The Pennsylvania State University

The Graduate School

The College of the Liberal Arts

THE INCONVENIENCE OF HAPHAZARD SAMPLING

A Thesis In

Sociology and Demography

by

Veronica Leigh Roth

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Arts

August 2011

The thesis of Veronica Leigh Roth was reviewed and approved* by the following:

David Johnson
Professor of Sociology, Demography, and Human Development and Family Studies
Thesis Advisor

Alan Booth
Distinguished Professor of Sociology, Human Development and Demography

Kurt Johnson
Director of the Survey Research Center

Barrett Lee
Professor of Sociology and Demography
Chair of the Graduate Program, Sociology

*Signatures are on file in the Graduate School

Abstract

Targeting specific populations can be an arduous task in survey research. While the ability of researchers to use intricate sampling techniques to build a geographically representative population has increased over the last fifty years, the ability to parse together a representative subsample of certain social characteristics has had less success. For instance, cluster samples can decrease costs of interviews by using fewer locations while yielding a representative sample. However, constructing a sampling frame of hidden populations, such as homeless youths or prostitutes, will still consist of using non-random techniques. Many people who share certain characteristics will likely fall in between these extremes. This study looks at a data set, compiled by Greenburg Quinlan Rosner and named Religion and America's Role in the World (RARW), which compared a nationally representative sample with a subsample of young evangelical Christians. The researchers used a random digit dial to recruit one-fourth of these young evangelicals into a telephone survey, and an internet opt-in, or convenience sample, to recruit the other three-fourths of the evangelicals into a web survey. This mixed-method, mixed-selection strategy poses many interesting methodological questions about representativeness, bias, and the use of post-adjustment strategies. By looking for skew in demographic variables between sample types, and then looking at the sample types in predicting outcome variables to detect bias, it will be possible to look into the first two issues. The use of post-adjustment strategies for dealing with non-random bias is analyzed. Finally, conclusions are presented about the veracity of results of non-probability samples and the advantages and disadvantages of these samples and their post-collection adjustment methods.[1]

[1] This paper uses two data sets, the 2008 General Social Survey and the Religion and America's Role in the World Survey, both of which are available from the Association of Religious Data Archives at www.theARDA.com

Table of Contents

List of Tables…………………………………………………………………………………...v

Introduction…………………………………………………………………………….....1

Literature Review………………………………………………………………………...2

Plan of Analysis………………………………………………………………………….11

Hypothesis……………………………………………………………………………….12

Variables…………………………………………………………………………………14

Methods………………………………………………………………………………….16

Results…………………………………………………………………………………...22

Discussion ……………………………………………………………………………….41

References……………………………………………………………………………….47

List of Tables

Table 1…………………………………………………………………………………………...11

Table 2…………………………………………………………………………………………...21

Table 3…………………………………………………………………………………………...21

Table 4…………………………………………………………………………………………...24

Table 5…………………………………………………………………………………………...27

Table 6…………………………………………………………………………………………...29

Table 7…………………………………………………………………………………………...31

Table 8…………………………………………………………………………………………...33

Table 9…………………………………………………………………………………………...36

Table 10…………………………………..……………………………………………………...37

Table 11……………………………………………………………………..……………….…...38

Table 12………………………………………………………………………………………….39

Introduction

Targeting specific low-incidence populations can be an arduous task in survey research. While the ability to collect survey samples representative of the population in geographically defined areas has increased over the last fifty years, generating surveys large enough to yield a representative subsample of persons with low incidence social characteristics has been less successful. For rare populations with characteristics that are weakly correlated with geographic clusters, building a sampling frame may be too expensive or impractical. Creating a nationally representative data set large enough to yield representative samples are usually too expensive to obtain using standard sampling strategies. A frequently used option in recruiting hard-to-reach populations into a survey is the use of non-representative and convenience samples.

This study looks at the Religion and America's Role in the World (RARW) data set, compiled by Greenburg Quinlan Rosner Research, which included a nationally representative sample of 1,000 respondents and a separate oversample of 400 young evangelical Christians. The researchers used a targeted random-digit dial telephone sample to recruit 100 of these young evangelicals into a telephone survey, and an internet opt-in panel, or convenience sample, to recruit 300 more evangelicals into a web survey.

This mixed-method, mixed-selection strategy raises methodological questions about representativeness, bias, and the use of post-collection adjustment strategies. In this study, I compare the young evangelical samples in the RARW to young evangelicals in the 2008 General Social Survey (GSS), which is a nationally representative sample. I also compare the two samples of evangelicals in the RARW that were selected with different modes. The comparison of the national sample to the combined probability and nonprobability samples of young evangelicals focuses on detecting differences in two outcome variables—feelings towards gays and lesbians and attitudes toward the legality of abortion. In the comparison of the two evangelical-only samples, I use the same variables, but focus on the differences of the probability and nonprobability samples under different model specifications and weighting approaches. Whether differences in demographic distributions of the samples bias the outcomes will be evaluated in both comparisons, using four differently constructed weights to ascertain which adjustment was superior at eliminating non-random bias. Finally, conclusions are presented about the validity of inferences made from nonprobability samples and the advantages and disadvantages of the post-collection adjustment methods to increase the representativeness of the sample in these approaches.

Literature Review

Although survey research is a well-established method, technology has brought a level of flexibility to the discipline that is both useful and problematic. When it comes to sampling a population, survey researchers usually employ a probability approach for reasons of representativeness and statistical necessity. Scheuder, Gregoire, and Weyer (1999/2001) argued that non-probabilistic samples have some utility in scientific research, although this is usually only for sophisticated, mixed-methods research, in which the nonprobability sample is employed out of necessity, or as a small-scale test of the overall research question. Because generalizing to a population with non-probabilistic samples is problematic, they have limited usefulness in quantitative research. In the social sciences, however, some populations are difficult to locate and sample, making nonprobability methods the only option. Rare populations, usually a small overall proportion of a population (e.g.,10% or less), may be found either through building a sampling frame or by finding a sampling frame known to include members of a rare population (Groves, et. al., p. 87, 2009). However, building a frame can be difficult and expensive, as is screening out unqualified respondents (ibid.). These sampling frames are most efficient and feasible when the population of interest is moderately to highly geographically clustered, allowing the researcher to target these geographical areas with a higher probability of selection while maintaining a representative sample. When the demographic distribution of the characteristic is unknown, or there is little geographical clustering, these approaches are generally not economically feasible.

Internet Surveys: Positives & Negatives:

Web-based surveys may seem like an attractive alternative method because recruitment can be targeted at certain websites, perhaps reaching large populations. Using the internet also controls costs. Unfortunately, the internet as a cost control measure may sacrifice sample coverage. As Couper (2001) points out, the sample is biased to internet users, producing coverage error, as those without internet access won't be in the sampling frame. Furthermore, Internet panels are biased toward volunteers, whom are not representative of people with internet access (Valliant & Dever, 2011).

Another issue apparent in internet surveys is satisficing—the condition under which a respondent uses minimal effort to answer a question, and therefore answers a question without fully comprehending what was asked (Krosnick, 1991). Of particular concern in internet surveys is the primacy effect, whereby the respondent simply chooses one of the first categories presented (Krosnick, 1991; Malhorta, 2008). Compared to RDD surveys, internet panels have more issues regarding satisficing, as the lack of interviewer and the opportunity to simply click through the survey gives the respondent an opportunity to fill out the form with minimal effort. This effect seems to be stronger for respondents who have not completed high school or those with a high school diploma (Malhorta, 2008).There is still an effect, however, for educated populations. Using a sample of university freshman, Heerwegh and Loosveldt (2008) found that, on average, web respondents completed the survey sixteen minutes faster than respondents who were randomly assigned to a face-to-face interview mode. Web respondents were also significantly more likely to respond “don’t know” or not respond to an item, and had a significantly smaller differentiation rate on scales (ibid.).

Another serious issue with internet panel surveys is the problem of respondent training. Toepel and colleagues (2008) found that people who have more experience on internet panels had higher inter-item correlation results among multiple items shown on the same screen. Serial survey takers seem to get better at minimizing their effort over time, aggravating the problem of satisficing and reducing the validity of the data. Although satisficing is an issue in all survey research, and not confined to the internet, the combination of not having an interviewer for pacing (Malhorta, 2008) combined with the increasing use of internet panels (Baker, et.al, 2010) makes this issue even more severe.

There are also advantages to internet surveys. In addition to lowered cost and more rapid fielding (Couper, 2001; Schonlau, et.al, 2004), they are known to reduce some types of measurement error. First, by negating the need for direct contact with an interviewer, web surveys can provide privacy, decreasing social desirability bias (Chang & Krosnick, 2009;Fricker et. al., 2005; Kreuter et.al, 2008). Internet surveys, compared to RDD surveys, can yield higher reliability and predictive validity (Chang & Krosnick, 2009), although survey researchers must be careful that these are not artifacts of respondent training or satisficing (Heerwegh & Loosveldt, 2008; Toepel, et.al, 2008). With better designs, i.e., those that include pacing and do not utilize grid responses, web surveys may actually reduce satisficing (Fricker et.al, 2005). Lastly, web surveys are visual, and respondents can take more time to answer questions than they may feel possible if they were talking to an interviewer (ibid.).

<u>Issues Surrounding Nonprobability Samples:</u>

Although probability-based internet surveys have mode effects that distinguish them from RDD samples (de Leeuw, 2005), .they are similar, in some respects, to RDD samples. They can be adjusted for non-response, and as long as internet coverage is also taken into account, they may also yield similar point estimates without weights (Malhorta & Krosnick, 2007). Many of the issues surrounding internet survey research is directly related to their use of nonprobability samples. Respondents may be recruited onto internet panels via website advertising, e-mail offers, or other means (Baker, et.al, 2010).However, all of these nonprobability responses involve substantial self-selection into the panel. This self-selection can lead to a sample that is more interested in the topic of the survey than the target population (Chang and Krosnick, 2009). This, in turn, can lead to a less representative sample, biased on the key variables in the study, and without a mechanism to correct for the selection bias, the inferences from the sample are likely to be biased (Groves, et.al, 2000).

It is interesting to note that, while nonresponse in internet samples is substantial, it is out of concern for lower response rates in surveys using more representative samples that researchers seem to turn to these nonprobability internet panels (Curtin, et.al, 2005; Baker, et.al, 2010). Yeager and colleagues (2009) explored the issues of comparability of probability and nonprobability samples, using a RDD, a probability web sample, and nonprobability web samples. When a web survey is used to collect data from a randomly selected population, they are comparable to RDD surveys in terms of representativeness on demographic and attitudinal measures (ibid.). While weighting theoretically improves the accuracy of estimates from nonprobability samples, the seven nonprobability samples used by Yeager et.al, (2009) did not show consistent improvement in accuracy with weighting.

Groves (2006) demonstrated that, for probability samples, non-response does not necessarily indicate the presence of bias, and that it is possible to have non-response bias when overall non-response rates are low, i.e., the bias is present even though most of the people contacted participated in the survey. It is also possible to have low non-response bias when non-response rates are high. Concerns over nonresponse in probability samples are healthy in the discourse of survey research. Although some researchers may think nonprobability samples are an attractive solution to these issues, there are many problems associated with this practice:

> "Because of falling response rates, legitimate questions are arising anew about the relative advantages of probability sample surveys. Probability sampling offers measureable sampling errors and unbiased estimates when 100 percent response rates are obtained. There is no such guarantee with low response rate surveys. . . Unfortunately, the alternative research designs for descriptive statistics, most notably volunteer panels, quota samples from large compilations of personal data records, and so forth, require even more heroic assumptions to derive the unbiased survey estimates." (Groves, p. 670, 2006)

Further complicating the matter of nonresponse and bias, Yeager and his colleagues (2005) found that the completion rates of the nonprobability samples were negatively correlated with data accuracy. Hill and colleagues (2007) argue that the use of large, biased samples, such as those found in an internet opt-in poll, can be preferable to small, unbiased samples. The authors did find that a weighted internet study yielded closer estimates to a weighted in-person interview than an unweighted RDD study. However, the comparison of the unweighted data set to two weighted data sets on univariate measures, where weighting is most effective in making samples more representative of the target population, makes direct comparisons of the representativeness of internet surveys to the phone survey less reliable. Although more research is necessary in this area to really understand the impact of nonresponse and nonprobability samples on bias, these findings raise interesting questions about the reality of convenience samples and the ideals of large sample sizes and low nonresponse rates.

Regardless of the accuracy of claims that internet samples possess superiority due to larger sample sizes, there are many other sampling issues to consider. Groves (p. 668, 2006) gives four issues that are inherent to all sampling: the possible bias of a sample relative to the population, possible bias of the sample on the survey variables, bias of those measured on the survey variables, and the ability to fulfill assumptions required by adjustment procedures. For the first issue, having an explicit sampling frame will aide in comparing the sample to the population. On the second issue, random sampling can reduce selection bias, including self-selection into a study. On the fourth issue, auxiliary variables are necessary to decide what kind of adjustments, such as weighting may be appropriate. The use of post-adjustment techniques and rich sampling frames can compensate for nonprobability methods, but this would require devoting more of the survey's questions to auxiliary variables (ibid.). Furthermore, researchers must understand the population and the topic of interest in enough depth to include auxiliary variables capable of detecting and reducing nonrandom bias.

Nonprobability Samples & Weighting:

When researchers use nonprobability methods in quantitative research, post-data collection adjustment is crucial to attempting to rectify nonrandom bias. Unfortunately, weighting a nonprobability sample to adjust for demographic differences from the population does not always yield unbiased estimates for outcome variables. For example, Fricker, and colleagues (2005) found less racial and educational diversity amongst web respondents, compared to phone respondents, and the authors were able to track response rates and assert that web response rates were considerably lower than found in the telephone survey. Weighted data reduced the significance of gender, race, and Hispanic origin, it did not reduce the significant differences in education, age, and scientific knowledge scores and opinions; knowledge and favorable opinions of science were significantly higher in the web response group (ibid.). The failure of weights to reduce bias in nonprobability samples has also been noted by other researchers (Schonlau et.al, 2004; Malhorta & Krosnick, 2007; Chang and Krosnick, 2009; Yeager et.al, 2009). Berrens et. al. (2003), however, found that no significant differences remained in opinions about environmental issues collected in an RDD telephone survey, compared to surveys collected using different internet sampling methodologies, once demographic characteristics were appropriately weighted on the internet sample. These mixed results indicate that, while web surveys can be used to adjust the sample on key demographic variables, nonrandom bias may still be present.

Rivers and Bailey (2009) argue that nonprobability internet panels, which are selected by enrolling panel respondents until researchers have a nonprobability quota sample, have similar biases to RDD surveys and both need post-collection adjustment. Post-stratification weights are needed to increase the representativeness of the sample (ibid.). One type of post-stratification adjustment is the use of propensity score weighting. Propensity score weighting is commonly used to adjust nonprobability web samples for non-response (Bethlehem, 2010). The main adjustment factor used in propensity methods is bias resulting from internet access. This method usually involves running logistic regression with a sample that includes both internet users and nonusers to predict the probability, or propensity, of internet access in order to identify key covariates that could be used for adjustment (Schonlau et.al, p. 301, 2009). The inverse of these propensity scores are used as weights. In a population with differential internet access, if the correct auxiliary variables are known and the data are missing at random, then propensity score weighting should yield samples more representative of population characteristics, because it takes into account differences between those with and without internet access. The use of propensity weights carry a cost, however, as they increase the design effect and standard errors. A simulation study by Bethlehem (2010) showed that the use of small, representative samples to provide a comparison with non-internet users increased the, standard errors, shrinking the effective sample size. Schonlau and colleagues (2009) used propensity score weights to adjust for differential internet access in the Health and Retirement Survey, and found that the design effect rose from 1.4 to 6.7. Contrary to the claims of Rivers and Bailey (2009) that matching provides an added benefit to yield representative results with weighting techniques, Yeager and colleagues(2009) found that their estimates were not more representative when nonprobability samples recruited respondents with quota samples when compared to samples recruited through pop-up ads, also known as river sampling. River sampling is the internet equivalent to having interviewers go to a public location and approach people at the location to join the study. Since web surveys are often conducted because they cost less per participant, and therefore allow more people to be sampled, the presence of large design effects can negate any gain from having more respondents, greatly reducing the economic advantage of the web survey.

Given the many issues of nonprobability internet samples, it is not surprising that arguments are beginning to form against the claims of internet superiority. As far as keeping down costs, Fricker and Schonlau (2002) argued that a random sample would need to be screened by phone or mail. Even though RDD-generated sampling frames could still save on labor costs for hiring phone interviewers, these are tempered by the cost of designing, deploying, and maintaining a functioning and secured web-site for the survey (ibid.). Generating a sampling frame using internet methods, perhaps by assembling a complete list of e-mail addresses linked to geographic locations, is not feasible and there is no way to save on building a probability sample at this time. Issues of post-collection adjustment will add considerable time to the production of results, and it seems doubtful that these problems are taken into account when a timeline is considered. When using nonprobability data in quantitative analysis, it may be necessary to look for bias in the non-random sample and to adjust that bias with unique confounds, specific to the model of interest, which would require a laborious post-adjustment protocol and limit the cost-effectiveness of nonprobability samples.

Although there has been an increased amount of research studying how the mode of data collection and the use of nonprobability approaches affect the validity of the inference made from survey data, conflicting findings and the need for more research when studying relatively rare subgroups necessitate more research in these areas. This study will add to this literature by exploring difference between data collection modes for a relatively rare population. According to data from the 2008 GSS, evangelical Christians aged 18-29 make up less than 8% of the population. Although they are not a hidden population, such as the homeless, this is not a population with an available unbiased sampling frame, and efforts to collect representative data on a sufficiently large sample of these individuals would require an extensive RDD effort that would be incredibly costly. For example, collecting data on 400 young evangelicals would require screening interviews with over 5,000 respondents. Because web surveys may be more cost effective and nonprobability samples are commonly used to survey hard-to-reach populations, it is important to explore the consequences of study bias in the context of surveying a rare population by nonprobability internet methods.

Plan of Analysis

In order to test the main hypotheses in this study, I used two data sets divided into five samples. Table 1 shows how the breakdown of the datasets and samples, as well as the variables used in the each.

Table 1. Samples & Variables Used in this Study

Dataset	Religion & America's Role in the World (RARW)			General Social Survey-2008	
Sample	National RDD [note a]	Young Evangelical RDD	Young Evangelical Nonprobability	National Comparison	Young Evangelicals
n	967	133	300	3,332	227
Variables					
Demographic					
Age	X	X	X	X	X
Education	X	X	X	X	X
Employment Status	X	X	X	X	X
Family Income	X	X	X	X	X
Gender	X	X	X	X	X
Race	X	X	X	X	X
Region	X	X	X	X	X
Rural/Urban	X	X	X	-	-
Attitudinal					
Biblical Literalist	X	X	X	X	X
Church Attendance	X	X	X	X	X
Political Views	X	X	X	X	X
Outcome					
Abortion	X	X	X	-	-
Gays & Lesbians	X	X	X	-	-

Note a: This sample originally contained 33 young evangelicals, but those cases were used in the Young Evangelical RDD and are excluded from this total.

According to Greeburg Quinlan Rosner (p. 23, 2008), the nonprobability sample recruited for RARW was “. . . drawn from an opt-in web panel that is designed to be demographically representative at a national level.” This statement likely means that the investigators bought a sample from an existing internet panel. The internet sample should then match the demographics of the national population, and not to characteristics of young evangelicals, unless post-collection adjustments were made.

The probability sample of young evangelicals was recruited through "an age-predicted random digit dial process" (ibid.). Respondents were then given screener questions to ascertain their age and faith. Post-collection adjustments would be necessary with this sample, as well, in order to assure representativeness of demographic characteristics such as race or gender.

No young evangelical observations were used in either the RARW national sample or the GSS national comparison sample. This was done to use the most possible leverage in testing the differences between the young evangelicals and the national RDD in the RARW-only analyses. The small proportion of young evangelicals in the RARW sample, only 3.3%, made no difference in the descriptive statistics for the RARW national sample, however I carried this restriction into all univariate and multivariate analyses for consistency.

I began my analysis with univariate measures to compare the RARW young evangelicals to young evangelicals in the GSS 2008. I was able to find or create similar demographic measures, as well as two religiosity measures and a political measure. I also compared the nationally representative RARW to the full sample of the GSS 2008, to look at issues of non-representativeness. Next I ran univariate measures for the evangelicals with different weights applied, to look at the impact of different weighting strategies on representativeness. I compared the RARW national sample to the RARW young evangelicals on the two outcome variables: feelings towards gays and lesbians and attitudes toward the legality of abortion. Then I moved to multivariate analysis using only the RARW samples. In this analysis, I investigated whether there was a difference between the national RDD and the young evangelicals when the demographic, religiosity, and political measures are controlled for in the model. For each outcome, I ran an unweighted model and four weighted models, each with a different weighting strategy, and compared the results. Finally, I used the same multivariate analysis models to compare the two different RARW young evangelical samples to test for differences in these two approaches to surveying young evangelicals.

Hypotheses

1. The nonprobability sample differs from a representative sample of evangelicals.

Given the works of Fricker and colleagues (2005) and Berrens and colleagues (2003), I expect that the nonprobability web sample will differ from the young evangelical RDD. On demographic measures, this should include characteristics related to education, race, and gender, which are known to be related to non-response. For example, I expect to find that the nonprobability sample, being an internet sample, is better educated and has more white respondents than the young evangelical RDD. On attitudinal measures, I expect the nonprobability web sample to be more religious or hold more conservative views on the outcome variables.

The weights are designed to adjust for differences in demographic and social factors among the samples. Therefore, if the variables used to construct the weights, are not directly related to the outcome variable in an analysis model, the weight will not have an impact on the regression estimates in the model. For example, if demographic characteristics are not related to the outcome variables, but the religiosity measures are, then a combined “demographic and religiosity” weight will adjust for the bias when a “demographic” weight will not. If there is bias due to inequalities of internet access, then weights that adjust both for demographic variables and the propensity to have access to the internet weight will best reduce differences in outcomes between the samples.

2. Inferences about sample differences in outcomes made by using Greenburg Quinlan Rosner's (2008) original weight, created using a national population and not taking into consideration different demographic profiles of young evangelicals, will differ when weights are used which take into account the demographic profile of this group.

The use of weights designed for a national population on the evangelical subsample should have led to incorrect estimates on the demographics of the young evangelicals. Furthermore, if those demographics are also related to outcome variables (see hypothesis 1) then the outcome variables will also differ with better-specified weights. I will also test multivariate models, to see if controlling for variables in an unweighted model, as well as in the weighted models, made a difference in inferences. Of the two outcome variables, Greenburg Quinlan Rosner (pp. 20-21) in their report briefly mention differences between evangelical and other groups in abortion attitudes. Feelings towards gays and lesbians was not used in their report. My intent, however, is to use their samples and measures collected in their surveys to test research questions related to biases that may stem from use of nonprobability samples.

Variables

Sample Type Variables: Given the use of multiple samples, variables are necessary to distinguish between the two datasets (GSS and RARW) and the three sample types (national, young evangelical probability, and young evangelical nonprobability). In analyses where the RARW data set was compared to the GSS, the variable of interest was a dummy for the *GSS*. For both data sets, the dummy *yevang* was used to distinguish between young evangelicals and the national RDD's. For the RARW data set, the variable *non-random* was used to distinguish between the nonprobability web panel of young evangelicals and the RDD recruited sample of young evangelicals.

Demographics: Demographic information in the RARW that was comparable to the GSS included educational attainment, race, income, gender, age. In RARW, age is a continuous variable that originally measured the year the respondent was born, but was recoded into a variable for age by subtracting the year the survey was carried out by the birth year. No day or month of birth was available for a more exact estimate of the respondent's age. For the RARW data set, dummy variables for White, Black, Hispanic, and all other races were created; the dummy for White was always the reference variable. However, because the GSS only codes for White, Black, and all other races, only the dummy variables for Black and another race were used in analyses comparisons between the GSS and RARW.

Education was coded into a four category variable: less than high school, high school graduate, some college, and bachelor's degree or more. Gender is a dichotomous variable that uses male as the reference category. Income was measured categorically. Originally, the categories for RARW were less than $10,000; $10,000 to under $20,000; $20,000 to under $30,000; $30,000 to under $50,000; $50,000 to under $75,000; $75,000 to under $100,000, and $100,000 or more. The GSS was recoded to match most categories, but the highest two categories in both datasets were recoded into a category for $75,000 or more dollars.

For employment respondents could choose full-time employee, part-time employee, unemployed, retired, student homemaker and some other category. Due to the retirement dummy variable being inapplicable to young evangelicals, as well as having a small number of responses in the national samples, I collapsed retirement into the "other" category. Full-time employment is always used as the reference variable in the analyses.

Attitudinal Variables: Political views were measured by the question "Thinking in political terms, would you say that you are conservative, moderate or liberal?" with answer choices for liberal, moderate, and conservative. Political views were used instead of party identification because Black evangelicals may hold conservative views while supporting the Democratic party (Robinson 2006), and these outcome variables are largely opinion-based, and do not contain wording about civil rights or legislation.

There were two religiosity variables that occur in both the RARW and the GSS. How often the respondent attended religious services was originally measured as an ordinal variable, but was recoded into a dummy variable (attended services once a week or more =1; attended less than once a week=0). A second religiosity variable indicated whether or not the respondent self-identified as a biblical literalist (literalist = 1; other = 0).

Outcome Variables: Two outcome variables were used in the analysis that were available only in the RARW dataset. The first assesses abortion attitudes in the item: "Do you think abortions should be legal in all cases, legal in most cases, illegal in most cases, or illegal in all cases?" The answer choices were on four-point scale, ranging from 1, equaling always legal, to 4, equaling always illegal. The second outcome assesses feelings toward gays & lesbians and was asked in a thermometer question that included ratings of several other groups. Respondents were asked to rate their feelings toward the specified group on a scale of 0-100, with 100 being the most favorable and zero the least. These questions were part of a planned missing design in the RARW, with 50% of respondents randomly assigned to receive this question.

Methods

<u>The RARW Sample</u>

I used multiple imputation to handle the missing data and allow the retention of more cases in the analyses. The most common source of missing data was from the planned missing design but most variables had a small percentage missing. I used the ICE procedure in STATA, imputing 12 data sets. Multiple imputation yields less bias than other methods, such as mean substitution, and allows for the use of the full sample in the multivariate analyses (Acock 2005, Allison 2002). The effect of the uncertainty from the missing data is modeled when multiple imputation is used, making maximal use of the information available without biasing the standard errors and significance tests.

The 2008 GSS

The 2008 General Social Survey had enough similar questions to those in the RARW and a sufficient sample size that I could create a subsample of the GSS containing just young evangelicals. The GSS evangelical was sample was selected by using the religious affiliation items in the GSS to match the Greenburg Quinlan Rosner (p. 3, 2008) definition of evangelicals: ". . . Protestants or members of another Christian religion who identify as fundamentalist, evangelical, charismatic, or Pentecostal, or who indicated they were born-again Christians." The age range was also restricted to respondents 18-29 years old, to match the RARW samples.

The GSS is a multi-stage probability sample, with respondent's selected from Standard Metropolitan Statistical Areas or counties (Smith, et.al, 2011). The sample is also stratified by region, age, and race, to ensure that the data is nationally representative of non-institutionalized adults, but with a lower cost than a simple random sample, which is necessary as the GSS uses in-home interviews. In 2008, the GSS had a panel component of 2,006 re-interviewed persons in addition to 2,023 new cases (Smith, et.al, 2011). In order to use as large a sample of young evangelicals as possible, I kept the panel cases in the sample, and included the weights provided to adjust for both non-response and panel attrition in all analyses. This ensures that my sample is still representative of young evangelicals in 2008.

Creation of Post-Stratification Weights

Adjusting for Demographic and Religiosity Variables

An important part of this study is the comparison of the conclusions drawn by probability and nonprobability samples after using weights to adjust for differences among the samples in demographic, religiosity, and internet use variables. I used the distributions on demographic and religiosity variables in the GSS to serve as the population values used to develop the post-stratification weights. Use of the GSS for estimates of the population characteristics of young evangelicals was necessary because the Census Bureau does not collect information on religion, and evangelicals may not be demographically comparable to the nation. The GSS is a representative probability sample and has a strong reputation as being a high quality data set, and so it is a good choice for estimating the population characteristics of this group.

A procedure called “raking” was used to create the weights. This procedure begins with the percentage distribution in each category of each of the variables used in the weights. It requires that all variables be divided into a relatively small number of categories, particularly when the sample sizes in the population estimates are relatively small (Battaglia et. al., 2009). The four variables used in the demographic weight were education, race, gender and region. The distribution into the response categories on these variables in the GSS was treated as the population and the distribution in the RARW was treated as the sample. Weights are created in a stepwise, iterative manner. Beginning with education, I took the percentages in I expected to find, given the GSS, and divided those by the proportion I actually found in RARW. I started with education, then race, then gender, and calculated region last. Once I calculated the education weight, I used it to weight the data, and then I generated the proportions for the race variable weighted by education. I next created a weight for race, and then multiplied by the education weight. I continued this same procedure for the rest of the demographic variables.

Because including variables later in the sequence may alter the weighted distributions of the earlier variables, additional iterations are often required to yield the same distributions on these variables in the population and sample. In this case two iterations were needed to produce similar proportions to the population for all the demographic variables.

I constructed a second set of weights that added the two religiosity variables: attended church weekly and biblical literalist. These two variables were appended to the first round of weighting, and then the second round of weights was recalculated. These weights were created to test the hypothesis set out by Groves (2006) that adjusting for certain attitudinal measures that may lead to self-selection could deal with error from the nonprobability selection method. For this weight, four iterations were needed to produce distributions similar to the GSS.

Propensity Score Weight

The young evangelical web sample was restricted to persons with internet access, and this could lead to selection bias. The standard way to adjust for this bias is to use propensity score weighting (Bethlehem, 2010; Schonlau et.al, 2009). A reference sample (in this case, the GSS) that contains similar questions to the web sample, but also includes persons without internet access, is used to estimate the propensity (likelihood) that a sample member with a given set of demographic characteristics has access to the internet. These propensities are then used to weigh each respondent by the inverse of their propensity to have internet access. For example, if a person is found to have a probability of .5 of having internet access, then they will be weighted by the inverse of this probability, in this case a weight of 2.0. This adjusts for the underrepresentation of persons with these characteristics in the web survey.

The internet propensity score weights were constructed by running a logistic regression on the 2008 GSS young evangelicals to predict their access to internet at home using demographic and religiosity measures also available in the RARW. The final equation only includes the variables that were significant or substantively important, such as race. The equation used to create the propensity of internet access was:

$$\hat{y} = a + b * age + b * education + b * south + b * part\ time + b * black + b * other\ race + b * family\ income$$

where $\hat{y}$ equals the log odds of having internet access. I exponentiated this to obtain the predicted odds, and converted that to a probability, or a propensity, for having internet access for each web respondent. This variable was multiplied by the demographic weight to create a weight that combines the demographic and internet propensity score weight.

Table 2, on the previous page, shows the percentage of GSS young evangelicals who have internet access by key demographic characteristics. Those who are wealthier, better educated, or were students were more likely to have internet access. Although the regional percentages may seem odd, there are a couple of possible explanations for why those in the Northeastern and Midwestern U.S. were less likely to have internet access relative to those in the South and West. As Table 4 (page 25) shows, those living in the Northeast and Midwest comprise less than a fifth of young evangelicals. Not only do most evangelicals live in the South and West, most of the white evangelicals tend to live in these regions. Further analyses (not shown) verified that respondents in the Northeast were most likely to be students, and mean family income was similar across the regions. Taken together, the increased number of Whites in the South and West seems to be the best explanation for the higher probability of internet access in these regions.

Table 2. Percentage of General Social Survey (2008) Respondents with internet access for Key Demographic Characteristics.

Family Income, in thousands of dollars		Work Status		Education		Region		Race		Age		Gender	
<10	65%	Full-time	74%	< HS	42%	NE	55%	White	80%	18-21	82%	Male	77%
10-<20	68%	Part-time	69%	HS Graduate	72%	MW	58%	Black	71%	22-25	73%	Female	73%
20-<30	66%	Student	85%	Some College	86%	S	77%	Other	56%	26-29	71%		
30-<50	66%	Keeping house	78%	Bachelor's	83%	W	81%						
50-<75	81%	Unemployed /other	70%										
75+	94%												

Table 3. Percentage of GSS (2008) Respondents with internet access for political and religiosity variables

Political Views		Attends Church Weekly or more?		Biblical Literalist	
Conservative	80%	No	74%	No	83%
Moderate	75%	Yes	76%	Yes	65%
Liberal	70%				

Table 3, on the previous page, shows the relationship between internet access and political and religiosity measures. Internet access appears unrelated to church attendance, but does differ by whether or not the respondent views the Bible as the literal word of God. There were also differences in access amongst those with different political views, with conservative evangelicals having higher rates of access than politically liberal evangelicals. This may also be due to racial differences, as Whites, who have higher levels of internet access than the other racial groups were more likely to be conservative than liberal, while African-Americans were most likely to be moderate, and slightly more likely than Whites to self-identify as liberals.

Multivariate Analyses

Once weights were constructed, two substantive models, one examining abortion attitudes and another regarding feelings toward gays & lesbians, were chosen to look at the impact of weighting. I looked at regression models for the national sample versus the young evangelicals, to test the hypothesis that weighting would not result in the replication of the Greenburg Quinlan Rosner results. To test the hypothesis that the samples are different, and that different weights would impact results, I ran regressions using only the young evangelical samples. I ran regressions without weights, with the demographic weight, and the combined demographic and attitudinal weight.

Results

First, univariate analyses were used to test the hypothesis that the nonprobability web sample of young evangelicals differs from the probability samples of young evangelicals in the RARW and the GSS (hypothesis 1). I upheld this finding, but with a surprising result. While the nonprobability sample differed from the GSS young evangelicals in many respects, they were slightly closer to the GSS estimates than the RARW probability sample of young evangelicals.

For the second hypothesis, that the inferences from the outcome variables would change when I used the weights I constructed, instead of Greenburg Quinlan Rosner's, I used both bivariate and multivariate tests. Acceptance or rejection of this hypothesis rested on multiple factors, including the outcome variable used and the type of analysis.

Abortion

When testing the difference between RARW young evangelicals and the national RDD, I did not find a difference in inferences using bivariate abortion variable models. Young evangelicals were always significantly less supportive of legalized abortion than the national RDD respondents. However, when I used OLS regression, I did find that using the GQR weight yielded a multivariate model that failed to detect a significant relationship between the national RDD and the young evangelicals. Furthermore, when I used OLS regression to test the difference between the young evangelical samples in the RARW, excluding the national RDD sample, I found that using the GQR weight led to the conclusion that the two samples were significantly different. This finding was not replicated with the other weights. Therefore, I would reject my second research hypothesis in the bivariate models, but not in the multivariate models.

Gays &Lesbians

In the bivariate measures, I found that using the GQR weight led to the conclusion that young evangelicals do not hold significantly different attitudes about gays and lesbians when compared to the national RDD. In bivariate models using the other weights, or not using a weight, I did find a significant difference. When I used OLS regression to test for a difference between the young evangelicals and the national RDD, I did not find a significant difference in any of the models. When I tested the young evangelical-only model, I was able to adjust for non-random bias using either the GQR weight or the demographic and attitudinal weight. There are differences in the coefficients of the attitudinal independent variables that provide evidence that the demographic and attitudinal weight better adjusts the overall model. I have evidence that upholds my second research hypothesis for the both the bivariate and multivariate models with this dependent variable.

Unweighted Univariate Statistics

Table 4. Unweighted Percentages for the GSS & RARW Young Evangelical Over-samples.

	GSS-2008	RARW-Young Evangelicals	RARW-Nonprobability	RARW-RDD
Education				
LT HS	16.7%	3.3%	3.7%	2.3%
HS	29.5%	19.8%	17.3%	25.0%
SC	39.2%	39.5%	42.0%	33.5%
BA/S+	14.5%	37.5%	37.0%	39.2%
Race				
White	60.4%	75.7%	76.3%	74.4%
Black	30.4%	10.7%	11.3%	9.3%
Other	9.3%	13.6%	12.4%	16.3%
Gender				
Male	41.4%	52.9%	52.3%	54.1%
Female	58.6%	47.1%	47.7%	45.9%
Region				
NE	5.7%	8.5%	8.7%	8.3%
MW	14.5%	25.4%	21.7%	33.8%
S	52.4%	51.0%	54.7%	42.9%
W	27.3%	15.0%	15.0%	15.0%
Attend Church Weekly or more?				
No	67.7%	44.6%	49.7%	33.1%
Yes	32.3%	55.4%	50.3%	66.9%
Biblical Literalist				
No	47.1%	48.7%	51.0%	43.5%
Yes	52.9%	51.3%	49.0%	56.5%
Political Views				
Conservative	38.0%	42.8%	39.7%	50.0%
Moderate	40.3%	40.3%	43.7%	32.6%
Liberal	21.7%	16.9%	16.7%	17.4%
Family Income, in thousands of dollars				
<10	15.8%	7.1%	7.7%	5.8%
10-<20	18.6%	11.4%	13.0%	7.7%
20-<30	14.4%	13.7%	12.7%	16.2%
30-<50	20.6%	24.2%	25.0%	22.3%
50-<75	18.7%	23.0%	23.3%	22.2%
75+	12.0%	20.6%	18.3%	25.8%
Employment Status				
Full-time	53.2%	51.7%	48.3%	59.4%
Part-time	15.9%	12.5%	11.3%	15.0%
Student	8.6%	18.2%	21.7%	10.5%
Keeping house	13.6%	12.0%	13.3%	9.0%
Unemployed/other	8.6%	5.5%	5.3%	6.0%
Age, Categorized				
18-21	18.5%	18.9%	20.0%	16.5%
22-25	33.0%	32.6%	34.7%	27.8%
26-29	48.5%	48.5%	45.3%	55.6%
N=	227	433	300	133

Table 4 shows the differences in the unweighted proportions of the GSS and the RARW samples. The RARW samples are further divided into the nonprobability (web) sample and the RDD sample. As the GSS conducts in-person interviews, and employs a multi-stage cluster strategy, the unweighted estimates should differ to some degree, given the mode effects. Since the random digit dial sample was designed to be nationally representative, I expected this sample to be closer to the GSS estimates, and yet, the RDD was furthest from the GSS estimates.

The nonprobability sample differed from the GSS on many demographic characteristics—family income, race, and education—in support of the first hypothesis. That the RARW probability sample was the least representative may be due to the small sample size, as 133 respondents may not yield the variance necessary to capture people who have relatively rare characteristics, such as not having graduated high school, amongst a relatively rare population-i.e., being a young evangelical. Looking at race, 30% of GSS young evangelicals are black, and less than 10% of GSS young evangelicals are of another race. For the RARW RDD, the proportion of African-Americans was a third of what it should have been, while the proportion of respondents who did not identify as either white or black was nearly twice what I would expect, given the GSS estimates. It is difficult to draw any conclusions without knowing details about the RDD sampling methodology, but the problems in the nonprobability study can be explained. The full sample of the GSS has 77% White respondents, 14% Black respondents, and about 8.5% all other races. For young evangelicals, the GSS had 60% White, 30% Black, and about 10% of young evangelicals belonged to some other race. However, because Greenburg Quinlan Rosner (2008) matched their non-random sample on demographic characteristics, their estimates are closer to the national estimates. The RARW young evangelical web sample was 76% White, 11% Black, and 13% of respondents belonged to another race. The young evangelical RDD was 79% White and 9% Black perhaps due to the fact that the GSS cluster samples better stratified for race than the RDD, but the small sample size may also be a factor.

For the political and religiosity measures, skew is again greatest amongst the RDD sample. Weekly church attendance was considerably higher amongst nonprobability sample respondents (50% versus 32% for the GSS), but the other attitudinal measures are much closer than I expected, given past research that nonprobability samples are biased on measures central to the study's objective (Chang & Krosnick, 2009; Groves, 2006). While I uphold the hypothesis that the nonprobability sample is different from the GSS young evangelicals, I unexpectedly found that the RARW probability sample is not representative on several characteristics.

The inconsistencies between the RARW and GSS probability samples are due to the sampling strategies. Greenburg Quinlan Rosner (2008) conducted an age-targeted RDD, although they gave no specific information on how a sampling frame was constructed. One possibility would be an oversample of geographical areas which are known to have higher proportions of young adults. Of course, only the use of screening questions would ascertain religious affiliation. Given the nature of Census design, this would lead to oversamples in areas where there are colleges and universities, urban areas, and possibly even oversamples of areas where military recruits live in the barracks. The use of a college or university would explain the distribution of respondent education, while the use of urban areas, where young professionals may cluster, would explain the age distribution. This type of oversample led to non-representation on a number of variables, and post-data collection adjustment was necessary for both the young evangelical samples in the RARW dataset.

Table 5. Two Nationally Representative Samples, excluding the RARW Evangelicals Oversample.

Education	GSS	RARW
LT HS	13.6%	6.5%
HS	29.7%	32.7%
SC	28.6%	27.3%
BA/S+	28.0%	33.6%
Race		
White	77.3%	72.5%
Black	12.9%	11.2%
Other	9.8%	16.3%
Gender		
Male	46.1%	51.0%
Female	53.9%	49.0%
Region		
NE	17.0%	19.3%
MW	20.6%	22.2%
S	37.4%	35.5%
W	25.0%	23.0%
Attend Church Weekly or more?		
No	72.3%	61.3%
Yes	27.7%	38.7%
Biblical Literalist		
No	68.0%	55.1%
Yes	32.0%	44.9%
Political Views		
Conservative	34.5%	41.5%
Moderate	38.2%	36.7%
Liberal	27.4%	21.9%
Family Income, in thousands of dollars		
<10	6.9%	7.1%
10-<20	10.2%	9.8%
20-<30	10.5%	11.1%
30-<50	17.7%	19.0%
50-<75	21.2%	20.7%
75+	33.5%	32.2%
Employment Status		
Full-time	50.1%	47.0%
Part-time	11.2%	9.8%
Retired	14.1%	22.3%
Student	3.3%	3.5%
Keeping house	12.3%	8.5%
Unemployed/other	9.1%	8.8%
Age		
18-29	21.0%	15.2%
30-39	18.0%	19.9%
40-49	19.5%	19.8%
50-59	18.2%	17.6%
60-69	12.1%	16.3%
70-79	7.3%	8.2%
80 & older	3.7%	3.0%
N=	3,332	967

Differences in the proportion of the variables for the national sample comparison (see Table 5) are also apparent, although generally less severe than those found amongst the young evangelicals. While the young evangelical samples were under-representative of Blacks, the national samples were over-representative of the other race category. This may be due to measurement error—the RARW interviewers asked one question regarding race and one focused on Hispanic origin, while the GSS uses one question for race and the categories White, Black, and other.

There was a bias toward conservatives, frequent church attenders, and biblical literalists in RARW. It was not as severe, however, as the skew apparent in the young evangelical samples. Taking into account sample size, as the GSS has more than three times as many respondents as the RARW, and the fact that these are unweighted percentages of surveys with different modes, these findings seem reasonable.

Weighted Univariate Statistics

Table 6. Weighted Percentages for the GSS & RARW Samples—Evangelicals Only.

	GSS 2008	RARW			
	Non-Response Weight, "wtnrcomb"	GQR weight	Demographic Only Weight	Demographic & Internet Propensity Score Weight	Demographic & Attitudinal Weight
Education					
LT HS	14.9%	3.0%	14.8%	20.0%	14.9%
HS	33.4%	29.7%	33.6%	32.9%	33.4%
SC	37.9%	38.7%	37.9%	35.3%	37.9%
BA/S+	13.8%	28.6%	13.7%	11.7%	13.8%
Race					
White	61.1%	60.0%	61.0%	61.9%	61.1%
Black	27.5%	10.5%	27.5%	25.2%	27.5%
Other	11.4%	29.6%	11.5%	12.9%	11.4%
Gender					
Male	46.1%	47.0%	45.9%	50.0%	46.1%
Female	53.9%	53.0%	54.1%	50.0%	53.9%
Region					
NE	5.6%	7.3%	5.6%	5.7%	5.6%
MW	12.1%	16.7%	12.1%	11.9%	12.1%
S	49.1%	50.8%	49.1%	50.8%	49.0%
W	33.2%	25.2%	33.2%	31.6%	33.2%
Attend Church Weekly or more?					
No	67.6%	40.8%	45.5%	48.2%	67.6%
Yes	32.4%	59.2%	54.5%	51.8%	32.4%
Biblical Literalist					
No	54.0%	40.6%	45.0%	44.1%	54.0%
Yes	46.0%	59.4%	55.0%	55.9%	46.0%
Political Views					
Conservative	40.6%	44.2%	40.5%	42.3%	37.2%
Moderate	35.3%	38.8%	36.9%	36.3%	36.5%
Liberal	24.1%	16.9%	22.6%	21.5%	26.3%
Family Income, in thousands of dollars					
<10	13.1%	12.8%	12.1%	14.8%	12.4%
10-<20	18.5%	16.6%	12.8%	14.3%	13.5%
20-<30	11.1%	15.2%	14.0%	14.1%	13.3%
30-<50	17.6%	18.0%	23.2%	21.9%	23.3%
50-<75	20.7%	22.8%	16.9%	16.5%	18.2%
75+	19.1%	14.6%	21.1%	18.4%	19.2%
Employment Status					
Full-time	44.1%	46.3%	40.3%	39.4%	40.4%
Part-time	22.9%	17.6%	15.2%	17.6%	14.3%
Student	11.7%	18.2%	22.7%	20.8%	21.5%
Keeping house	10.4%	9.0%	14.9%	15.7%	16.6%
Unemployed/other	10.9%	8.9%	6.9%	6.5%	7.2%
Age, Categorized					
18-21	30.2%	20.2%	27.6%	24.1%	26.1%
22-25	32.4%	29.1%	33.5%	36.3%	33.5%
26-29	37.5%	50.7%	38.9%	39.6%	40.4%

The second hypothesis states that inferences made using weights specific to the young evangelicals differ from inferences made using the Greenburg Quinlan Rosner (GQR) weight. As shown in Table 6, the internet propensity score weight tends to yield estimates furthest from the

GSS. This is because the inverse probability weights are built to converge with the GSS estimates, while the internet propensity score weight was built to correct for differential internet access. For those variables that were not used in the construction of the weights, weighting did not necessarily make these estimates more representative. The demographic and internet propensity weight did bring family income slightly closer to the GSS estimates, as it up-weighted those with family incomes of under ten thousand dollars, although other categories were actually further from the GSS estimates compared to the demographic only weight. Yet, the mean of family income for the demographic and internet weight was 3.689, which is comparable to the GSS mean of 3.6889. The mean for the RARW sample with only the demographic weight was 3.87, though all are preferable to the unweighted RARW mean, which was 4.07.

The GQR weight was similar to the GSS on gender, and also for the proportion White. The proportion Black, however, was still about 1/3rd of the GSS estimates, and the other race category is three times higher. In 2008, the American Community Survey (Census Bureau, 2009) estimated the population was 66% non-Hispanic White, 12% non-Hispanic Black, and 22% would belong to the other race category. The GQR weight is close to these estimates, and I believe that, in addition to matching the web sample on demographics (Greenburg Quinlan Rosner, 2008) also constructed a weight on the characteristics of the national population. Overall, the use of the GQR would lead to different inferences about the characteristics of evangelicals than the use of the weights specific to evangelicals, in accordance with the second hypotheses.

Bivariate Analysis

Table 7. Descriptive Statistics : Support for Legality of Abortion and Feelings towards Gays & Lesbians for the National RDD and both the probability and nonprobability young evangelicals (YE), Religion & America's Role in the World only

	National RDD	All YE	Non-probability YE	RDD YE	Demographic Only Weight	Demographic & Internet Propensity Score Weight	Demographic & Religiosity Weight
Support for the Illegality of Abortion							
Legal in all cases	19.7%	9.5%	9.3%	9.9%	8.9%	9.7%	9.5%
Legal in most cases	31.0%	21.6%	21.0%	22.9%	19.9%	21.0%	25.6%
Illegal in most cases	31.8%	45.2%	45.7%	44.3%	41.1%	39.4%	39.7%
Illegal in all cases	12.8%	23.7%	24.0%	22.9%	30.1%	30.0%	25.2%
Feelings Toward Gays & Lesbians- 0= Unfavorable, 100= Favorable							
Mean	49.32	37.67	34.33	45.2	35.97	34.34	36.72
Standard Error	(.3608)	(.4585)	(.5288)	(.8699)	(.7618)	(.8124)	(.806)
Feelings Toward Gays & Lesbians, in quintiles							
Unfavorable	25.0%	37.5%	41.2%	29.4%	40.2%	42.6%	39.9%
	8.4%	10.2%	9.7%	11.4%	10.2%	9.7%	9.6%
Neutral	30.2%	26.8%	27.2%	26.0%	22.8%	22.3%	23.3%
	11.6%	9.3%	9.0%	9.9%	12.3%	11.7%	11.9%
Favorable	24.9%	16.1%	12.9%	23.4%	14.4%	13.6%	15.3%

In their analysis of support of the legality of abortion, Greenburg Quinlan Rosner (2008) found that young evangelicals were more likely to support making abortion illegal in all circumstances when compared to the nationally representative random digit dial sample. Table 7, on the previous page, shows the impact of weighting on the proportions. The demographic only weight and the demographic and internet propensity score weight actually increase the differences in the category for "illegal in all cases." The demographic and religiosity weight yielded estimates closer to the unweighted young evangelical estimates in the most extreme categories. Some responses are shifted in the middle two categories, with a slightly higher percentage of evangelicals believing abortion should be legal in most cases. Unweighted, 31% of young evangelicals believed abortion should be legal. The demographic weight yields an estimation of nearly 29%, and the demographic and religiosity weight yields an estimation of just over 35%. Weighting never changed the significantly lower support for legal abortion among young evangelicals, in contrast to my second hypothesis.

The question about feelings towards gays and lesbians was not analyzed by Greenburg Quinlan Rosner. However, this is another polarizing social issue, in which one would expect young evangelicals to differ from the national population, holding more socially conservative views. This expectation holds, unless the GQR weight is used. The RDD sample of young evangelicals had a distribution closest to the national sample. I suspect that this is largely due to social desirability bias. The quintiles show that the web sample held the least favorable views. When weights were applied, responses tended to become slightly more extreme, and those in the neutral category dissipated from nearly 27% to 23% or less. For this dependent variable, I found evidence that my conclusions differed if I used the GQR weight, in accordance with the second hypothesis.

Multivariate Analysis

Greenberg Quinlan Rosner did not publish any multivariate models based on this dataset. However, given the sophisticated tools available for looking at research questions, I decided to use OLS regression to further explore the differences between both the young evangelical and the national samples, as well as the between the nonprobability and probability samples.

National RDD and Young Evangelical Comparison

Table 8, on the next page, shows a series of multiple regressions for the dependent variable concerning feelings toward gays and lesbians. Although I found significant differences in the bivariate models, the dummy variable denoting whether or not the respondent was a young evangelical is not significant in any OLS model. Even when the data are not weighted, young evangelical attitudes do not significantly differ from the national population, once demographic and religiosity variables are used as controls. The only significant predictors were education, conservatism, attending church weekly or more, and being a biblical literalist. Being a biblical literalist was not significant when I used the GQR weight, which is likely due to the fact that evangelicals are not properly weighted in this model.

Table 8. Ordinary Least Squares Regression of full RARW sample—DV: Feelings Towards Gays & Lesbians (0=Unfavorable, 100=Favorable)

	Unweighted	GQR weight	Demographic Weight	Demographic & Internet Access Weight	Demographic & Religiosity Weight
	b (se)	b (se)	b (se)	b (se)	b (se)
R is a Young Evangelical	-6.67	-8.23	-9.66	-10.39	-10.72
	(3.85)	(6.10)	(5.30)	(5.66)	(5.54)
Race/Ethnicity (White is Reference)					
Black	-6.46	-12.38	-9.50	-8.73	-9.88
	(4.46)	(6.69)	(6.01)	(6.06)	(6.28)
Hispanic	-2.45	-2.14	-3.59	-2.45	-4.11
	(4.30)	(7.45)	(7.07)	(7.02)	(7.17)
All other races	10.68	13.21	13.29	13.26	13.83
	(7.77)	(8.65)	(8.93)	(8.87)	(9.01)
Education	5.48***	5.25**	5.68**	5.81**	5.83**
	(1.56)	(1.87)	(1.79)	(1.82)	(1.76)
Family Income	-0.16	0.15	-0.25	-0.34	-0.48
	(1.06)	(1.56)	(1.35)	(1.31)	(1.37)
Female	-4.64	-4.74	-4.23	-4.38	-4.16
	(2.65)	(3.58)	(3.37)	(3.30)	(3.44)
Rural	-6.52*	-5.07	-6.13	-6.71*	-6.16
	(2.94)	(3.52)	(3.35)	(3.38)	(3.40)
R is from the South	-1.80	-0.06	-0.07	-0.21	-0.30
	(2.09)	(3.37)	(2.87)	(2.87)	(2.84)
R is a Conservative	-20.08***	-18.72***	-18.05***	-17.87***	-18.09***
	(2.63)	(3.54)	(3.24)	(3.19)	(3.30)
R Attends Church Weekly or More	-6.18**	-7.28*	-6.78*	-6.19*	-6.81*
	(2.28)	(3.20)	(3.03)	(3.02)	(3.09)
R is a Biblical Literalist	-5.60**	-4.71	-5.38*	-5.69*	-5.53*
	(2.32)	(2.79)	(2.69)	(2.73)	(2.70)

n=1,394
* p-values < .05, ** p-values < .01, *** p-values < .001
Note: Also controlling for age and dummy variables for employment status—part-time, student, homemaker, and unemployed/other with full-time worker as the reference.

In both the unweighted model and the model using demographic and internet access weight, living in a rural area was a significant predictor of lower favorability toward gay and lesbian persons. Models using the demographic weight and demographic and attitudinal weight have marginally significant (p-value<.1) coefficients. As the GSS does not include a variable for currently residing in metropolitan or rural area, I did not account for this in the internet access weight. People in the rural United States are less likely to have internet access at home, as are women, minorities, and those with lower incomes (Strover, 1999; Wilson et.al, 2003); it is possible that the internet access weight is capturing this relationship since the equation included terms for race and income. The best adjusted model was the internet propensity score weighted OLS. I support my second hypothesis that my conclusions would have changed had I used the GQR weight. Had I used an unweighted model, however, my effect sizes and significance tests would be similar to the internet access weighted model.

While the bivariate and multivariate models for feelings towards gays and lesbians, showed differential support for the differences between young evangelicals and the national sample, findings for attitudes about the legality of abortion were robust. As shown in Table 9, on the next page, the coefficient for being a young evangelical was significant, unless the GQR weight was used. Conservatism, church attendances, and being a biblical literalist are significant predictors. The dummy coefficient for living in a rural area is not significant in the unweighted and marginally significant ($p<.1$) in the internet access weighted models. The higher coefficient for biblical literalists may have eroded the rural effect, as the primary determinants of abortion attitudes are religiosity (Himmelstein, 1986).

Table 9. Ordinary Least Squares Regression of full RARW sample—DV: Attitude towards Legality of Abortion (1=Almost always legal, 4= Almost always illegal)

	Unweighted b (se)	GQR weight b (se)	Demographic Weight b (se)	Demographic & Internet Access Weight b (se)	Demographic & Religiosity Weight b (se)
R is a Young Evangelical	0.32***	0.12	0.33**	0.29*	0.34**
	(0.07)	(0.12)	(0.12)	(0.13)	(0.12)
Race/Ethnicity (White is Reference)					
Black	-0.10	-0.02	-0.06	-0.07	-0.08
	(0.08)	(0.11)	(0.10)	(0.10)	(0.10)
Hispanic	-0.25*	-0.14	-0.17	-0.24	-0.15
	(0.10)	(0.15)	(0.15)	(0.16)	(0.15)
All other races	0.01	0.11	0.13	0.13	0.14
	(0.13)	(0.18)	(0.18)	(0.18)	(0.18)
Education	-0.06	-0.04	-0.07	-0.06	-0.07
	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)
Family Income	-0.01	-0.02	-0.01	-0.01	-0.02
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Female	0.10*	0.08	0.11	0.12	0.12
	(0.05)	(0.07)	(0.06)	(0.07)	(0.07)
Rural	0.07	0.15*	0.14*	0.14	0.14*
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
R is from the South	0.05	0.08	0.09	0.08	0.11
	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)
R is a Conservative	0.39***	0.40***	0.38***	0.36***	0.37***
	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)
R Attends Church Weekly or More	0.52***	0.48***	0.49***	0.48***	0.49***
	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)
R is a Biblical Literalist	0.13**	0.13*	0.14*	0.15*	0.14*
	(0.05)	(0.07)	(0.06)	(0.06)	(0.06)

n=1,394

* p-values < .05, ** p-values < .01, *** p-values < .001

Note: Also controlling for age and dummy variables for employment status—part-time, student, homemaker, and unemployed/other with full-time worker as the reference.

Table 10. Ordinary Least Squares Regression of full RARW sample, showing the young evangelicals broken into the probability and nonprobability samples (RARW national RDD is the reference)

	Unweighted b (se)	GQR weight b (se)	Demographic Weight b (se)	Demographic & Internet Access Weight b (se)	Demographic & Religiosity Weight b (se)
DV: Feelings Toward Gays & Lesbians					
Young Evangelical RDD	3.28	-0.98	-3.93	-3.89	-5.79
	(5.41)	(6.71)	(9.38)	(9.44)	(10.63)
Nonprobability Sample	-12.60**	-13.85**	-12.28*	-12.39*	-13.05*
	(3.76)	(4.89)	(5.69)	(6.21)	(5.77)
DV: Attitudes Toward Illegality of Abortion					
Young Evangelical RDD	0.23*	0.24	0.49**	0.49**	0.55**
	(0.10)	(0.13)	(0.16)	(0.16)	(0.19)
Nonprobability Sample	0.37***	0.32**	0.26*	0.22	0.30*
	(0.08)	(0.11)	(0.13)	(0.14)	(0.13)

n=1,394
* p-values < .05, ** p-values < .01, *** p-values < .001
Note: Also controlling for age, race, gender, family income, living in a rural area, living in the south, being a conservative, attending church weekly or more, being a Biblical literalist, and dummy variables for employment status—part-time, student, homemaker, and unemployed/other with full-time worker as the reference.

While it is important to note that young evangelicals are different from the national population on the abortion measure, as Greenburg Quinlan Rosner stated, it is also instructive to break the young evangelical samples out into probability and nonprobability samples. Table 10 shows the regressions coefficients broken into these groups (full tables available by request). The non-significant difference for the RDD young evangelicals and the significant differences for nonprobability sampled young evangelicals provides support for the conclusion that a mode effect impacted estimates of this dependent variable.

For the abortion measure, both groups significantly differ from the national sample, except in the case of the GQR weight and in the case of the internet access weight. Given how robust the results were for the differences in young evangelical support of abortion, these findings are contradictory to expectations. This also contradicts previous literature that upholds the difference between evangelicals and non-evangelicals opinions about abortion (Evans, 2002;

Wilcox, 1992). In support of the second hypothesis, the demographic and demographic and religiosity weights do yield more accurate inferences than the GQR weight, although the internet propensity weight does not.

Probability and Nonprobability Young Evangelicals Comparison

Table 11. Ordinary Least Squares Regression of Young Evangelical-only Samples—DV: Feelings Towards Gays & Lesbians (0=Unfavorable, 100=Favorable)

	Unweighted b (se)	GQR weight b (se)	Demographic Weight b (se)	Demographic & Internet Access Weight b (se)	Demographic & Religiosity Weight b (se)
Nonprobability Sample	-16.20*	-11.79	-13.86*	-14.17*	-12.54
	(6.41)	(7.36)	(6.66)	(7.16)	(6.81)
Race/Ethnicity (White is Reference)					
Black	2.00	4.54	2.67	4.15	0.03
	(8.92)	(9.14)	(10.59)	(10.53)	(9.28)
Hispanic	-3.24	6.41	-4.18	1.98	-7.65
	(9.82)	(11.51)	(13.46)	(14.14)	(12.37)
All other races	-6.34	-1.29	-11.74	-12.27	-9.31
	(12.49)	(21.52)	(19.21)	(16.84)	(17.91)
Education	4.86	6.70	3.54	3.17	4.82
	(2.89)	(4.17)	(4.24)	(4.64)	(4.50)
Family Income	-1.88	-0.39	-2.64	-2.82	-3.95
	(1.71)	(2.40)	(2.39)	(2.37)	(2.35)
Female	-9.15	-8.13	-8.09	-9.83	-8.65
	(6.14)	(6.99)	(7.62)	(7.50)	(7.78)
R is from the South	-4.10	-9.80	-5.86	-5.15	-6.87
	(5.86)	(7.37)	(7.47)	(7.19)	(7.27)
R is a Conservative	-17.90**	-17.51*	-15.75	-15.25	-17.34
	(6.37)	(7.54)	(8.24)	(8.43)	(8.48)
R Attends Church Weekly or More	-3.25	3.87	2.63	3.47	3.15
	(4.93)	(6.10)	(6.24)	(6.78)	(7.06)
R is a Biblical Literalist	-10.90*	-8.87	-13.27*	-13.34*	-13.13*
	(5.08)	(6.96)	(6.20)	(6.60)	(6.40)

n= 432

* p-values < .05, ** p-values < .01, *** p-values < .001

Note: Also controlling for age, living in a rural area, and dummy variables for employment status—part-time, student, homemaker, and unemployed/other with full-time worker as the reference.

I restricted the sample to look at only the young evangelicals in order to analyze whether non-random bias influences interpretations on the outcome variables, in accordance with the second hypothesis. For the feelings towards gays & lesbians measure (see table 11), the nonprobability sample remained significantly different from the evangelical RDD oversample until both the demographic and religiosity variables were used in the weight, upholding the research hypothesis. While the GQR weight accomplishes this same feat, it does not properly account for the racial or educational composition of young evangelicals. Therefore, this finding is likely due to both the increased standard errors and improper specification of the model. For example, the coefficient for the biblical literalist term drops nearly 23% when the GQR weight was applied, but the coefficient increases more than 19%, from the unweighted model, when the weights tailored for young evangelicals are applied. The dummy variable for being a conservative was no longer significant when I applied the weights I constructed. I would also expect political affiliation to be a significant predictor in these models, and the similarity in coefficients suggests that this non-significance is a result of inflated standard errors. Had the young evangelicals been selected using a probability method, a researcher would be able to run the model without weights, using the data more efficiently. Unfortunately, the significant term for the nonrandom sample makes the unweighted model unusable.

Table 12. Ordinary Least Squares Regression of Young Evangelical-only Samples—DV: Attitude towards Legality of Abortion (1=Almost always legal, 4= Almost always illegal)

	Unweighted b (se)	GQR weight b (se)	Demographic Weight b (se)	Demographic & Internet Access Weight b (se)	Demographic & Religiosity Weight b (se)
Nonprobability Sample	0.18	0.36**	-0.08	-0.13	-0.02
	(0.09)	(0.10)	(0.14)	(0.1)5	(0.16)
Race/Ethnicity (White is Reference)					
Black	-0.35*	-0.29	-0.25	-0.31	-0.24
	(0.14)	(0.16)	(0.18)	(0.19)	(0.22)
Hispanic	-0.37*	-0.43**	-0.51	-0.78	-0.29
	(0.15)	(0.16)	(0.27)	(0.33)	(0.23)
All other races	0.15	-0.13	0.16	0.24	0.18
	(0.20)	(0.23)	(0.32)	(0.28)	(0.28)
Education	-0.09	-0.10	-0.18*	-0.14	-0.22**
	(0.05)	(0.07)	(0.07)	(0.09)	(0.07)
Family Income	-0.01	0.01	0.03	0.04	0.05
	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)
Female	0.18	0.19	0.34**	0.40**	0.39**
	(0.09)	(0.10)	(0.12)	(0.13)	(0.12)
R is from the South	0.02	-0.06	0.06	0.03	0.13
	(0.08)	(0.09)	(0.12)	(0.13)	(0.13)
R is a Conservative	0.28**	0.26*	0.14	0.07	0.10
	(0.09)	(0.10)	(0.12)	(0.14)	(0.13)
R Attends Church Weekly or More	0.46***	0.51***	0.34**	0.25*	0.35**
	(0.09)	(0.11)	(0.11)	(0.12)	(0.12)
R is a Biblical Literalist	0.22*	0.12	0.32*	0.42**	0.29*
	(0.09)	(0.14)	(0.12)	(0.14)	(0.14)

n= 432

* p-values < .05, ** p-values < .01, *** p-values < .001

Note: Also controlling for age, living in a rural area, and dummy variables for employment status—part-time, student, homemaker, and unemployed/other with full-time worker as the reference.

As shown in Table 12, the difference between the nonprobability and probability models was significant with the incorrect weighting strategy. This weight also yielded a model in which young evangelicals were not significantly different from the national RDD (see Table 9), contrary to expectations (Evans, 2002; Wilcox, 1992). Furthermore, while the nonrandom sample term was not significant in the unweighted models, inferences made from other coefficients change if I used a weighted model, in accordance with my second hypothesis.

Female evangelicals were less supportive of keeping abortion legal, compared to male evangelicals, and the coefficients doubled from the unweighted model to the appropriately weighted models. They were robust findings, unlike those for education, church attendance, and biblical literalism. While the coefficients decreased for the religiosity variables, there was a dependence on which weight was applied. While the variables still split enough variance in the demographic weight model, they are obviously competing for variance. Since the coefficients are not stable between the weighing strategies, this is not simply an issue of inflated standard errors. Using internet propensity score weight may seem easy to justify, given the nonprobability internet sample. However, the non-significance of the web sample relative to the national RDD sample (see Table 10) is counter to expectations and I am reticent to suggest that, for this dependent variable, the internet access weight was beneficial. While I do find evidence that my conclusions differ depending on which weighting strategy is used (hypothesis 2), ascertaining the best weighting strategy is difficult, and depends on which dependent variable I analyzed.

Discussion

Although the RARW data set contains less than 500 young evangelicals, it demonstrated many of the issues inherent to dealing with nonprobability samples. Non-representation, difficulties with post-collection adjustment, and uncertainty regarding the "right" model, i.e., one that best captures the attitudes of young-evangelicals were endemic in this project. I will briefly review my findings and hypotheses, and discuss the benefits and drawbacks of the nonprobability sample, and conclude with a discussion of nonprobability samples and the consequences of post-adjustment strategies.

Hypotheses

The nonprobability sample differed from the GSS young evangelicals, supporting the first hypothesis. The proportions for the RARW young evangelical RDD were also skewed on many variables, and in some cases were less representative than the nonprobability sample. For the second hypothesis, that the inferences made using the GQR weight would change when weights constructed specifically for young evangelicals were applied, I found a large amount of evidence to support the hypothesis, but this was not consistent across the analyses. The bivariate models for attitudes about abortion and being a young evangelical did not support the second research hypothesis. The multivariate models for the dependent variable feelings toward gays and lesbians comparing young evangelicals to the RARW national RDD also did not differ by weight. However, the multivariate models for the abortion outcome variable, and the bivariate measures for the feelings toward gays and lesbians dependent variable supported the second hypothesis. The issue of proper model specification for the abortion models—either the comparisons of the young evangelicals to the national RDD or the comparison of the young evangelical samples—is more difficult to resolve.

Benefits of the Nonprobability Sample

The small percentage of evangelicals aged 18-29 in the population—estimations from the 2008 GSS suggest slightly less than 8% of non-institutionalized adults fall into this category—makes internet panels and nonprobability samples attractive. The probabilistic sampling strategies to get to this population would certainly have drawbacks. Greenburg Quinlan and Rosner used an age predicted RDD to randomly sample 100 young evangelicals, the expense of making the phone calls and labor of screening respondents must have been costly. That the young evangelical RDD was as skewed as the nonprobability sample on demographics, and was more skewed on the religiosity variables, may be indicative of just this struggle. It may also be the result of finding a way to oversample on age, but in such a way that led to a biased sample of young evangelicals. If I am correct on how the oversample was conducted, a potential problem may lie in the fact that educated, wealthier, urban young evangelicals were more likely to get into the young evangelical RDD, but have opinions that are not representative of all young evangelicals. If this is true, then there is a possibility that a carefully constructed and appropriately adjusted nonprobability sample may yield less bias then a small probability survey. While I agree with Hill and colleagues (2007) that there is potential for nonprobability samples, I would advocate that their use is limited and only researchers with a wealth of knowledge of their target sample should attempt to collect nonprobability datasets for inference.

Combining the young evangelical RDD and nonprobability sample was imperative to creating appropriate weights. I tried using simple inverse probability weights, only using race, gender, region, and education, and after ten rounds the RDD-only young evangelical sample failed to converge. I only needed two rounds for the demographic weight when I used both the web and RDD samples, and when I added the two religiosity measures, I only needed four rounds. Not weighting the RDD would have been problematic for univariate measures, since the sample is so badly skewed.

Drawbacks of the Nonprobability Sample

Adjustment for non-random bias requires auxiliary variables to correct for participation bias due to heightened interest on topics central to the study (Chang and Krosnick, 2009; Groves 2006). However, this is only possible if the auxiliary variables also exist in a gold standard data set. While Greenburg Quinlan Rosner had a randomly selected sample of young evangelicals, the small sample size and skew on racial and educational variables made using this sample for weights indefensible. While it is possible to use the GSS to measure some of the variables, such belonging to an evangelical faith or how often one attends church, it may be more difficult on political attitudes, which are dissimilar with the exception of the question regarding whether the respondent identifies as conservative, moderate or liberal. Because bias could be attributable to political measures as well as religious measures, the lack of auxiliary variables is problematic. Including items on how often the respondent voted or volunteered for a campaign, for example, would have been helpful as far as this data set is concerned.

Weighting the samples was time consuming and led to a less efficient use of the sample. Design effect (DEFF) coefficients for the young evangelical-only comparisons ranged from about .92 to 2.3, depending on which model was used. Not surprisingly, the demographic and attitudinal weight tended to have a higher design effect, while the demographic only weight had somewhat smaller effects. DEFF coefficients for the variables family income and Black were larger than the other variables. In the demographic and attitudinal weight model for feelings toward gays and lesbians, the young evangelical sample needed to be 130% larger than it was to yield the appropriate amount of variance for the term Black coefficient, without weights. Given the design effects for this data set, a larger nonprobability sample may have coped better with the larger standard errors inherent to weighted analysis.

Although weights may not always be necessary in certain instances (Winship and Radbill, 1994) they were necessary in this analysis. Not only was the skew apparent in the univariate analyses, but the shrinking coefficient of the non-random term shows that the addition of weights for the "feelings toward gays and lesbians" models did more than simply increase standard errors to interfere with significance. However, as this was only necessary in one model, and not for the abortion dependent variable, it raises the question of how useful auxiliary variables will be in the analysis. If attitudinal weights are only necessary in certain models, all analyses will require tests of non-random bias and the ability of the weight to correct for bias, if it is detected. Although this work is feasible, disclosure on the creation of weights for datasets like the RARW is imperative.

Conclusion

The internet panel may have some cost-savings in data collection, and while the cost of post-data collection adjustment is unknown, attempts to cope with making a non-representative sample representative is an arduous task (Fricker and Schonlau, 2002). It would behoove the discipline if survey researchers began to discuss respondent burden in contrast to analyst burden, so that people who fund collection or use secondary data sets will have a better idea of the trade-offs and costs of probability and nonprobability procedures. This study illustrated the drawbacks of dealing with a nonrandom sample, and more importantly, the need for post-data collection adjustment. For those target populations whose characteristics of interest are not geographically clustered, however, a nonprobability sample may be the only realistic option. This study also illustrated how post-collection adjustments could be used, and also showed how knowledge of the target population could assist in making a nonprobability sample a viable alternative to probability samples.

Survey research seems poised to generate nonprobability samples for quantitative research. We may blame declining response rates (Curtin et.al, 2005) or the proliferation of internet access (Couper, 2001). Regardless, we need to come to terms with the fact we are not trading quantity of respondents and generalizability for quality of data and internal validity, as our qualitative peers have done, but are trading quality of respondent and external validity for quantity of observations. While it is possible to deal with some of these issues, nonprobability samples and web-based survey research remains problematic, and is not posed to stand apart from probability methods. Although researchers may wish to utilize internet panels, they should do so with the expectation that considerable post-data collection adjustment will be necessary, and they must have a probability sample and demographic and auxiliary variables to carry out those adjustments. More importantly, a failure to use an existing data set to construct the best possible nonprobability sample only complicates matters. By applying a rigorous strategy of data collection and adjustment, the ability to use nonprobability samples for survey research would give those who study certain hard-to-reach populations a chance to conduct cost-effective research. Until the rigorous work of adjustment has been accurately completed, however, a nonprobability survey sample has to be considered unusable in the production of knowledge.

Sources Cited

Acock, A. C. (2005) Working with Missing Values *Journal of Marriage and Family* 67: pp. 1012-1028

Allison, P.D. (2002) Missing Data Sage Publications: Thousand Oaks, CA.

Baker, R.; Blumberg, S.; Brick, J.M.; Couper, M.P.; et.al. (2010) AAPOR Report on Online Panels. Available from www.aapor.org.

Battaglia, M.P., Izrael, D., Hoaglin, D.C., & Frankel, R. (2009) Practical Consideration in Raking Survey Data. Available from www.abtassoc.us.

Bethlehem, J. (2010) Selection Bias in Web Surveys. *International Statistical Review* 78(2) 161-188.

Berrens, R. P.; Bohara, A. K.; Jenkins-Smith, H.; Silva, C.; Weimer, D. L. (2003) The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples *Political Analysis* 11: pp. 1-22.

Chang, L. and Krosnic, J.A. (2009) National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality *Public Opinion Quarterly* 73(4): pp. 641-678.

Couper, M. P. (2001) Web Surveys: A Review of Issues and Approaches *Public Opinion Quarterly* 64: pp. 464-494.

Curtin, R.; Presser, S. and Singer, E. (2005) Changes in Telephone Survey Nonresponse over the Past Quarter Century. *Public Opinion Quarterly* 69(1): 87-98.

Evans, J. H. (2002) Polarization in Abortion Attitudes in U.S. Religious Traditions, 1972-1998 *Sociological Forum*, Vol. 17, No. 3 (Sep., 2002), pp. 397-422

Fricker, R.D. and Schonlau, M. (2002) Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature *Field Methods* 14: 347-367.

Fricker, S.; Galesic, M.; Tourangeau, R.; Yan, T. (2005) An Experimental Comparison of Web and Telephone Surveys *Public Opinion Quarterly* 69: pp. 370-392.

Greenburg Quinlan Rosner (2008) Religion and America's Role in the World Available from www.pbs.org

Groves, R.M.; Singer, E. and Corning, A. (2000) Leverage-Saliency Theory of Participation: Description and an Illustration. *Public Opinion Quarterly* 64(3): 299-308.

Groves, R.M. (2006) Nonresponse Rates and Nonresponse Bias in Household Surveys *Public Opinion Quarterly* 70(5): pp. 646-675.

Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau (2009) Survey Methodology (Second Edition). Wiley: p. 87

Heerwegh, D. and Loosveldt, G. (2008) Face-to-Face Versus Web Surveying in a High-Internet-Coverage Population: Differences in Reponse Quality. *Public Opinion Quarterly* 72(5): 836-846

Hill, S.J.; Lo, J.; Vavreck, L.; Zaller, J. (2007) The Opt-in Internet Panel: Survey Mode, Sampling Methodology and the Implications for Politcal Research. Available from www.ucla.edu.

Himmelstein, J.L. (1986) The Social Basis of Antifeminism: Religious Networks and Culture. *Journal for the Scientific Study of Religion* 25(1): 1-15.

Kreuter, F.; Presser, S.; Tourangeau, R. (2008) Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly* 72(5): 847-865.

Krosnick, J.A. (1991) Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* 5: 213-236.

de Leeuw, E.D. (2005) To Mix or Not to Mix Data Collection Modes in Surveys *Journal of Official Statistics* 21(2) pp. 233-255.

Malhorta, N. (2008) Completion Time and Response Order Effects in Web Surveys. *Public Opinion Quarterly* 72(5) 914-934.

Malhorta, N. & Krosnick, J.A. (2007) The Effect of Survey Mode and Sampling on Inferences about Politcal Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples. *Political Analysis* 15: 286-323.

Rivers, D. and Bailey, D. (2009) Inference From Matched Samples in the 2008 U.S. National Elections. Available from amstat.org.

Robinson, C. (2006) From Every Tribe and Nation? Blacks and the Christian Right *Social Science Quarterly* 87(3): pp. 591-601

Schreuder, H.T.; Gregoire, T. G.; Weyer, J. P. (1999/2001) For What Applications Can Probability and Non-Probability Sampling be Used? *Environmental Monitoring and Assessment* 66: pp. 281-291.

U.S Census Bureau (2009) Table 3: Annual Estimates of the Resident Population by Sex, Race, and Hispanic Origin for the United States: April 1, 2000 to July 1, 2008 (NC-EST2008-03) Available from www.census.gov

Schonlau, M.; Zapert, K.; Simon, L.P.; Sanstad, K.H.; et.al. (2004) A Comparison Between Responses From a Propensity-Weighted Web Survey and an Identical RDD Survey. *Social Science Computer Review* 22: 128-138

Schonlau, M.; van Soest, A.; Kapteyn, A.; Couper, M. (2009) Selection Bias in Web Surveys and the Use of Propensity Scores. *Sociological Methods & Research* 37: 291-318

Smith, Tom W.; Marsden, Peter V; Michael Hout and Jibum Kim. (2011). General Social Surveys, 1972-2010: Cumulative Codebook. National Opinion Research Center. Available from norc.org

Strover,S. (1999) Rural Internet Connectivity. Available from www.utexas.edu.

Toepel, V.; Das, M. and van Soest, A.(2008) Effects of Design in Web Surveys: Comparing Trained and Fresh Respondents. *Public Opinion Quarterly* 72(5): 985-1007.

Valliant, R. and Dever,J.A. (2011) Estimating Propensity Adjustments for Volunteer Web Surveys. *Sociological Methods & Research* 40: 105-137.

Wilcox, C. (1992) Race, Religion, Region and Abortion Attitudes *Sociological Analysis* 53(1): pp. 97-105

Wilson, K.R.; Wallin, J.S.; Reiser, C. (2003) Social Stratification and the Digital Divide. *Social Science Computer Review* 2003(21): 133-143

Winship, C. and Radbill, L. (1994) Sampling Weights and Regression Analysis. *Sociological Methods & Research* 23(2): 230-257.

Yeager, D.S.; Krosnick, J.A.; Chang, L.; Javitz, H.S.; et.al. (2009) Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. and colleagues 2009 Unpublished paper.