

*Originally published in July of 2008*

## **The EDD Buyer-Fool Hypothesis**

Prepared as a “self-study” OPED to determine the characteristics of a truly objective survey, the following information is presented to highlight potential challenges that may need to be addressed to ensure the true objectivity of one of the industry’s most quoted electronic discovery surveys - that being the Annual Socha-Gelbmann Electronic Discovery Survey.

### **What is a hypothesis?**

A hypothesis is an explanation for a phenomenon which can be tested in some way which ideally either proves or disproves the hypothesis.

### **What hypothesis is being tested in this article?**

The EDD Buyer-Fool Hypothesis made by the authors of the 2008 Electronic Discovery Survey, that states “Anyone who makes buying decisions primarily on these rankings [EDD Survey] or any other generalized rankings is a fool.”

### **Reasons for agreement with the hypothesis?**

- Selection Bias
- Ethical Violations
- Conflict Of Interest
- Available Data Challenges

### **What is Selection Bias?**

Selection bias is a distortion of evidence or data that arises from the way that the data are collected. It is sometimes referred to as the selection effect. The term selection bias most often refers to the distortion of a statistical analysis, due to the method of collecting samples. If the selection bias is not taken into account then any conclusions drawn may be wrong.

## **Types of Selection Bias that may validate hypothesis:**

- **Self Selection Bias:** Self-selection bias is possible whenever the group of people being studied has any form of control over whether to participate. Participants' decision to participate may be correlated with traits that affect the study, making the participants a non-representative sample. Self-Selection Bias: Fact that organizations could decline to participate in survey. Anecdotally, at least one vendor stated that they declined to participate in the survey.
- **Sampling Bias:** An extreme form of biased sampling occurs when certain members of the population are totally excluded from the sample (that is, they have zero probability of being selected).
- **Biased Sampling:** Assumption that not all members of known 600 vendors were given a chance to participate in the survey.
- **Assumption Basis:** Only 107 of 600+ vendors responded. Anecdotally, at least one vendor asked to participate in survey as they were not invited to participate.

## **What might determine an ethical violation in relation to surveys?**

Based on the Council of American Research Organization's (CASRO) Code of Standards and Ethics for Survey Research, a research organization's report to a Client or the Public should contain, or the Research Organization should be ready to supply to a Client or the Public on short notice, key information about a survey to include:

1. A description of the sample design, including the method of selecting sample elements, method of interview, cluster size, number of callbacks, Respondent eligibility or screening criteria, and other pertinent information.
2. A description of results of sample implementation including (a) a total number of sample elements contacted, (b) the number not reached, (c) the number of refusals, (d) the number of terminations, (e) the number of non-eligible's, (f) the number of completed interviews.
3. A description of any weighting or estimating procedures used.
4. A description of any special scoring, data adjustment or indexing procedures used. (Where the Research Organization uses proprietary techniques, these should be described in general and the Research Organization should be prepared to provide technical information on demand from qualified and

technically competent persons who have agreed to honor the confidentiality of such information).

5. Estimates of the sampling error and of data should be shown when appropriate, but when shown they should include reference to other possible sources of error so that a misleading impression of accuracy or precision is not conveyed.

### **Types of Ethical Violations that may validate hypothesis:**

Assumption is that the current survey may not provide one or more of the aforementioned elements (1 thru 5 above) to the public, thus constituting a incongruence between the CASCO Code of Standards and Ethics and the actual presentation of the survey. Assumption Basis: Past Survey Presentation.

### **What is a “Conflict of Interest?”**

A conflict of interest is a situation in which someone in a position of trust, such as a lawyer, insurance adjuster, a politician, executive or director of a corporation or a medical research scientist or physician, has competing professional or personal interests. Such competing interests can make it difficult to fulfill his or her duties impartially. A conflict of interest exists even if no unethical or improper act results from the “conflict”.

### **Type of Conflict of Interest that may validate hypothesis:**

Financial Conflict of Interest: Survey authors have three interrelated financial interests in the electronic discovery area that include Personal Consulting Practices (providing services for fees to clients that include electronic discovery vendors), Participation in the Electronic Discovery Reference Model (receive funds from participating organizations to include electronic discovery vendors) and Provision of the Annual Electronic Data Discovery Survey (which contains information gathered from entities to include electronic discovery vendors and is provided for a fee to requesting entities). These interests create known situations where survey participants may be members of the EDRM and/or may be consulting clients of the report authors. Based with on past survey results, these financial interests have not been fully disclosed and survey results have not been validated by a third party to ensure no impropriety in face of financial conflict of interest.

## **What might be challenges associated with the survey data?**

Based on data provided in the 2008 Survey preview, the data points listed below provide some challenges for those seeking to use the Survey as a primary source for buying decisions:

- Only 107 EDD services and software providers interviewed of universe of > 600. In order for the sample set to provide a confidence level of 95% and a confidence interval of 4, at least 300 vendors would have needed to have been “surveyed”. Additionally, from a weighting perspective – the rankings appear to weigh vendor self assessments as the primary source of rankings as opposed to the actual end user consumers of EDD services and software.
- Only 29 law firms and 19 corporations interviewed in respective universes that are well above at least 600 respectively. Thus these numbers also do not provide a confidence level of 95% or confidence level of 4. Additionally, from a weighting perspective, a very limited number of these (law firms/corporations) actual end user consumers of EDD services and software are included in the survey thus potentially having disproportionately low contribution to overall survey rankings.

## **Conclusion:**

Based on the presented Selection Bias, Ethical Violations, Conflict Of Interest and Available Data Challenges, there appears to be enough evidence to suggest The EDD Buyer-Fool Hypothesis is accurate and could be represented as a Theory based on hypothesis testing results.

## **References:**

600 Vendors <http://www.law.com/jsp/legaltechnology/pubArticleLT.jsp?id=1191229387053>

Sample Size Calculator - <http://www.surveysystem.com/sscalc.htm>