

12-2008

## Evaluation of the stability and validity of participant samples recruited over the Internet.

Daniel Z Lieberman

Follow this and additional works at: [https://hsrc.himmelfarb.gwu.edu/smhs\\_psych\\_facpubs](https://hsrc.himmelfarb.gwu.edu/smhs_psych_facpubs)

 Part of the [Mental and Social Health Commons](#), [Psychiatry Commons](#), and the [Psychiatry and Psychology Commons](#)

---

### APA Citation

Lieberman, D. (2008). Evaluation of the stability and validity of participant samples recruited over the Internet.. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, 11 (6). <http://dx.doi.org/10.1089/cpb.2007.0254>

This Journal Article is brought to you for free and open access by the Psychiatry and Behavioral Sciences at Health Sciences Research Commons. It has been accepted for inclusion in Psychiatry and Behavioral Sciences Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact [hsrc@gwu.edu](mailto:hsrc@gwu.edu).

Evaluation Of Subject Samples Enrolled Over The Internet Using Two Different  
Recruitment Strategies

Running title: Evaluation of Internet Subject Samples

Daniel Z. Lieberman, M.D.

Department of Psychiatry and Behavioral Sciences

George Washington University

2150 Pennsylvania Ave, NW, Rm 8-421

Washington, DC 20037

Tel: (202) 741-2899

Fax: (202) 741-2891

Email: [dlieberman@mfa.gwu.edu](mailto:dlieberman@mfa.gwu.edu)

## Abstract

Research that is conducted via the Internet has the potential to reach important clinical populations of subjects who would not participate in traditional studies. Concerns exist, however, about the validity of samples recruited in this manner, especially when subjects are anonymous, and never have contact with study staff. It is not possible to make clinical diagnoses, and subjects may not approach questionnaires with seriousness or even truthfulness. This study evaluated two anonymous samples that were recruited over the Internet to test on online program designed to help problem drinkers. The two studies were conducted three years apart, and significantly different recruitment strategies were utilized. Despite these differences, the two samples were highly similar in demographic and clinical features. Correlations that have been found between variables in traditional non-anonymous studies, were also found in both samples, supporting the validity of the data that was collected. Appropriate skepticism is required when critically evaluating Internet studies. Nevertheless, the results of this study indicate that it is possible to obtain stable, valid data from anonymous subjects over the Internet, even when there are significant differences in the way the subjects are obtained.

## Introduction

Users of the Internet represent a valuable resource to researchers. Communication efficiencies associated with this medium can make subject recruitment easier, and populations who would not participate in traditional research can be accessed and studied. Participating in a research study via the Internet is convenient for subjects, and can allow the preservation of anonymity when sensitive topics are involved.

This new form of conducting research can also have important liabilities. The actual composition of the sample being studied, particularly in the case of anonymous participants, is difficult to know when subjects are participating in a study via the Internet. Researchers lose control over the context in which data are collected <sup>1</sup>. For example, subjects working in a non-clinic or laboratory environment might invest less time and energy in accurately completing questionnaires. An Internet subject may take on a fictional persona <sup>2</sup>, or give false information for a variety of other reasons as well. They may distrust the strategy used to protect their identity, and withhold information they do not want attributed to them. They may not take questionnaires seriously, and provide incorrect information for entertainment, or they may answer the questions without giving them the amount of consideration they require.

In order to evaluate the stability of study samples recruited via the Internet, and to test the validity of the information obtained, this study compared two separate samples of problem drinkers who were recruited to test an automated intervention for alcohol abuse that was delivered over the Internet. Both samples were composed of anonymous

individuals with drinking problems who participated exclusively via the Internet. They did not have contact with the study staff at any time.

## Materials and Methods

### Participants

Subjects were recruited for two separate studies that were done approximately three years apart. Study 1 was designed to evaluate the clinical and demographic characteristics of individuals who visited the Web site, and completed the program<sup>3</sup>. Levels of satisfaction with the program were measured, and correlated with clinical and demographic characteristics. No advertising was used to recruit this sample. Users of the site came from Google, Yahoo!, Excite, Alta Vista, and other search engines, which were noted to list the site on the first page of results when relevant search terms were used.

Study 2 measured participants' level of interest in conventional treatment pre- and post-intervention in order to evaluate the effectiveness of the alcohol abuse program in increasing motivation for change (Lieberman and Huang, unpublished data). In this study, the previous recruitment strategy was not effective at enrolling subjects.

Consequently, paid advertising exclusively purchased from Google was used to recruit subjects. Most of the data collection instruments were the same in the two studies allowing a comparison of the two groups.

Although participants had significant drinking problems, they were not treatment-seeking. Most recognized that they were experiencing adverse effects from their alcohol consumption, but were not yet at the point of wanting to change their behavior. The Web site offered an evaluation that would help them better understand the role of alcohol in their lives. No commitment to change was required to participate in the evaluation. The evaluation was designed to mimic a well-established intervention that increases motivation for change by highlighting the negative aspects of a patient's alcohol use.

Participants registered for the study anonymously, and provided no identifying information. Internet Protocol (IP) addresses, which can potentially be used to identify a user, were not recorded by the application. Informed consent was obtained by having potential participants read an information screen, and then selecting a link to click to indicate whether or not they chose to participate. The study was approved by the George Washington University institutional review board.

## Application

An open source application was developed which was modeled on the Drinkers' Check-up<sup>4</sup>. The Drinkers' Check-up involves a multi-step assessment that encourages a patient to see how alcohol consumption is negatively affecting his or her life<sup>5</sup>. It is hypothesized that this experience increases awareness of, and concern about, these consequences<sup>6</sup>. The process is non-judgmental, non-threatening, and objective. The application source code is freely available for use and modification under a Creative Commons license

(<http://creativecommons.org/licenses/by-nc-sa/3.0/>), and can be obtained from the corresponding author.

The application guided participants through a series of questionnaires including the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) <sup>7</sup>, the Alcohol Use Disorders Identification Test (AUDIT) <sup>8</sup>, the Family Tree Questionnaire <sup>9</sup>, the Decisional Balance Questionnaire <sup>10</sup>, and questionnaires designed to obtain non-identifying demographic information, and history of alcohol consumption. Following the evaluation, participants were given individualized feedback on the outcome.

The SOCRATES evaluates a patient's readiness to change in the context of the transtheoretical model of health behavior change <sup>11</sup>. Patients are scored based on their recognition of having an alcohol problem (recognition subscale), their level of concern about whether alcohol is having a negative effect on their lives (ambivalence subscale), and the degree to which they have initiated behavior change (steps subscale). Patients are given statements to read, and then asked to rate each statement on a five-point Likert scale ranging from strongly disagree to strongly agree. The range of possible subscale scores is recognition: 7-35, ambivalence: 4-20, and steps: 8-40. Higher scores represent greater motivation for change.

The AUDIT is a ten-item questionnaire designed to distinguish light drinkers from those with problem drinking. The items included in the AUDIT were chosen to reflect three dimensions of drinking: alcohol intake; alcohol dependence, such as difficulty in

controlling drinking, neglect of alternative interests, and physiological withdrawal; and adverse consequences. Possible scores range from 0 to 40, with higher scores indicating greater alcohol-related psychopathology.

The Family Tree Questionnaire provides patients with a consistent set of cues for identifying blood relatives with alcohol problems by using a family tree diagram. Patients are asked to classify relatives into one of the following categories: never drank, social drinker, possible problem drinker, definite problem drinker, no relative (applicable only for siblings), or don't know/don't remember <sup>12</sup>.

#### Data Analysis

Results were evaluated both with and without the Bonferroni correction to adjust for multiple comparisons. Clinical and demographic variables that would provide information on who the subjects were, and how they were experiencing their alcohol use were measured and compared. The demographic variables were age, gender, years of education, and employment status. Clinical variables were age of first use of alcohol, age of onset of problem drinking, motivation scores as measured by the SOCRATES, amount of alcohol consumed per week, and AUDIT score. The reported helpfulness of the program modules were also compared.

Summary statistics are presented as means and standard deviations for continuous variables, and percentages for discrete variables. The t-test was used to compare

continuous variables, and the chi-square for discrete ones. In order to assess the validity of the data collected from anonymous subjects, Pearson's correlation was used to evaluate characteristics within the online groups that have been previously found to be associated with one another in conventional studies of non-anonymous subjects. Differences in drinking between genders were also analyzed as a measure of putative validity.

## Results

Over a 24 month period between February 2001 and February 2003, Study 1 enrolled 1,157 subjects. Study 2 enrolled 445 subjects over a period of six months between December, 2005 and May, 2006. Demographic and clinical data from the two samples are shown in Table 1. Subjects from Study 2 were, on average, three years older than those in Study 1. These subjects also rated the AUDIT feedback as being more helpful. In other respects the two groups were similar. Motivation scores, alcohol consumption, and ages of first use and onset of problem drinking were all very similar.

Gender and employment status are shown in Table 2. In both studies about half of the subjects were female. Alcohol use disorders tend to be twice as prevalent in men compared to women<sup>13</sup>, therefore women were over-represented in both online samples, suggesting that there are advantages for women to seek help via the Internet as opposed to the use of traditional clinics. Consistent with the higher than average educational level of this population, most subjects were employed full-time. There was little difference

between the two samples. In both samples, approximately one quarter of the subjects had a father with a drinking problem.

Table 3 shows the correlations between variables in the two studies that have been found to be significantly related to one another in conventional, face-to-face, non-anonymous studies. These variables were measured in order to evaluate the putative validity of the data collected. Overall, each of the relationships was found to be highly significant. The one exception is the relationship between level of alcohol consumption, and the perceived helpfulness of the module that provided feedback on how a subject's drinking compared to national averages. Interestingly, both of the two studies failed to find a correlation between these two variables. Also of note, relationships in one study that had a stronger correlation, as measured by a higher  $r$  value, tended to be stronger in the other study as well. Examples include Recognition x AUDIT, Recognition x Ambivalence, and Ambivalence x AUDIT.

Gender differences are shown in Table 4. Men reported consuming significantly more drinks per week than did women in both studies. In one of the studies, but not the other, men reported an earlier age of onset of problem drinking. There were no differences in the AUDIT scores between men and women indicating that, on average, the two genders experienced similar alcohol-related psychopathology.

## Discussion

The Internet has a clear sampling advantage for populations that are difficult to access. In this case, the sample of interest represented a population of problem drinkers who did not perceive a need for treatment, and would be difficult to bring into a traditional clinic environment. This has been referred to as a “hidden population” of problem drinkers. Being able to administer a preventive intervention to this population has important public health implications (Lieberman and Huang, in press).

There were two important differences in the recruitment of the two samples. Firstly, there was a difference in time of about three years between the two studies. The Internet is evolving rapidly in a number of ways, including the development of a broader user population. It would not have been surprising if the characteristics of users registering for the two studies changed dramatically. The fact that the two samples were so similar provides evidence that stable, representative subject pools can be accessed via the Internet even under anonymous study conditions.

The other difference was that Study 1 recruited from a number of different search engines as a result of being listed in the first page of results, while Study 2 required advertising, and used only Google. It is not clear why the site dropped in the search engine rankings, but a number of factors may have contributed. Most simply, the number of sites indexed by the search engines increased rapidly between 2001 and 2005. Additionally, commercial sites became increasingly aware of the monetary value of being ranked highly by search engines. Search engine optimization became a sophisticated business that attracted a great deal of resources and ingenuity. Our site remained mostly static

between the two studies, and did not use any specific strategies to influence search engine behavior.

Despite the possibility that Google users differ from users of other search engines, or that people who click on advertisements are different from those who follow non-sponsored links in the results section of a search, the differing recruitment methods did not have an effect on the subject pool obtained. This finding suggests that a number of different strategies can be used to recruit subjects via the Internet without affecting the generalizability of the results.

The most important clinical characteristics that were measured were the number of drinks consumed per week, the severity of alcohol-related psychopathology as measured by the AUDIT, and the level of motivation for change as measured by the SOCRATES. In spite of the differences associated with acquiring the two samples, Table 1 shows that these variables were nearly identical across the two groups. Gender and employment status were remarkably similar, and only age differed significantly. Subjects in the second study were, on average, three years older than those in the first study.

In spite of the numerous unknown factors that had the potential to impact negatively on the integrity of the data, both studies revealed internal correlations between variables that replicated relationships observed in other studies which utilized non-anonymous subjects in conventional research environments. For example, a frequently identified relationship between age of onset of alcohol use and age of onset of alcohol problems <sup>14</sup> was also

found in the two Internet samples. The number of drinks consumed per day was significantly correlated with severity of alcohol related problems, and recognition of having an alcohol problem <sup>15-17</sup>. Although the subscales of the SOCRATES measure different components of motivation, they would be expected to be correlated based on where an individual is on the stages of change continuum <sup>11</sup>. This correlation was present in our data suggesting that the SOCRATES questions were answered with a degree of thoughtfulness and accuracy by the participants. In both studies greater ambivalence and recognition of having a drinking problem was significantly associated with perceived helpfulness of the feedback derived from the scale.

Also consistent with findings from other studies, men reported significantly more drinks per week than women <sup>18</sup>. Study 2, but not Study 1 found that men also reported an earlier age of onset of problem drinking <sup>19</sup>. Like other studies that have found that women experience more alcohol-related pathology at lower levels of consumption, there were no significant differences between the AUDIT scores of men and women <sup>20</sup>.

The validity of the data collected from the subjects is supported not only by the consistency with traditional studies but also by the rather striking similarity between the correlation coefficients of many of the variables tested. Mitigating the inherent disadvantages of an online sample, is the ability to collect a large sample size. In Study 1 the sample was over one thousand. Advantages of a larger sample size include the potential for more stable data, and a more detailed analysis. The high statistical power

obtained with a large sample can allow multiple tests to be run on the same dataset without losing significance following a Bonferroni or other correction.

In order to fully take advantage of the strengths of Internet recruitment, it is important to identify ways to enhance the integrity of the study sample. In this case, the program itself provided an incentive for subjects to provide accurate information. The appeal of the program lay in the promise of receiving individualized feedback following a multi-stage assessment of drinking parameters. The more time and effort that a user devoted to ensuring the accuracy of the answers, the more valuable the feedback became.

It is possible that the similarities found could be due to participants from Study 1, also participating in Study 2. Because participants were anonymous, it is not possible to rule this out. Nevertheless it is unlikely due the fact that the assessment battery was the same for both studies. It took a significant amount of time and work to complete the questionnaires, and most people would not find it worthwhile to do it twice.

In general, researchers and clinicians who critically evaluate studies that utilize anonymous subjects recruited via the Internet need to exert caution in interpreting the results of the studies due to the unknown factors associated with this type of study sample. The loss of control over the subjects that is inherent in Internet mediated research must be fully appreciated, and taken into account. However, the liabilities associated with these samples may not be much greater than the weaknesses associated with traditional recruitment methods, such as the trade off between internal validity and generalizability.

The results of this study suggest that valid and reliable data can be obtained, even when the underlying population is accessed via different methods and over extended periods of time.

## References

1. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M (2003) *Psychological research online: opportunities and challenges*. Pittsburgh, PA: Carnegie Mellon University.
2. Smith M, Leigh B. World-Wide Web survey research: Benefits, potential problems and solutions. *Behavior Research Methods, Instruments & Computers*. 1999;29(2):274-279.
3. Lieberman DZ. Clinical characteristics of individuals using an online alcohol evaluation program. *Am J Addict*. Mar-Apr 2005;14(2):155-165.
4. Lieberman DZ. Determinants of satisfaction with an automated alcohol evaluation program. *Cyberpsychol Behav*. Dec 2003;6(6):677-682.
5. Miller WR, Rollnick S (2002) *Motivational Interviewing. Second edition ed.* New York: The Guilford Press.
6. Miller WR, Sovereign RG, Krefe B. Motivational Interviewing with Problem Drinkers: II. The Drinker's Check up as a Preventive Intervention. *Behavioral Psychotherapy*. 1988;16:251-268.
7. Miller WR. Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). Albuquerque, NM: Center on Alcoholism, Substance Abuse, and Addictions; 1992.
8. Babor TF, Grant M. From clinical research to secondary prevention: International collaboration of the Alcohol Use Disorders Identification Test (AUDIT). *Alcohol Health & Research World*. 1989;13:371-374.

9. Mann RE, Sobell LC, Sobell MB. Reliability of a family tree questionnaire for assessing family history of alcohol problems. *Drug Alcohol Depend.* 1985;15:61-67.
10. Carey KB, Maisto SA, Carey MP, Purnine DM. Measuring readiness-to-change substance misuse among psychiatric outpatients: I. Reliability and validity of self-report measures. *Journal of Studies on Alcohol.* 2001;62(1):79-88.
11. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot.* Sep-Oct 1997;12(1):38-48.
12. Vogel-Sprott M, Chipperfield B, Hart DM. Family history of problem drinking among young male social drinkers: reliability of the Family Tree Questionnaire. *Drug & Alcohol Dependence.* 1985;16(3):251-256.
13. Grant BF, Dawson DA, Stinson FS, Chou SP, Dufour MC, Pickering RP. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991-1992 and 2001-2002. *Drug Alcohol Depend.* Jun 11 2004;74(3):223-234.
14. Hingson RW, Heeren T, Winter MR. Age of alcohol-dependence onset: associations with severity of dependence and seeking treatment. *Pediatrics.* Sep 2006;118(3):e755-763.
15. Glanz J, Grant B, Monteiro M, Tabakoff B. WHO/ISBRA Study on State and Trait Markers of Alcohol Use and Dependence: analysis of demographic, behavioral, physiologic, and drinking variables that contribute to dependence and seeking treatment. *International Society on Biomedical Research on Alcoholism. Alcohol Clin Exp Res.* Jul 2002;26(7):1047-1061.

16. U.S. Department of Health and Human Services. Epidemiology of alcohol use and alcohol-related consequences, in Ninth Special Report to the U.S. Congress on Alcohol and Health. Washington, DC: U.S. Department of Health and Human Services; 1997.
17. Hupkens CL, Knibbe RA, Drop MJ. Alcohol consumption in the European community: uniformity and diversity in drinking patterns. *Addiction*. Oct 1993;88(10):1391-1404.
18. Dawson DA, Archer L. Gender differences in alcohol consumption: effects of measurement. *Br J Addict*. Jan 1992;87(1):119-123.
19. Gomberg ES. Women and alcohol: use and abuse. *J Nerv Ment Dis*. Apr 1993;181(4):211-219.
20. Frank JB, Weihs K, Minerva E, Lieberman DZ. Women's mental health in primary care. Depression, anxiety, somatization, eating disorders, and substance abuse. *Med Clin North Am*. Mar 1998;82(2):359-389.

Reprint requests should be directed to Daniel Z. Lieberman, M.D. at  
dlieberman@mfa.gwu.edu

Table 1. Comparison of continuous clinical and demographic variables

	Study 1		Study 2		P
	Mean	SD	Mean	SD	
Age	35	11.3	38	12	.001
Age of first use of alcohol	17	14.2	16	3.8	.46
Age of onset of problem drinking	26	25	26	11.6	.82
Years of education	14	4.1	14	2.3	.22
SOCRATES Recognition score	20.5	7.6	20.7	7.3	.71
SOCRATES Ambivalence score	13.3	4.2	13.3	4.1	.96
SOCRATES reported helpfulness	3.2	1.0	3.4	1.0	.006
Drinks per week	31.8	24.6	29.8	24.7	.15
Consumption feedback helpfulness	3.7	1.0	3.8	1.0	.88
AUDIT	17.7	8.8	17	8.4	.13
AUDIT reported helpfulness	3.6	1.2	3.8	0.9	<.001

Table 2. Comparison of discrete demographic variables

	Study 1	Study 2			
	percent	percent	Chi-square	df	P
Female subjects	48	46	.40	1	.53
Employment			3.48	4	.48
Unemployed	9	9			
Part-time	10	9			
Full-time	65	68			
Homemaker	6	6			
Student	10	8			
Father with a drinking problem	25	27	.50	1	0.48

Table 3. Relationship between variables found to be associated with one another in non-anonymous studies, assessed in order to evaluate the validity of the anonymously collected data.

	Study 1		Study 2	
	r	P	r	P
Age of first use x Age of first problem	0.52	<.001	0.25	<.001
Drinks per week x AUDIT	0.61	<.001	0.61	<.001
Drinks per week x Consumption feedback helpfulness	-0.05	0.14	-0.09	0.1
AUDIT x AUDIT helpfulness	0.26	<.001	0.13	0.02
Recognition x Drinks per week	0.39	<.001	0.31	<.001
Recognition x AUDIT	0.66	<.001	0.65	<.001
Recognition x Ambivalence	0.84	<.001	0.86	<.001
Recognition x SOCRATES helpfulness	0.23	<.001	0.31	<.001
Ambivalence x Drinks per week	0.31	<.001	0.3	<.001
Ambivalence x AUDIT	0.60	<.001	0.62	<.001
Ambivalence x SOCRATES helpfulness	0.23	<.001	0.26	<.001

Table 4. Gender differences in drinking behaviors

Study 1	Men		Women		P
	Mean	SD	Mean	SD	
Age of onset of problem drinking	26.4	28.6	25.5	21.4	.61
Number of drinks per week	35.9	26.9	27.1	20.9	<.001
AUDIT	18.3	8.9	17.1	8.6	.02
Study 2	Men		Women		P
	Mean	SD	Mean	SD	
Age of onset of problem drinking	23.4	8.9	27.8	13.7	.001
Number of drinks per week	36.1	29.1	22.4	15.7	<.001
AUDIT	17.8	8.5	16.1	8.2	.04