 1998, Expert Affidavit and Deposition, *Statistical Sampling and Survey Research*.
- DFW v. Continental Air Lines, Texas, 1998, Expert Deposition and Testimony.
- Randall's Food Markets, Inc., v. Fleming Companies, Inc., The American Arbitration Association Dallas, Texas, June, 1998, Expert Affidavit, *Statistical Sampling*.
- Randall's Food Markets, Inc., v. Fleming Companies, Inc., The American Arbitration Association Dallas, Texas, February 1998, Expert Report, *Statistical Sampling*.
- Donald E. Haney v. Timesavers Inc., et al. United States District Court, District of Oregon, January 1998, Expert Testimony, *Patent Infringement*.
- Merck-Medco Managed Care Inc. v. Rite Aid Corporation et al. Northern District of Maryland, May 1997, Expert Deposition, *Antitrust*.
- Donald E. Haney v. Timesavers Inc., et al. United States District Court, District of Oregon, July 1997, Expert Report, *Patent Infringement*.
- Kenneth Heubert Williams v. Honri Vashon Hunt et al., State of Michigan in the Circuit Court for the County of Oakland, May 1997, Expert Deposition, *Value of Life*.
- Merck-Medco Managed Care Inc. v. Rite Aid Corporation et al. Northern District of Maryland, April 1997, Expert Report, *Antitrust*.
- Robinson Rubber et al. V. Hennepin County, Minnesota, United States District Court, District of Minnesota, Fourth Division, April 1997, Expert Deposition, *Antitrust*.
- Robinson Rubber et al v. Hennepin County, Minnesota, United States District Court, District of Minnesota, Fourth Division, April 1997, Expert Report, *Antitrust*.

- Massachusetts Wholesalers of Malt Beverages, Inc., v. Commonwealth of Massachusetts et al, Suffolk Superior Court, 1996, Expert Testimony, *Financial Damages*.
- Luke Brothers v. S. P. Krusell, US District Court, District of Massachusetts, July 1996, Expert Affidavit, *Antitrust Price-Fixing*.
- Luke Brothers v. S. P. Krusell, US District Court, District of Massachusetts, August 1996, Expert Affidavit, *Antitrust Price-Fixing*.
- Daras v. Texaco Inc, 1993, Affidavit.
- Environmental Protection Agency: Navajo Generating Station, 1991, Public Comment, *Valuation of Environmental Damages*.

## PROFESSIONAL EXPERIENCE

2009 – Present	National Managing Director, Advanced Analytical Consulting Group, Inc.
2002 - 2009	National Leader Economic and Statistical Consulting, Deloitte FAS L.L.P.
2001 - 2002	Global Leader of Economic Consulting, Arthur Andersen L.L.P.: SFE
1998 – 2001	National Leader of Economic Consulting, Arthur Andersen L.L.P.: SFE
1996 – 1998	Director of Economics, Arthur Andersen L.L.P.:CRCO
1995 - 1996	Economist, Arthur Andersen L.L.P.
1991 - 1995	Senior Associate, Charles River Associates
1988 - 1991	Associate Economist, The RAND Corporation
1985 - 1988	Computer Advisor, University of Chicago Computation Center
1982 - 1985	Research and Teaching Consultant, SPSS Inc.
1981 - 1982	Research Consultant, Needham, Harper Worldwide Advertising

## PROFESSIONAL HONORS and ACTIVITIES

- Earhart Fellowship for graduate research in economics, 1982 - 1983
- Hewlett Grant for research in developing countries, 1985 - 1986; renewed, 1986 - 1987

- CBS Bicentennial Scholarship for research on events leading to the American Revolution, 1986 - 1987
- Homer and Alice Jones Fellowship, University of Chicago, 1987 - 1988
- American Economics Association, 1988- Present
- Population Association of America, 1988-1991

## **PAPERS, Presentations, AND PUBLICATIONS**

Daniel S. Levy. "New Econometric Techniques for Transfer Pricing." Presented at the American Bar Association Annual Meetings, August, 1997.

Daniel S. Levy et al. "Economics and the New Transfer Pricing Regulations: Achieving Arm's Length Through the Invisible Hand." Special Report to *Transfer Pricing Reporter*, Vol. 4, No. 2, May 24, 1995.

Daniel S. Levy and Deloris R. Wright. "In the OECD and the United States, It's the Arm's-Length Principle that Matters: Comparison of New Transfer Pricing Regulations." *International Transfer Pricing Journal* 1, No. 2, January 1995.

Robert Fagan, Manjusha Gokhale, Daniel S. Levy, Peter Spinney, and G.C. Watkins. "Estimating DSM Program Impacts for Large Commercial and Industrial Electricity Users." Presented at 1995 International Energy Program Evaluation Conference, Chicago, IL, August 1995.

Talk on the EPA's decision to require the Navajo Generating Station to reduce emissions to protect visibility in the Grand Canyon. Panel on "Valuation of Environmental Resource Damages," CRA conference on *Economists' Perspectives on Legal Issues Today: Estimating Damages*, Boston, MA, April 23, 1992.

Daniel S. Levy et al. "Conceptual and Statistical Issues in Contingent Valuation: Estimating the Value of Altered Visibility in the Grand Canyon." (MR-344-RC). Santa Monica, CA: RAND Corporation, 1995. Draft submitted to the Environmental Protection Agency, March 1991.

Daniel S. Levy and D. Friedman. "The Revenge of the Redwoods?: Reconsidering Property Rights and Economic Allocation." *The University of Chicago Law Review* (April 1, 1994). Reprinted in *Land Use and Environment Law Review* 26 (September 1995).

Lois Davis, Susan Hosek, Daniel S. Levy and Janet Hanley, "Health Benefits for Military Personnel: An Overview of Their Value and Comparability to Civilian Benefits" (WD-5875-FMP). Santa Monica, CA: RAND Corporation, February 1992.

D. Buddin, J. Hanley, Daniel S. Levy, and D. Waldman. *Promotion Tempo and Enlisted Retention* (R-4135-FMP). Santa Monica, CA: RAND Corporation, August 1991.

Daniel S. Levy et al. "Comments On Contingent Valuation of Altered Visibility in the Grand Canyon Due to Emissions from the Navajo Generating Station." Presented to the Environmental Protection Agency, April 18, 1991.

Daniel S. Levy. "The Economic Demography of the Colonial South." Ph.D. Thesis, Department of Economics, University of Chicago, 1991.

J. DaVanzo and Daniel S. Levy. "Influences on Breastfeeding Decisions in Peninsular Malaysia." Presented at *The Yale Conference on the Family, Gender Differences, and Development*, September 1989.

Daniel S. Levy. "Long-Run Geographic and Temporal Changes in Mortality in the Colonial South." Presented at the annual meeting of the Population Association of America, Baltimore, 1989. Submitted 1995 to *Social Science History*.

Daniel S. Levy. "The Economic Determinants of Family Sizes in Colonial Maryland: Evidence from Colonial Legislators of Maryland." Presented at the Social Science History Association, Chicago, 1989.

Daniel S. Levy. "The Epidemiological Causes of Changing Political Life Expectancies." Manuscript, 1989.

Daniel S. Levy. "The Life Expectancies of Colonial Maryland Legislators." *Historical Methods* 20, No. 1 (Winter 1987): 17-27.

David W. Galenson and Daniel S. Levy. "A Note on Biases in the Measurement of Geographic Persistence Rates." *Historical Methods* 19, No. 4 (Fall 1986): 171-179.

#### Appendix 4 – Information Considered

1. *ORACLE USA, INC., a Colorado corporation, ORACLE INTERNATIONAL CORPORATION, a California corporation, ORACLE EMEA LIMITED, an Irish private limited company, and SIEBEL SYSTEMS INC., a Delaware corporation, Plaintiffs, v. SAP AG, a German corporation, SAP AMERICA, INC., a Delaware corporation, TOMORROWNOW, INC., a Texas corporation, and DOES 1-50, inclusive, Defendants, Fourth Amended Complaint for Damages and Injunctive Relief.* In United States District Court, Northern District of California, San Francisco Division, Filed August 19, 2009.
2. Casella, George and Roger L. Berger, Statistical Inference. Second Edition. Pacific Grove, California: The Wodsworth Group/Duxbury. 2002.
3. Cochran, William G., Sampling Techniques. Third Edition. New York, New York: John Wiley & Sons, Inc. 1977.
4. Efron, B. & R.J. Tibshirani, An Introduction to the Bootstrap. Boca Raton, Florida: Chapman & Hall. 1993.
5. Levy, Paul S. and Stanley Lemeshow, Sampling of Populations. Fourth Edition. New York, New York: John Wiley & Sons, Inc. 2008.
6. Rice, John A., Mathematical Statistics and Data Analysis. Third Edition. Belmont, California: Brooks/Cole. 2007.
7. ORCLX-AACG-000003.
8. ORCLX-AACG-000004.
9. ORCLX-AACG-000005.
10. ORCLX-MAN-000060.
11. ORCLX-MAN-000061.
12. ORCLX-MAN-000205.
13. ORCLX-MAN-000206.
14. Conversations with Mandiant – various dates.