

Not by the Book: *Facebook* as Sampling Frame

Christine Brickman Bhutta, Ph.D.

February 24, 2011

Social networking sites and online questionnaires make it possible to do survey research faster, cheaper, and with less assistance than ever before. The methods are especially well-suited for snowball sampling of elusive subpopulations. This note describes my experience surveying thousands of Catholics via Facebook in less than a month, at little expense, and without hired help. Although the respondents were disproportionately female, young, educated, and religiously active, their responses preserved key correlations found in standard surveys conducted by Gallup and the GSS. I relate my methods to existing Web-based methods and offer concrete suggestions for future work.

Keywords: social networking Web sites, *Facebook*, Internet sampling, snowball sampling, chain-referral sampling, coverage error, inference

Online social networking sites (SNSs) offer new ways for researchers to conduct studies quickly, cheaply, and single-handedly – especially when seeking to construct “snowball” samples for exploratory work. *Facebook* is currently the SNS best suited for this type of research, thanks to size (currently exceeding 500 million users worldwide), features, intensive use, and continuing growth. Each *Facebook* user is directly linked to his or her personal “friends,” while also having access to membership in one or more of the millions of *Facebook* groups that connect other users throughout the world. *Facebook* groups are virtual communities linking people with some shared interest, attribute, or cause. Researchers can readily sample populations of interest by working through existing groups or creating new ones.

Although researchers and journalists devote much attention to social networking, I have yet to locate any work that exploits SNSs as a *tool* of research. Existing SNS research focuses on questions related specifically to the phenomenon of online social networking: what functions do SNSs serve for those who use them, and what benefits do users derive (Joinson 2008)? Is the

accumulation of social capital one of those benefits (Ellison, Steinfield, and Lampe 2007)? Do SNS users behave differently or look differently than non-users (Hargittai 2007)? What privacy concerns do the rise of SNSs raise (Dwyer, Hiltz, and Passerini 2007; Jones and Soltren 2005), and to what extent do these concerns influence online behavior (Tufekci 2008)? Can online social interactions predict tie strength (Gilbert and Karahalios 2009)?

My work shifts the emphasis from research *about* SNSs to research *through* SNSs. By working through existing networks of *Facebook* users, I recruited nearly 4,000 baptized Catholics to participate in an exploratory investigation of the role of affective bonds in religious commitment. Because existing surveys lack questions that adequately gauge these affective bonds, I needed to construct a new instrument. And because I was testing a new theory, a non-representative sample was appropriate for my investigation (and necessary due to my lack of financial means).

This methodological note describes my experiences using *Facebook* for survey research and argues for its usefulness in certain contexts. Within five days of releasing a twelve-minute online survey to a *Facebook* group of potential volunteers, I harvested 2,788 completed questionnaires. Within a month, the total number of respondents increased to about 4,000. Total monetary costs averaged less than one cent per survey – vastly less than the cost of surveys obtained through mail, phone, or even E-mail.¹ Moreover, the responses became available for review the moment they were entered. Hence, if the survey turned out to contain any substantial errors or omissions, I could repair the damage within minutes. Although the respondents by no means constituted a random sample of the relevant (Catholic) population, their responses preserved many of the statistical relationships obtained by traditional means. These and other advantages described below suggest that, in some contexts, *Facebook* may be a useful tool for

exploratory work and for rapid pre-testing of surveys destined for dissemination via traditional methods. It may likewise be an effective tool for reaching some hidden populations. And while the present study was quantitative in nature, the recruitment method can easily be used for qualitative studies as well.

The paper proceeds as follows. Section one reviews the relevant literature on chain-referral sampling and electronic survey methods, highlighting strengths and limitations of both of these methods. Section two follows with a description of the *Facebook* features that make it an effective tool for snowball sampling. Section three discusses attempts to recruit study volunteers, and section four details the results of those efforts by offering survey and sample statistics. Section five addresses issues related to sample representativeness, and section six offers suggestions for others interested in replicating this method.

1. Related Research

Chain-referral sampling first emerged in response to the neglect of social structure and interpersonal relationships in survey research methods. As Coleman (1958) notes, most early analyses overlooked the role of relationships, “never including (except by accident) two persons who were friends” (28). Snowball sampling is a chain-referral technique that accumulates data through existing social structures. The researcher begins with a small sample from the target subpopulation and then extends the sample by asking those individuals to recommend others for the study. Chain-referral techniques have the added benefit of providing relatively easy access to “hidden” subpopulations that are almost impossible to sample by standard (phone, mail, or door-to-door) methods, due to their small size or distrust of outsiders. Examples include studies of prostitutes (Faugier and Sargeant 1997), the homeless (Anderson and Calhoun 1992), AIDS

victims (Martin and Dean 1990), members of the LGBT community (Browne 2005), drug users (Griffiths, Gossop, and Strang 1993; Biernacki and Waldorf 1981), and religious “cults” (Lewis 1986).

Sample bias is the principal downside of the chain-referral approach. On the one hand, study volunteers may try to protect their friends by not referring them, a problem known as “masking.” On the other hand, “referrals occur through network links, so subjects with larger personal networks will be oversampled, and relative isolates will be excluded” (Heckathorn 1997:175). Thus Faugier and Sargeant’s (1997) study of prostitutes undersampled women who were new to the business or who had been ostracized by their peers. Participants may also recruit inappropriate volunteers, especially if they misinterpret the study’s design or purpose (Biernacki and Waldorf 1981). And response rates are difficult to define, much less estimate, when participation spreads through forwarded surveys and undocumented invitations.

Despite these limitations, no one disputes the value of chain referral methods for studies of elusive subpopulations and exploratory work (Penrod et al. 2003; Faugier and Sargeant 1997). Moreover, new techniques can help to overcome some of the problems discussed above.²

Facebook and other social networking sites allow us to carry chain-referral methods into the age of the Internet, while also exploiting the strengths of Web-based questionnaires. Elimination of labor, paper, and postage drastically reduce survey costs, making large studies affordable to conduct (Weible and Wallace 1998; Schmidt 1997). Whereas postal delivery time slows the pace of studies conducted with paper surveys, instantaneous data transmission greatly reduces the time needed to carry out Web-based surveys (Evans and Mathur 2005; Wilson and Laskey 2003). Web-based surveys are faster because they eliminate the need to manually input data into data analysis programs (Evans and Mathur 2005). Technological advances allow

researchers to construct complex skip patterns that reduce response burden and perhaps increase response rates (Shropshire, Hawdon, and Witte 2009). Online surveys can also increase the willingness to answer sensitive questions (Tourangeau 2004) and reduce socially-desirable responding (Chang and Krosnic 2009).

SNS sampling shares most of the limitations associated with other forms of Web-based research. We cannot reach those who lack the requisite computer skills and equipment (Couper et al. 2007; Best and Krueger 2002). As a result, studies that use online surveys under-represent those with limited financial resources, members of some racial and ethnic groups, older people, and the less educated (Couper et al. 2007; Best and Krueger 2002). Nor are we likely to reach many people with serious concerns about Internet privacy (Evans and Mathur 2005). The layout and readability of surveys can vary across hardware and software (Evans and Mathur 2005). Electronic surveys can easily reach unintended recipients and are more readily taken multiple times (Smith and Leigh 1997). And response rates tend to be lower than those associated with phone, mail, and interviews (Converse et al. 2008; Greene, Speizer, and Wiitala 2008; Cole 2005; Evans and Mathur 2005; Griffis, Goldsby, and Cooper 2003; McDonald and Adam 2003).

Other factors mitigate some of these concerns. By using software that logs the IP (Internet Protocol) address of the computer upon which the survey was taken, researchers can identify repeat responders (Gosling et al. 2004). Moreover, as the Internet penetration rate continues to increase, Web-based samples become increasingly representative of the population of interest. Finally, researchers can capitalize on the strengths of various methods by utilizing mixed-mode approaches (Greene et al. 2008).

Bearing in mind all these considerations, let us turn to a specific SNS-based project.

2. Facebook as a Sampling Frame

Facebook's size, growing popularity, and features make it the preferred SNS for constructing a snowball sample in the United States. According to the web analytics company, Compete, *Facebook* surpassed *MySpace* as the most popular SNS in the United States in December of 2008, as measured by the number of unique visitors in a given month.³ Throughout 2009, the gap continued to grow, fueled largely by a dramatic increase in *Facebook* site visits but also by a slow but persistent decrease in visits to *MySpace*. As of October 2010, *Facebook* had more than 133 million unique visitors each month, compared to just over 58 million unique *MySpace* visitors.

The value of *Facebook* as a snowball sampling frame extends beyond its size. Equally important is how quickly, easily, and diffusely users communicate information with each other – both directly and indirectly. The average *Facebook* user has 130 “friends” on the site (Facebook 2010a), and interactions occur through private messages and public (“wall”) postings. When Jack posts a message to Jill’s wall, that information is visible to those with access to Jill’s page.

But thanks to other *Facebook* features, other users might acquire the information in the wall posting even without visiting Jill’s page. Every *Facebook* user has a “News Feed” that contains aggregated content posted by friends, along with photo tags, friend requests, event RSVPs and group memberships. Depending on one’s privacy settings, the message that Jack posts to Jill’s wall might show up in the feeds of Jill’s friends, even if Jill’s friends are not friends with Jack. See the discussion section for more information on privacy settings.

Another *Facebook* feature that is relevant to constructing snowball samples is the *Facebook* group. Users can create new groups and join existing groups based on anything, ranging from specific interests, special events, or shared workplaces, regions, high schools, or

colleges. There is no monetary cost to joining or creating a group, and a given user can belong to as many as 200 groups at one time. Group administrators control the content and the membership of the group. Among other things, they decide whether a group is “open” (anyone can join and invite anyone else to join), “closed” (administrators must approve requests of non-members who desire to join the group), or “secret” (membership is by invitation only).

Administrators also have the ability to send mass messages to all group members in groups that do not exceed 5,000 members.

Starting with one or more groups or network of friends, researchers can create snowball samples by gathering respondents via links to additional friends and groups. To illustrate the potential of this simple approach, consider the result of one enterprising *Facebook* user who administers the open group, “Six Degrees of Separation: The Experiment.” In order to maximize the number of group members, he invited all his friends to join and encouraged all of them to do likewise *ad infinitum*. The group recently numbered more than six million.

3. Recruiting Study Participants

How, then, might a researcher navigate this elaborate web of relationships to recruit study volunteers? To some extent this depends on the population of interest and the nature of the study. In this section, I share details of my approach, based on my needs.

My study investigated the role of affective bonds in the religious commitment of baptized Catholics in the U.S. The strategy was straightforward and contained just three basic steps. First, I created a new *Facebook* group that potential study volunteers could join. Second, I populated the group by soliciting the help of my personal network of friends and the administrators of existing *Facebook* groups of Catholics. Third, once the group reached

sufficient size (not knowing what to expect, I was modestly hoping for 500-1,000 members), I sent the survey link to the study volunteers via the “message all group members” feature. In the message, I encouraged volunteers to send the survey link to other friends who were not members of the group, including those who do not have *Facebook* profiles.

I began my search in December 2008 by creating a new group named, “Please Help Me Find Baptized Catholics!” The group’s description explained the purpose of the group, outlined eligibility requirements, and provided instructions on how to be involved. Though I wished only to survey baptized Roman Catholics, the text of my group page invited any viewer to join the group and likewise encouraged them to forward invitations to all their *Facebook* friends and groups. This strategy was designed both to maximize sample size and to avoid the biases associated with sampling down social chains composed entirely of Catholics.

The group page also provided a platform where I attempted to legitimize and humanize the study, two ingredients shown to increase survey response rates (Dillman 2007; Kittleson 2003). On the group’s page, I included my institutional affiliation and my university E-mail address. I also included the names, E-mail addresses, and the university Web sites of the members of my dissertation committee. I posted a scholarly review of the research that justifies non-probability sampling to the group’s discussion board, along with a list of answers to anticipated questions. I also responded to additional questions that individuals posted to the group’s wall, and this information was available for all to see. Each post that I made included a picture of myself, putting a human face on the project.

With the research group in place, I then turned to administrators of existing groups of Catholics for help in advertising my group and recruiting study volunteers. The keywords “Catholic,” “Catholics,” and “Catholicism” returned thousands of results, so I selected the fifty

largest groups that best-represented Catholics generally. To that end, I excluded groups that appeared to have large proportions of foreign members and groups with narrow membership criteria, such as those created for specific *Facebook* networks, college alumni groups, or ethnic groups.

I then contacted the administrators of this subset of existing Catholic *Facebook* groups, soliciting their help in recruiting volunteers for the study. Each administrator received a personal message that explained the purpose of the research and asked them to send a message to the members of their groups with an invitation to join the research group. I also encouraged them to contact me with any questions or concerns (Figure 1).

[Figure 1 about here]

Over the course of three days, I sent personal messages to forty-three of the fifty administrators of the Catholic groups. On the first day of solicitations, I contacted nine administrators,⁴ and six of them responded positively. As a result, my group initially grew quite quickly. Over the next few days, however, my requests for help yielded fewer responses. Of the twenty-five administrators contacted on the second day, eighteen failed to follow up, and no one responded to messages sent on the third day. I suspect that this rapid decline in responses was a consequence of my own initial success. As my group grew, the administrators of other groups may have concluded I no longer needed their help. In any case, in light of the rapid growth of my own group and the rapid decline in responses from other group administrators, I decided not to contact the last seven of fifty administrators.

Table 1 summarizes the outcome of these efforts. Of the forty-three administrators contacted, fifteen agreed to help. Assuming each member of each group became aware of the

study, and assuming no single person was a member of more than one group, the maximum recruitment potential from these fifteen groups was 37, 463.

The number of members who actually learned about the study, however, was almost certainly much lower. Although I asked administrators to send messages to their members, five elected to post the information to their groups' pages instead. Based on personal communications with these administrators, I learned that they did so for one of two reasons: either they considered messaging their members obtrusive, or their group size exceeded 5,000 members, inhibiting their ability to send mass messages. Because viewing the posting required users to visit the group page, and because many *Facebook* users appear to join groups that they rarely if ever return to,⁵ this approach left users less likely to learn about the study. Nevertheless, I preferred some assistance to no assistance. Moreover, one person who posted the link administers a group with over 30,000 members. Even if the rate of awareness in his group was low, the potential for recruitment from his group in absolute numbers was substantial.

[Table 1 about here]

After working through existing Catholic groups, I sent a mass message to all of my 200 *Facebook* friends, seeking their help in recruiting volunteers. This message went directly to each person's inbox. Though I failed to collect systematic information on the results of this process, the feedback that I received via their reply messages suggests that many of them passed my group link along to others. A few also posted information about the study on their personal *Facebook* pages.

As mentioned above, *Facebook* is designed in such a way that group administrators lose the ability to send mass messages to their membership if the group size exceeds 5,000. Because

I planned to use mass mailings to direct people to the online survey, I closed the initial group when membership approached 4,500 and opened a second group for additional volunteers. After 2,500 people joined the second group,⁶ I closed it and opened a third and final group. Altogether, nearly 7,500 people joined one of the three groups over the course of one month and received the link to the survey.

Figure 2 traces the recruitment process for the study. Recruitment occurred in two stages: I first corralled volunteers into my *Facebook* research groups; I then disseminated the survey link via mass message. Boxes with double lines indicate action taken directly by me. My 200 *Facebook* friends and the forty-three administrators of existing Catholic *Facebook* groups were the only individuals I directly solicited for assistance. Volunteers arrived in the research groups through several paths. Some were members of the Catholic *Facebook* groups from which I recruited. Others were my personal *Facebook* friends. Presumably, still others arrived via additional degrees of separation: they were *Facebook* friends of my friends; they were friends of others who were already in (or aware of) the research group; they were friends of friends of friends. It is also possible that some arrived independently. For example, suppose a new *Facebook* user was looking for a Catholic group to join. That person might stumble upon my research group by doing a keyword search of existing *Facebook* groups.

[Figure 2 about here]

After populating the research group, I released the survey to group members by direct message (Figure 3). Some – likely many – of the survey takers were members of my research groups, but others were not. Presumably, some group members forwarded the link to other *Facebook* friends. Others forwarded the link via E-mail to friends who do not have *Facebook* accounts.

[Figure 3 about here]

Figure 2 also aims to clarify why I recruited study volunteers in two stages, rather than one. With so many existing Catholic *Facebook* groups to tap into, why not simply circulate the link to the survey? Why bother with the intermediary step of creating the research group?

Creating the research group allowed me to capitalize on the indirect information channels built into *Facebook*. Much information is circulated among networks very passively through public wall posts and feeds. When an individual joins a group, that information appears in that person's profile and, depending on one's privacy settings, may appear in other users' feeds.⁷ If Jill joins my research group, her friend, Jack, might learn about the study when he visits her page and views her wall. He might also learn about the study without even going to Jill's page if her status update appears in his feed. Thus, simply by joining my group, that individual might recruit friends on my behalf. If instead I opted to merely circulate the link to the survey, that individual must actively forward the link to others.

And as previously noted, by using the research group as a platform where I tried to both legitimize and humanize the study, I hopefully increased the survey response rate (Dillman 2007; Kittleson 2003).

4. Survey and Sample Statistics

As with most studies that employ chain-referral sampling, I am unable to calculate a response rate. One might be tempted to compare the number of members in the research groups with the number of survey takers, but this would create an inflated figure of uncertain size. Not only would we ignore those who took the survey without being in the research group, we would

likewise ignore all those individuals who were invited to the research group but who chose not to join.

However, I am able to calculate the study's completion rate. The Web-based survey software (QuestionPro) collects basic statistics that indicate the response rate of those who start the survey. The survey was started 4,709 times, yielding 4,016 completed surveys for a completion rate of 85.3 percent, comparable to rates achieved in other Web-based studies.⁸ Only 18 responses came from ineligible participants, leaving a total of 3,998 usable surveys. On average, respondents completed the survey in twelve minutes, and the high completion rate suggests this was a reasonable request of one's time.

Speed of response represents a significant advantage of Web-based surveys,⁹ one that this research captures quite well. Figure 4 plots the number of completed surveys within the first month of its release. Although I kept the survey link active for 100 days, the vast majority of completed surveys arrived within days.¹⁰ Members of groups two and three received the survey link at time zero; members of group one received the survey link on day three. Just five hours after releasing the survey, I tallied 426 completed surveys, a little more than 10 percent of the total number of completed surveys. Within five days, the number grew to 2788, 70 percent of the total. Within ten days, the figure increased to 80 percent. Participation continued to taper off, and 90 percent of the data arrived by the end of the first month.¹¹

[Figure 4 about here]

To assess the representativeness of my sample, I compared the *Facebook* adults who were raised Catholic to their counterparts in the 2008 GSS. Compared to the general population of U.S. adult Catholics and ex-Catholics, the *Facebook* sample is younger on average (44.7 years versus 30.5 years), more female (53.0 percent versus 69.9 percent), and less likely to identify as

Latino/Hispanic (32.5 percent versus 5.7 percent). The *Facebook* sample is also much better educated and more religious. Two-thirds of *Facebook* respondents have at least a bachelor's degree (versus almost one-fourth of GSS respondents), and 65.9 percent claim to attend Mass at least once per week (versus 26.7 percent in the GSS). Table 2 summarizes these comparisons.

[Table 2 about here]

Religious activity was especially striking among the *Facebook* males – 27.6 percent report attending Mass *more than* once per week, versus 21.6 percent for females (Table 3). This result directly contradicts not only the GSS data, but also a huge and varied body of research demonstrating greater religiosity among women than men for all aspects of religious behavior, all regions of the world, and all know eras (Miller and Stark 2002; Beit-Hallahmi 1997).

[Table 3 about here]

The sampling frame and sampling technique help explain the bias. Like *Facebook* users generally, the research sample is younger and better-educated than the general population. The religiosity bias, however, likely stems from targeting *Facebook* groups of Catholics. While some groups cater specifically to inactive or “lapsed” Catholics, the majority appear to attract more devout individuals. The latter also appear to be larger and have more group activity.¹² From a sociological standpoint, this is to be expected. To the extent that a group exists to bind together those that share a common identity, the notion of a *Facebook* group of indifferent Catholics is an oxymoron: they have little to rally around. Thus, because I used the groups as my starting point for the snowball sample, I obtained volunteers who were disproportionately religious. This bias persisted even after weighting the data by age and gender (results not shown).

5. Preserving Correlations of Interest

In short, the *Facebook* respondents cannot possibly serve as a representative sample of the general Catholic population. (Pollsters should view *Facebook* findings with extreme caution.) However, like many researchers, I wanted to understand the relationships between certain population characteristics rather than the prevalence of individual characteristics. For example, rather than wanting to know how frequently Catholics attend religious services on average, I hoped to expose the factors that influence this behavior.

As others have shown, biased samples drawn from the Web often preserve measures of statistical relationships quite well. For example, Best et al. (2001) separately analyzed a random telephone sample and an Internet sample and found that the same factors influence political attitudes in both. Consequently, the authors believe they “would have reached the same conclusions about the determinants of particular political attitudes by relying on a diverse convenience sample of Internet users as [they] would have by using a more expensive, time-consuming, probabilistic telephone sample” (143). Bainbridge (2007) also replicated correlations of interest found in the GSS with those in a non-probability, Web-based sample.

But, as both Best et al. (2001) and Bainbridge (1999) point out, not all non-random samples preserve correlations of interest. Specifically, researchers must be certain that the source of the bias in the data does not correlate with the relationships of interest. With Internet surveys, for instance, researchers must assume that the relationships of interest are not influenced by patterns of Internet use – an assumption that obviously fails when studying something like attitudes toward technology or Internet privacy.

Therefore, the relevant question of the current investigation is this: does the relationship between Catholic commitment and other variables of interest – the topic of my dissertation – correlate with *Facebook* usage?

To address this question, Table 4 reports the *Facebook* correlations between variables of interest in my study and compares them to corresponding correlations derived from two high-quality probability-sample surveys: the General Social Survey and the Gallup Poll of Catholics. As noted above, my dissertation explored the role of affective bonds in sustaining religious commitment. Because existing surveys do not gauge affective bonds (hence the *Facebook* study), I could not compare this exact relationship across samples. However, we might understand affective religiosity as a dimension of religiosity, much like religious orthodoxy and religious social networks are dimensions of religiosity (Lenski 1961). Consequently, the first nine variables in Table 4 offer the best available proxies for affective bonds. The outcome variable is a common measure of religious commitment – a dichotomous variable indicating whether the respondent claims to attend Mass at least weekly. The table also correlates Mass attendance with standard demographic controls, which have also been shown to influence religious commitment.

As shown in Table 4, the *Facebook* data preserved the correlations between the variables of primary interest. In both *Facebook* and Gallup samples, respondents are more likely to attend Mass weekly if they expect Catholics to adhere to traditional church teachings, if they are registered in a Catholic parish, and if they believe Catholicism contains a greater share of the truth than other religions do. The match-up is equally good for *Facebook* and GSS.

[Table 4 about here]

The correlations between Mass attendance and many demographic variables are more difficult to evaluate due to frequent lack of agreement between the GSS and Gallup data. For example, the GSS data capture the tendency for women to be more religious than men, but *both* the Gallup and *Facebook* data fail to preserve this relationship. Gallup and the GSS similarly disagree over the relationship between both education and being Latino and religious attendance, again making it difficult to evaluate the corresponding correlations in the *Facebook* data.

One area where GSS and Gallup agree while *Facebook* differs is the relationship between age and Mass attendance. Although a strong positive correlation exists in both probability samples, the corresponding *Facebook* correlation is zero.

Closer analysis reveals a curvilinear relationship between age and attendance in the *Facebook* data (Figure 5). Although overall rates of Mass attendance are much higher in the *Facebook* sample than in the GSS and Gallup, the correlation for those aged 30 and older is positive and significant in all three samples (the *Facebook* correlation is 0.14, $p < 0.001$). A statistically significant, negative correlation (-0.10 , $p < 0.001$) exists for those between age 18 and 29 in the *Facebook* sample.

[Figure 5 about here]

Social network theory may explain this unusual finding among young respondents. Contemporary research has repeatedly demonstrated that social ties and interaction are key to religious commitment (Cornwall 1989; Welch 1981). Applying this insight to the college setting, we might expect heightened religious involvement among those who join faith-based campus groups where social ties can flourish. However, after graduation religious commitment may suffer as these social networks dissolve. Anecdotal data lends credence to this hypothesis. An extensive study of young adult Catholics revealed that significant numbers of Catholic

college students who were active in Campus Ministry were often frustrated after graduation because they found it difficult to find a parish experience as vital and engaging as the collegiate one (Hoge et al. 2001). This frustration may lead some to attend Mass less frequently.

Traditional surveys may not detect this pattern of religious involvement if members of these faith-based campus groups comprise a relative minority of the population. Perhaps the pattern is evident in the *Facebook* data because the sample over-represents these individuals.

Certainly this is just a working hypothesis, but the data illustrate a case where *Facebook* usage *does* correlate with a pattern of Catholic commitment. If age were central to the investigation, *Facebook* sampling would not be a viable approach (at least for those under age 30).

For the present study, however, the correlations that held were more important than those that failed. As noted, whereas the demographic variables were control variables in the model, the religion variables (the first nine variables in the table) offer the best available proxies for affect, the variable of interest in the study. Moreover, because of the unusual relationship between age and attendance among people under 30, I ran the dissertation model separately for younger (under age 30) and older (30 and older) respondents and found very similar results.¹³

Table 4 also reveals the limitations of probability-based studies and underscores the importance of considering cost-benefit trade-offs when selecting the best method of data collection. Although the *Facebook* data failed to preserve the correlation between gender and Mass attendance found in most empirical studies, this relationship was likewise absent in the Gallup data. Correlations between Mass attendance and both education and being Latino likewise differed in the GSS and Gallup samples. While these findings do not make the *Facebook* method superior to a probability-based technique, they highlight the fact that even

national samples often fall short of true randomness (Bainbridge 2002; 1999), yet they are also quite expensive to conduct. In some studies of certain populations, a non-probability approach may therefore offer an attractive cost-benefit trade-off.

In short, the viability of *Facebook* sampling depends on both the population of interest and the particular research question. Those interested in replicating this method are advised to first consider whether they can reasonably expect that *Facebook* usage does not influence the relationship between the variables of interest. If, *ex ante*, they anticipate that a *Facebook* sample will produce the same correlations found in the population, the *Facebook* approach may be worthwhile. Those who then choose to move forward can subsequently evaluate the veracity of this assumption by including questions from probability-based surveys. If the assumption holds, the researcher has a useful alternative to more traditional means of pre-testing surveys or for conducting exploratory studies. Moreover, the preliminary results may provide good evidence in support of a working hypothesis that helps the researcher secure a grant for a traditional study. In the unfortunate case that the assumption fails, the monetary cost to the researcher is small because *Facebook* sampling is relatively inexpensive.

6. Discussion

Because my method of data collection appears to be unprecedented, I could not avail myself of any established procedures or “Best Strategies.” I relied instead on standard insights, personal intuition, and pure luck. I have summarized my experience in the hope of helping others exploit the unique strengths of sampling through social networking sites while avoiding or at least understanding their limitations.

The researcher should bear in mind that the suggestions below are based on *Facebook*'s current functionality and popularity. At the present time, *Facebook* offers a useful platform for social scientific research, but any one of many factors could easily undermine its utility. Although *Facebook* has experienced incredible growth since its inception, there is no guarantee that this will continue. Just as *Facebook* once usurped *MySpace*, so might another innovation undermine *Facebook*'s popularity. Furthermore, *Facebook* frequently changes its feature set, often with little or no announcement, such as when it removed its geography-based networks in 2009 (Facebook 2010b). If *Facebook* were to similarly remove the "message all members" feature that is available to group administrators, this would eliminate what was the primary recruitment method for my study.

Facebook users can also alter the viability of the site for research by changing their privacy settings. In fact, concern about privacy has been a direct result of some of the changes to the *Facebook* site, including the addition of the "News Feed" in 2006 (Boyd 2008) and, more recently, changes to users' default privacy settings (Boyd and Hargittai 2010). Although studies show considerable variation in how frequently users alter their privacy settings, (Madden and Smith 2010; Stutzman and Kramer–Duffield 2010; Lewis, Kaufman, and Christakis 2008; Ellison et al. 2007), one longitudinal study finds that engagement with privacy settings is on the rise, having increased significantly between 2009 and 2010 (Boyd and Hargittai 2010).

For all of these reasons, scholars who plan to pursue a similar method are advised to evaluate the current features of the site before undertaking their investigations.

6a. Recruitment

Keeping the size of my groups under 5,000 members was critical to the success of the study. As noted above, group administrators lose the ability to send mass messages to members when the group size exceeds this number. Had this occurred, I would need to rely on my volunteers to return to the group page to access the survey link when it became available. Like many *Facebook* groups, however, “Please Help Me Find Baptized Catholics,” had little activity. The groups that keep members coming back are those that keep the content fresh – frequently updating the group's “Recent News,” posting new photos and videos, adding current events and links – so members have a reason to return. Even had I attempted this, there is little chance I would captivate the interests of all 7,500 members, prompting them to return to the page on a regular basis. Therefore, closing the group well before the size reaches the 5,000 member threshold is a vital strategy for a successful study.¹⁴

The size restriction also suggests the method of corralling study volunteers into a research group is not viable for large studies. I was only able to gather 7,500 members by creating multiple groups, but the process of managing membership quickly became inefficient. In theory, I could have acquired more volunteers by continually opening new groups and moving members across them, but I found this unmanageable. Therefore, I do not recommend this approach to those who seek more than a few thousand respondents.

Those with financial means might circumvent this problem by taking out a targeted *Facebook* ad. Researchers can target study volunteers by location, sex, age, keyword, relationship status, job title, workplace, or college. *Facebook* ads display in the right side column on *Facebook* pages, and up to three ads may show at one time on any given page. There is no set cost for *Facebook* ads; rather, advertisers compete with one another by bidding for advertising

space. Rather than creating a research group, one could directly link the target audience to the survey by clicking on the advertisement.

The experience of at least one researcher suggests targeted ads provide a useful way to recruit survey respondents, although it remains unclear whether this approach leads to a more or less representative sample. This researcher was interested in attitudes toward fate, and he set up a daily budget, which determined how many people would receive the ad. He budgeted \$1,000 each day for 30 consecutive days and always reached the budget by day's end. Although his sample was even more biased toward females and young respondents than was the sample in the Catholic study,¹⁵ this disparity may reflect the topic of the investigation more than the recruitment method. For instance, perhaps women are relatively more interested in fate than in Catholicism and hence more willing to complete a survey about fate. In any case, the experience of the fate researcher points to a potential for heightened bias of which others should be cognizant, should they choose to use *Facebook* ads for recruitment.

6b. Assessing and Mitigating Sample Bias

Researchers planning to use *Facebook* as a sampling frame can also undertake measures that enable them to assess – and perhaps mitigate – bias that will necessarily arise. Above I noted that including several items in my survey pulled directly from two probability-based surveys enabled me to assess the extent which *Facebook* usage correlated with relationships of interest. These benchmarks also enabled me to measure the extent of the bias in the data, as well as to evaluate the ability of survey weights to improve its representativeness. Although weights were ineffectual in the present study, other researchers might achieve different results. Indeed, if the experiences of those who undertake more traditional Web-based studies provide any guide,

we should expect mixed success (Dever, Rafferty, and Valliant 2008; Loosveldt and Sonck 2008; Valliant and Lee 2005).

The diversity of *Facebook* groups lends itself to a stratified sampling approach that may increase the representativeness of the sample. Earlier I mentioned that I only sampled groups that appeared representative of the target population: all individuals baptized in the Catholic Church in the United States. In retrospect, this was an ineffectual approach because the sampling frame itself (i.e. *Facebook* users) is biased in several ways, and my sample reflected that bias. Had I employed a stratified approach, I might have mitigated that bias by targeting specific groups, such as Latino Catholics and groups comprised of older respondents.

6c. Other SNS's?

One might also wonder whether or not expanding the sampling frame to include other SNSs would lead to a more representative sample. Although *Facebook* is currently the world's largest SNS, many others are also quite large and appeal to a different demographic. According to data from [compete.com](http://www.compete.com), *MySpace* receives over 58 million unique visitors each month, and *LinkedIn* receives just under 14 million. A recent study by Lenhart (2009:6) summarizes demographic variations in the types of people who tend to use each of these three social networking sites. While both *Facebook* and *MySpace* appeal to a younger audience (median age of 26 and 27, respectively), *MySpace* users are more likely to be Hispanic or black, to be female, and to have a high school education or some college. *Facebook* users tend to be white and to have a college degree. *LinkedIn* users are also likely to be white, male, and well-educated, but this SNS appeals to working professionals who are relatively older (median age of 40). Perhaps

I could have increased the representativeness of Latinos and older respondents by including *MySpace* and *LinkedIn* users, respectively.

How effectively I would have reached those individuals, however, remains uncertain. A variety of factors leave me pessimistic. Perhaps most importantly, neither *LinkedIn* nor *MySpace* equips group administrators with the “message all members” feature as *Facebook* does. Even individuals who expressed interest in the study would have to return to the group page to receive the survey link. Both *LinkedIn* and *MySpace* also lack the equivalent of *Facebook*’s “feeds” that transmit information passively and diffusely among networks. *Facebook* has more than seven times¹⁶ as many groups as *MySpace*, leaving the researcher many fewer sources of potential volunteers on *MySpace*. Finally, I am unsure how interested *LinkedIn* users would have been in my study, given that the site exists primarily for professional networking.

In short, while in theory opening the sampling frame to include users of other SNSs should have increased the representativeness of my sample, in practice I believe I would have achieved limited success, given the nature of my study and its design. The extent to which other SNSs present viable options for other researchers depends largely on their needs and the scope of their studies.

6d. Other Best Strategies and Recommended Uses

Releasing the survey in multiple waves can mitigate traffic-related problems and offers an opportunity to make changes to unforeseen problems. My experience offers an imperfect attempt to avoid these pitfalls. As displayed in Figure 1, when I first released the survey, I sent the link to approximately one-third of all group members; the remaining volunteers received the survey after three days. Within minutes of disseminating the link in the first wave, I received an

E-mail from a volunteer identifying a problem with one of the questions that I did not catch, even after pre-testing the survey.¹⁷ I corrected the problem immediately, but several minutes passed before the change took effect, presumably due to the large volume of individuals on the Web site at the time. Fortunately, the erroneous item was not critical to my primary analysis (I did not use the variable in my model), so it was not necessary for me to exclude these cases. Future researchers can learn from this mistake by releasing the survey only to a few dozen respondents initially. This experience offers another lesson as well – researchers conducting surveys via the Web must be mindful of the limitations of the Web-based software they choose.

The *Facebook* method has other potential uses besides exploratory studies. Given its low cost, the quick turnaround time, and the ability to revise items on the fly, I strongly recommend the use of *Facebook* for pretesting survey items.

Facebook sampling may also be a viable option for research of some hidden populations. Because it is a chain-referral method, *Facebook* sampling provides access to some populations that are absent from standard samples because they are too small or too difficult to reach. Consider a hypothetical study of members of one Protestant denomination: the Lutheran Church-Missouri Synod. A researcher who turns to the 2008 GSS will find just thirty members of this group, but I found over 14,000 members spread out over sixteen *Facebook* groups. Suppose instead that a health researcher is trying to locate victims of thyroid cancer and their families. A quick keyword search returned several dozen thyroid cancer groups on *Facebook* comprised of thousands of members. Thanks to widespread use and niche groups, *Facebook* offers researchers a way to easily reach many otherwise hidden populations.

However, the *Facebook* approach may not work when the hard-to-reach population is stigmatized. Many of these groups – prostitutes and drug users, for example – are unlikely to

have a public presence on *Facebook* or other SNSs, so it would be difficult for a researcher to locate them. Even if one could identify the population of interest, establishing trust in an online medium would likely pose quite a challenge to the researcher. For example, I found several groups of Jehovah's Witnesses, but the administrators of these groups may be skeptical of an unsolicited message like the one I sent to my Catholic groups. However, if the researcher had a pre-existing relationship with a member of the stigmatized group, then *Facebook* or another SNS may still provide a useful platform to navigate close, intimate networks.

Although the present study restricted the sampling frame to the United States, *Facebook* is an increasingly attractive option for international research as well – particularly in nations with diffuse Internet use. Among today's active *Facebook* users, 70 percent live outside the United States (Facebook 2010a). Whereas international research greatly increases the cost of conducting survey research via traditional methods like mail and telephone, the *Facebook* method poses no additional costs. The best candidates for international research will be nations with high rates of Internet diffusion generally, and high rates of *Facebook* use specifically. Coverage error poses an increasing concern the lower the penetration rate.

As a final note, my research project aroused no serious concerns from the Institutional Review Board (IRB) at my university. Anonymity, confidentiality, and data security transmission present challenges to researchers who collect data via the Internet. The potential to record the IP address threatens to reveal the identity of subjects, and the data are susceptible to hacking and corruption (Benfield and Szlemko 2006). I was able to allay human rights concerns with limited effort. In terms of the survey itself, my IRB simply required that my survey begin with an informed consent page and that I ask respondents to indicate that they were at least 18 years old. They also required me to submit information on the privacy and data security efforts

taken by QuestionPro, the company that provided the Web-based survey software for the study. Presumably, the ease in which I received IRB approval reflected the safety and security of QuestionPro software,¹⁸ rather than a lax review process.

7. Conclusion

Over thirty-five years ago, Mark Granovetter (1973) illustrated how weak ties transmit information more quickly and more diffusely than do strong ties. Those to “whom we are weakly tied are more likely to move in different circles from our own and thus will have access to information different from that which we receive” (1371). Weak ties are the bridges between small clusters of close friends, linking us together to form an elaborate web of social relationships.

Facebook offers researchers a way to capitalize on the strength of these weak ties. According to *Facebook's* internal statistics, the average user has 130 friends (Facebook 2010a). While some of these relationships constitute “strong ties,” many are also acquaintances, old friends from high school or college, even total strangers (DiMicco et al. 2008; Joinson 2008; Lampe, Ellison, and Steinfeld 2008). Millions of *Facebook* groups also exist on the site, linking users to countless others whom they do not even know.

The present study offers a modest attempt to exploit these features, which make *Facebook* a useful tool for snowball sampling. By navigating through existing groups of Catholics and by tapping into personal friends, I recruited nearly 4,000 individuals to participate in my research study. Data collection was extremely fast and incredibly cheap. Within five days 2,700 individuals completed the survey, and a modest license fee for the Web-based software was the only expense. The data also preserved the statistical relationships between the variables

of interest, despite being biased in several ways. While snowball sampling via *Facebook* is no substitute for probability-based techniques, the fact that the relevant correlations between variables hold suggests *Facebook* may be a valuable tool for exploratory research of certain populations.

Because the present study is the first of its kind, it remains unclear how broadly this technique can be applied. As similar studies of other populations are conducted, we can better evaluate the merits of *Facebook* as a general tool for social scientific research.

Notes

1. Raziano et al. (2001) report a cost of \$10.50/mail survey and \$7.70/E-mail survey. Harewood et al.'s (2001) study reported a cost per completed survey of \$0.71 for E-mail, \$2.54 for mail, and \$2.08 for telephone. Cobanoglu, Warde, and Moreo (2001) report a unit cost of \$2.61 for mail surveys and \$1.20 for faxed surveys *excluding* the cost of coding the resulting information into machine-readable form.
2. See Heckathorn's (1997) discussion of respondent-driven sampling (RDS), which attempts to overcome problems of non-representativeness.
3. Compete's "Site Comparison" feature allows users to compare statistics for millions of Web sites. I generated the following report, comparing *MySpace* and *Facebook* usage: <http://siteanalytics.compete.com/facebook.com+myspace.com/> [accessed February 10, 2010].
4. While the exact number was arbitrary, my intent was to contact a subset of administrators and wait for their responses to see how well the message was received. For example, perhaps I poorly explained the projected or omitted a character in the survey group link. By waiting for a few responses, I left an opportunity to revise outreach efforts based on their feedback.
5. This is merely my impression based on personal experience. Many groups I belong to have little activity – members infrequently post messages to the wall or discussion board; the administrators rarely change the content of the page or send messages to members.
6. After closing a group, new members can continue to join if they are responding to an existing group invitation. I became aware of this after closing the first group and becoming alarmed when the group approached 5,000 members. In fact, in order to keep the first group below the threshold, I sent the group members a message and asked some of them to leave the first group and join the second group. Because I wanted to avoid this problem with the second group, I closely monitored its growth and the number of outstanding invitations to the group. I closed group two with just 2,500 members and over 2,000 outstanding group invitations. At this point, I realized I could likely continue to snowball users into several more research groups, but I was content with the amount of current interest in the study. Therefore, I created group three as a "closed" group – users who desired to join must send a request to the group administrator for permission. As a result just over one hundred members comprised this group.
7. Users can adjust their privacy setting so information like this does not appear. However, the default setting makes this information public in a user's profile and available in the feeds. See the discussion section for more information on privacy settings.
8. Survey2000 had a completion rate of 80.5 percent (Witte, Amoroso, and Howard 2000). Couper and his colleagues report a drop-out rate of 10.6 percent in their online survey (Couper et al. 2004). Knowledge Networks claims an average completion rate of 88 percent for their Web surveys (Schonlau, Fricker, and Elliott 2002). Finally, an analysis of 67 Web surveys yielded an average drop-out rate of 16 percent (Manfreda and Vehovar 2002).
9. Those who use mixed modes to gather data typically report faster speed of response with electronic surveys than with traditional modes. In Cobanoglu et al.'s (2001) study, the average response speed for mail and Web-based surveys was 16.46 days and 5.97 days, respectively. Schaefer and Dillman (1998) reported rates of 14.39 days and 9.16 days for mail and E-mail modes, respectively.

10. Pan (2009) reports similar response patterns. He administered four separate Web-based surveys and found that “the response percentages of the first days are generally larger than 70 percent” (8).
11. I also acquired volunteers to participate in an open-ended, follow-up IM interview from among the survey respondents. After submitting the survey, respondents received an invitation to participate in a follow-up interview if they met certain criteria. Over 250 expressed interest in participating by E-mailing me.
12. I base this on the information obtained from the Catholic groups I targeted. For example, one group is entitled “You May Be a Hardcore Catholic If...” At the time of writing, it had nearly 30,000 members, and six different individuals posted a message to the wall in the past twenty-four hours. Information posted on the group’s page provides further evidence that devout Catholics comprise the group. One posting warns that “[a]ny posts on the wall and discussion board that are offensive, crude, use foul language, or misrepresent the Catholic Church and its teachings will be deleted.” In contrast, the group entitled, “Ex-Catholics,” has just ninety-two members, and the last wall posting was made three weeks ago.
13. In the full sample, the coefficient on affect is 1.30 ($p < 0.001$); for the samples of younger and older respondents, the coefficients are 1.32 ($p < 0.001$) and 1.35 ($p < 0.001$), respectively. Age is not statistically significant in any of the three cases.
14. Even after closing a group, new members can continue to join if they are responding to an outstanding group invitation from existing members. Several hundred additional users joined after I closed my groups.
15. In the study that used targeted *Facebook* ads, 85 percent of the sample was aged 35 or younger and 88% was female. The corresponding figures in my study were 73 percent and 70 percent, respectively.
16. At last count, MySpace had 4.7 million groups, and Facebook had over 35 million.
17. The problematic item asked respondents to “select all of the ranges that reflect the ages of your children.” Although this was a multiple response item, I inadvertently set up the software to only accept a single response. Apparently, no one in my pre-test sample had children in multiple age ranges.
18. Although QuestionPro does collect IP addresses, the company’s privacy policy explicitly states that this information does not identify individual users. QuestionPro also has a firewall designed to prevent hackers from entering the system.

References

- Anderson, Leon and Thomas P. Calhoun. 1992. "Facilitative Aspects of Field Research." *Sociological Inquiry* 62(4): 490-498.
- Bainbridge, William S. 2007. "Expanding the Use of the Internet in Religious Research." *Review of Religious Research* 49(1) 7-20.
- . 2002. "Validity of Web-Based Surveys: Explorations with Data from 2,382 Teenagers." Pp. 51-66 in *Computing in the Social Sciences and Humanities*, edited by O.V. Burton. Chicago: University of Illinois Press.
- . 1999. "Cyberspace: Sociology's Natural " *Contemporary Sociology* 28(6): 664-667.
- Beit-Hallahmi, Benjamin. 1997. "Biology, Density and Change: Women's Religiosity and Economic Development." *Journal of Institutional and Theoretical Economics* 153: 166-178.
- Benfield, Jacob A. and William J. Szlemko. 2006. "Internet-based Data Collection: Promises and Realities." *Journal of Research Practice* 2(2). Retrieved February 17, 2010 (<http://jrp.icaap.org/index.php/jrp/article/viewFile/30/122>).
- Best, Samuel J. and Brian Krueger. 2002. "New Approaches to Assessing Opinion: The Prospect for Electronic Mail Surveys." *International Journal of Public Opinion Research* 14(1): 73-92.
- Best, Samuel J., Brian Krueger, Clark Hubbard, and Andrew Smith. 2001. "An Assessment of the Generalizability of Internet Surveys." *Social Science Computer Review* 19(2): 131-145.
- Biernacki, Patrick and Dan Waldorf. 1981. "Snowball Sampling: Problems and Techniques of Chain Referral Sampling." *Sociological Methods and Research* (10)2: 141-163.
- Boyd, Danah. 2008. "Facebook's Privacy Trainwreck: Exposure, Invasion, and Social Convergence." *Convergence* 14(1): 13-20.
- Boyd, Danah and Eszter Hargittai. 2010. "Facebook Privacy Settings: Who Cares?" *First Monday* [Online] 15(8). Retrieved October 5, 2010 (<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086/2589>).
- Browne, Kath. 2005. "Snowball Sampling: Using Social Networks to Research Nonheterosexual Women." *International Journal of Social Research Methodology* 8(1): 47-60.

- Chang, Linchiat and Jon A. Krosnick. 2009. "National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly* 73(4):641-678.
- Cobanoglu, Cihan, Bill Warde, and Patrick Moreo. 2001. "A Comparison of Mail, Fax and Web-based Survey Methods." *International Journal of Market Research* 43(4): 441-452.
- Cole, Shu Tian. 2005. "Comparing Mail and Web-Based Survey Distribution Methods: Results of Surveys to Leisure Travel Retailers." *Journal of Travel Research* 43(May): 422-430.
- Coleman, James. 1958. "Relational Analysis: The Study of Social Organizations with Survey Methods." *Human Organization* 17(4): 28-36.
- Converse, Patrick, Edward Wolfe, Xiaoting Huang, and Frederick Oswald. 2008. "Response Rates for Mixed-Mode Surveys Using Mail and E-mail/Web." *American Journal of Evaluation* 29(1): 99-107.
- Cornwall, Marie. 1989. "The Determinants of Religious Behavior: A Theoretical Model and Empirical Test." *Social Forces* 68(2): 572-592.
- Couper, Mick, Arie Kapteyn, Matthias Schonlau, and Joachim Winter. 2007. "Noncoverage and Nonresponse in an Internet Survey." *Social Science Research* 36(1): 131-148.
- Couper, Mick P., Roger Tourangeau, Frederick G. Conrad, and Scott D. Crawford. 2004. "What They See is What We Get: Response Options for Web Surveys." *Social Science Computer Review* 22(1): 111-127.
- Dever, Jill, Ann Rafferty and Richard Valliant. 2008. "Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias?" *Survey Research Methods* 2(2): 47-62.
- Dillman, Don. 2007. *Mail and Internet Surveys: The Tailored-Design Method*. Hoboken: John Wiley and Sons.
- DiMicco, J., D. Millen, W. Geyer, C. Dugan, B. Brownholtz, and M. Muller. 2008. "Motivations for Social Networking at Work." *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*: 711-720.
- Dwyer, Catherine, Starr Roxanne Hiltz, and Katia Passerini. 2007. "Trust and Privacy Concern within Social Networking Sites: A Comparison of *Facebook* and *MySpace*." Paper presented at the Thirteenth Americas Conference on Information Systems, Keystone, Colorado, August 9 – 12.

- Ellison, Nicole, Charles Steinfield, and Cliff Lampe. 2007. "The Benefits of *Facebook* Friends': Social Capital and College Students' Use of Online Social Network Sites." *Journal of Computer-Mediated Communication* 12(4): 1143-1168.
- Evans, Joel R. and Anil Mathur. 2005. "The Value of Online Surveys." *Internet Research* 15(2): 195-219.
- Facebook. 2010a. *Facebook Statistics*. <http://www.facebook.com/press/info.php?statistics> [accessed September 27, 2010].
- Facebook. 2010b. "Growing Beyond Regional Networks." *The Facebook Blog*. Retrieved October 5, 2010 (<http://blog.facebook.com/blog.php?post=91242982130>).
- Faugier, Jean and Mary Sargeant. 1997. "Sampling Hard to Reach Populations." *Journal of Advanced Nursing* 26(4): 790-797.
- Gilbert, Eric and Karrie Karahalios. 2009. "Predicting Tie Strength with Social Media." Paper presented at the 27th International Conference on Human Factors in Computing Systems, Boston, MA. April 4 – April 9.
- Gosling, Samuel, Simine Vazire, Sanjay Srivastava, and John Oliver. 2004. "Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions about Internet Questionnaires." *American Psychologist* 59: 93-104.
- Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78(6): 1360-1380.
- Greene, Jessica, Howard Speizer, and Wyndy Wiitala. 2008. "Telephone and Web: Mixed-Mode Challenge." *Health Services Research* 43(1): 230-248.
- Griffis, Stanley, Thomas Goldsby, and Martha Cooper. 2003. "Web-Based Surveys: A Comparison of Response, Data, and Cost." *Journal of Business Logistics* 24(2): 237-258.
- Griffiths, P., M. Gossop, and J. Strang. 1993. "Reaching Hidden Populations of Drug Users by Privileged Access Interviewers: Methodological and Practical Issues." *Addiction* 88(12): 1617 – 1626.
- Hargittai, Eszter. 2007. "Whose Space? Differences among Users and Non-Users of Social Network Sites." *Journal of Computer-Mediated Communication* 13(1): 276-297.
- Harewood, G. C., R.F. Yacavone, G.R. Locke, and M.J. Wiersema. 2001. "Prospective Comparison of Endoscopy Patient Satisfaction Surveys: E-mail versus Standard Mail versus Telephone." *The Journal of Gastroenterology* 96(12): 3312-3317.

- Heckathorn, Douglas D. 1997. "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations." *Social Problems* 44(2): 174-199.
- Hoge, Dean R., William D. Dinges, Mary Johnson, and Juan L. Gonzales. 2001. *Young Adult Catholics: Religion in the Culture of Choice*. Notre Dame: University of Notre Dame Press.
- Joinson, Adam. 2008. "Looking at, Looking up or Keeping up with People?: Motives and Use of *Facebook*." Paper presented at the 26th International Conference on Human Factors in Computing Systems, Florence, Italy. April 5 – April 10.
- Jones, Harvey and Jose Hiram Soltren. 2005. "*Facebook*: Threats to Privacy." Project MAC: MIT Project on Mathematics and Computing. Retrieved August 12, 2009 (<http://www-swiss.ai.mit.edu/6.805/student-papers/fall05-papers/facebook.pdf>).
- Kittleson, Mark J. 2003. "Suggestions for Using the Web to Collect Data." *American Journal of Health Behavior* 27(2): 170-172.
- Lampe, Cliff, Nicole Ellison, and Charles Steinfield. 2008. "Changes in Use and Perception of *Facebook*." *Proceedings of the 2008 Conference on Computer-Supported Cooperative Work*: 721-730.
- Lenhart, Amanda. 2009. "Adults and Social Network Websites." *Pew Internet and American Life Project*. Retrieved February 10, 2010 (<http://www.pewinternet.org/Reports/2009/Adults-and-Social-Network-Websites/2-Main-Findings/2-Half-of-social-network-users-have-profiles-on-MySpace.aspx?r=1>).
- Lenski, Gerhard. 1961. *The Religious Factor*. Garden City: Doubleday.
- Lewis, James R. 1986. "Reconstructing the 'Cult' Experience: Post-Involvement Attitudes as a Function of Mode of Exit and Post-Involvement Socialization." *Sociological Analysis* 47(2): 151-159.
- Lewis, K., Kaufman, J. and Christakis, N. 2008. "The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network." *Journal of Computer-Mediated Communication* 14: 79–100.
- Loosveldt, Geert, and Nathalie Sonck. 2008. "An Evaluation of the Weighting Procedures for an Online Access Panel Survey." *Survey Research Methods* 2(2): 93-105.
- Madden, Mary and Aaron Smith. 2010. "Reputation Management and Social Media." *Pew Internet and American Life Project*. Retrieved November 19, 2010 (<http://pewinternet.org/Reports/2010/Reputation-Management.aspx>).

- Manfreda, Katja Lozar and Vasja Vehovar. 2002. "Survey Design Features Influencing Response Rates in Web Surveys." Paper presented at the International Conference on Improving Surveys, Copenhagen, Denmark. August 25 – 28.
- Martin, John L. and Laura Dean. 1990. "Developing a Community Sample of Gay Men for an Epidemiologic Study of AIDS." *American Behavioral Scientist* 33(5): 546-561.
- McDonald, Heath and Stewart Adam. 2003. "A Comparison of Online and Postal Data Collection Methods in Marketing Research." *Marketing Intelligence and Planning* 21(2): 85-95.
- Miller, Alan S. and Rodney Stark. 2002. "Gender and Religiousness: Can Socialization Explanations Be Saved?" *American Journal of Sociology* 107: 1399-1423.
- Pan, Bing. 2009. "Online Travel Surveys and Response Patterns." *Journal of Travel Research* 47(4): 1-16.
- Penrod, Janice, Deborah Preston, Richard Cain, and Michael Starks. 2003. A Discussion of Chain Referral as a Method of Sampling Hard-to-Reach Populations." *Journal of Transcultural Nursing* 14(2): 100-107.
- Raziano, D.B, R. Jayadevappa, D. Valenzula, M. Weiner, and R. Lavizzo-Mourey. 2001. "E-mail versus Conventional Postal Mail Survey of Geriatric Chiefs." *Gerontologist* 41: 799-804.
- Schaefer, David R. and Don A. Dillman. 1998. "Development of a Standard E-mail Methodology: Results of an Experiment." *Public Opinion Quarterly* 62(3): 378-397.
- Schmidt, William C. 1997. "World-Wide Web Survey Research: Benefits, Potential Problems, and Solutions." *Behavior Research Methods, Instruments, and Computers* 29(2): 274-279.
- Schonlau, Matthias, Ronald D. Fricker and Marc N. Elliott. 2002. *Conducting Research Surveys via E-mail and the Web*. Santa Monica, CA: Rand Corporation.
- Shropshire, Kevin O., James E. Hawdon, and James C. Witte. 2009. "Web Survey Design." *Sociological Methods & Research* 37(3): 344-370.
- Smith, M. A. and B. Leigh. 1997. "Virtual Subjects: Using the Internet as an Alternative Source of Subjects and Research Environment." *Behavior Research Methods, Instruments, and Computers* 29: 496-505.
- Stutzman, F. and Kramer-Duffield, J. 2010. "Friends Only: Examining a Privacy-Enhancing Behavior in Facebook." Paper presented at the ACM Conference on Human Factors in Computing Systems, Atlanta, GA. April 10 – April 15.

- Tourangeau, Roger. 2004. "Survey Research and Societal Change." *Annual Review of Psychology* 55(1): 775–802.
- Tufekci, Zeynep. 2008. "Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites." *Bulletin of Science, Technology & Society* 28(1): 20-36.
- Valliant, Richard and Sunghee Lee. 2005. "Economic Characteristics of Internet Users vs. Nonusers and Implications for Web-Based Surveys." *Webuse & Society* 1(8): 34-51.
- Weible, Rick and John Wallace. 1998. "Cyber Research: The Impact of the Internet on Data Collection." *Marketing Research* 10(3): 19-23.
- Welch, Kevin W. 1981. "An Interpersonal Influence Model of Traditional Religious Commitment." *The Sociological Quarterly* 22(1): 81-92.
- Wilson, Alan and Nial Laskey. 2003. "Internet-Based Marketing Research: A Serious Alternative to Traditional Research Methods?" *Marketing Intelligence & Planning* 21(2): 79-84.
- Witte, James C., Lisa M. Amoroso, and Philip E. N. Howard. 2000. "Research Methodology: Method and Representation in Internet-Based Survey Tools - Mobility, Community, and Cultural Identity in Survey2000." *Social Science Computer Review* 18(2): 179-195.

Figure 1. Message to Group Administrators

Hi, [Administrator's name] -

You don't know me, but I'm a Ph.D. candidate in [department] at [university name], and I'm seeking your help on my dissertation. I'm trying to reach baptized Catholics to see if they'd be interested in taking part in a study. I was hoping you might be willing to send a message to the members of your Facebook group, [group name], with a link to another group: [link to my research group].

I understand that you might be reluctant to help me out, so I'm happy to answer any questions you have about my dissertation.

Thanks!

[my name]

Table 1. Results of Group Solicitation Efforts

<i>Response</i>	<i># Groups</i>	<i>Membership</i>	<i>Recruitment Potential</i>
Sent Message	10	4,614	4,614
Posted Link	5	32,849	32,849
<u>No Response</u>	<u>27</u>	<u>119,848</u>	<u>0</u>
Totals	42 ^a	157,311	37,463

^aForty-three administrators were contacted but one group was misclassified. The group appeared to be comprised of Catholics but it was not.

Figure 2. Where Did the Survey Takers Come from?

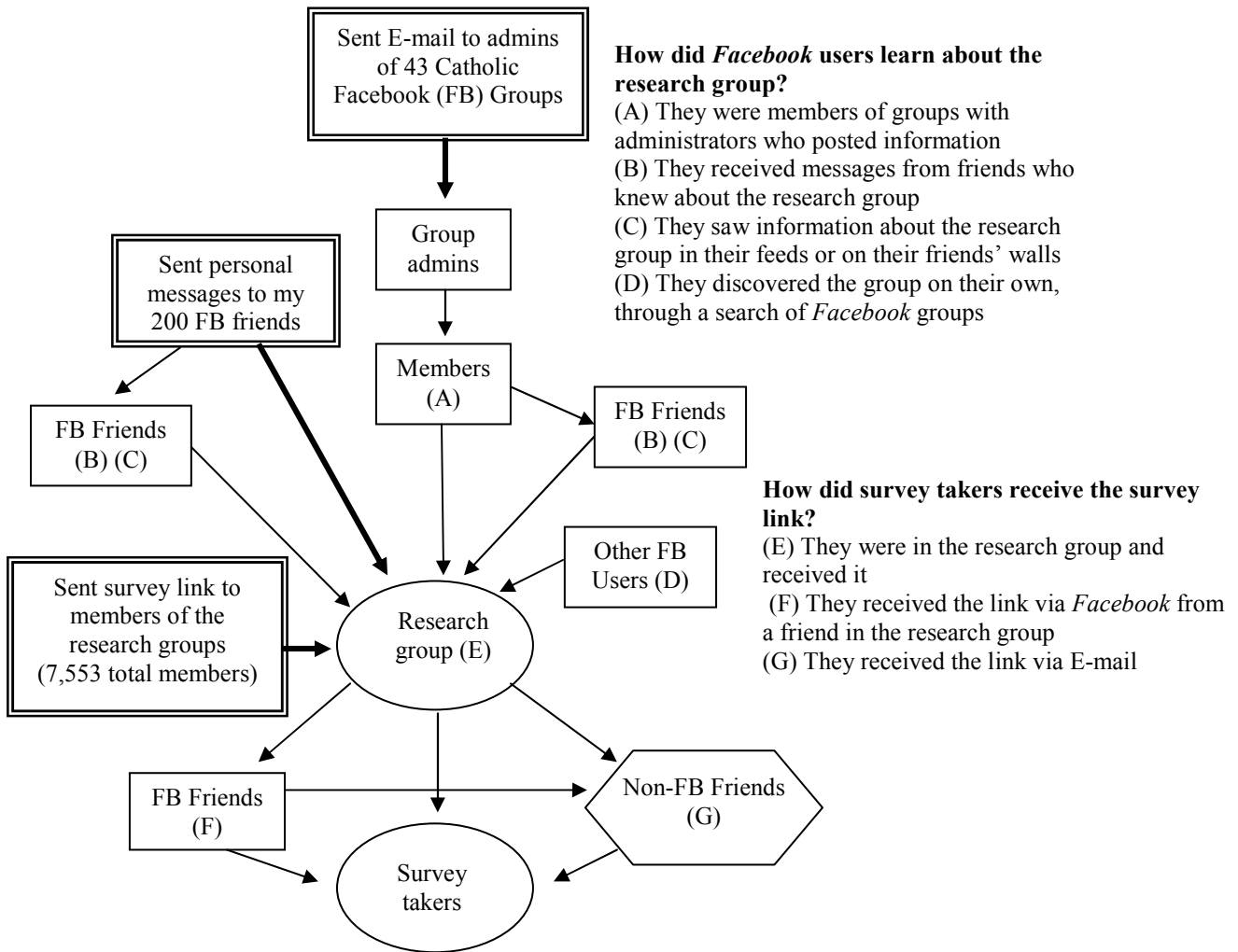


Figure 3. Message to Research Group Members

Hello, group members!

At long last, the survey is ready! As a reminder, to be eligible, you must either:

- 1) have been baptized into the Roman Catholic Church in the United States, or
- 2) have converted to Roman Catholicism in the United States.

This INCLUDES both former Catholics and inactive Catholics.

If you have further questions about the study or eligibility requirements, check out my post on the group's discussion board ([web address]).

Click here to access the survey:

[survey link]

And please pass the link along to other Catholics – both current and former. To send it to friends who are NOT on Facebook, just copy and paste the link to an email.

Thank you!

[my name]

Figure 4. Number of Completed Surveys during First Month of Survey Release

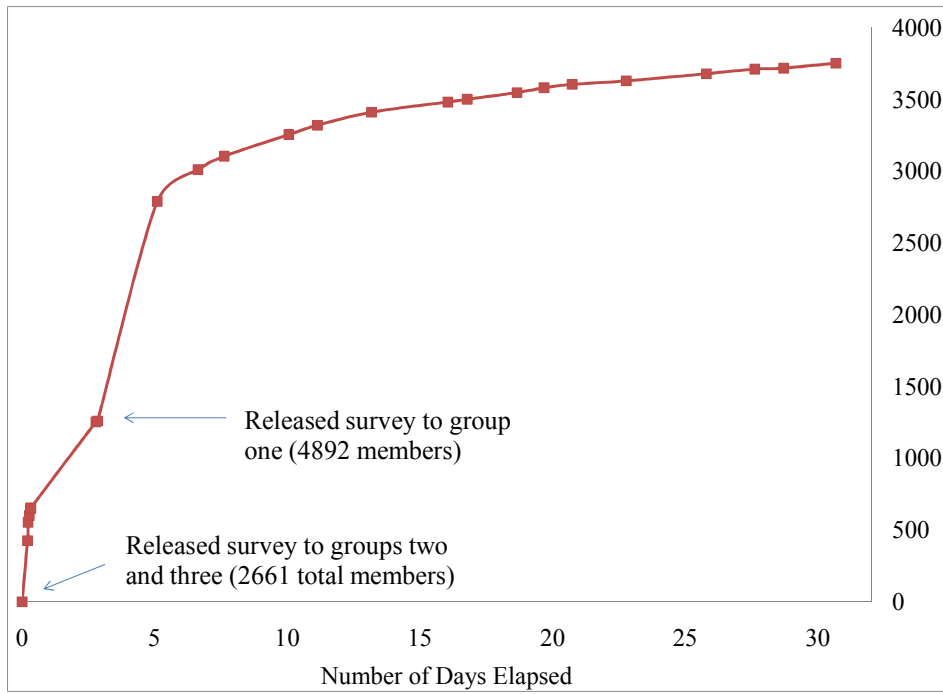


Table 2. Sample Characteristics: *Facebook* sample vs. GSS 2008
(Adults raised Catholic)

	Facebook <i>sample</i> (N=3598)	GSS (N=631)
Demographics		
Mean Age	30.5	44.7
% Female	69.9	53.0
% Latino	5.7	32.5
% College Grad	66.6	23.6
Mass Attendance (%)		
Never or seldom	13.8	42.0
Sometimes	20.3	31.3
Once/week	42.5	22.4
<u>>Once/week</u>	<u>23.4</u>	<u>4.3</u>
Total	100.0	100.0

Table 3. Attendance at Religious Services by Gender: *Facebook* Samples vs. GSS 2008
(Adults raised Catholic)

	Facebook			GSS	
	Females (N=2516)	Males (N=1082)		Females (N=343)	Males (N=288)
Never or seldom	12.9%	16.1%	Never or seldom	39.7%	44.6%
Sometimes	21.5%	17.5%	Sometimes	31.8%	30.8%
Once/week	44.0%	38.8%	Once/week	22.6%	22.2%
<u>>Once/week</u>	<u>21.6%</u>	<u>27.6%</u>	<u>>Once/week</u>	<u>6.0%</u>	<u>2.4%</u>
Total	100.0%	100.0%	Total	100.1%	100.0%

Note: column totals may not equal 100.0 percent due to rounding

Table 4. Correlations with Current Religious Attendance

	Facebook	'05 Gallup	'04,'06,'08 GSS ^d	
Catholics should... ^b	Have a Catholic wedding	0.29***	0.20***	
	Attend Mass each week	0.44***	0.38***	
	Help the poor	0.21***	0.13***	
	Believe in Transubstantiation	0.24***	0.16***	
	Registered in parish	0.31***	0.40***	
	Believe Catholicism is more true	0.46***	0.25***	
	Childhood Salience ^c	0.07***		0.21***
	Catholic parents	-0.02		-0.04
	Catholic spouse	0.28***		0.16*
	Age	0.01	0.30***	0.26***
	Sex	0.02	-0.03	-0.08***
	Education	-0.06***	-0.06+	0.10***
	Latino	-0.03	-0.02	-0.10***
Presence of children	0.03+		0.14***	

+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

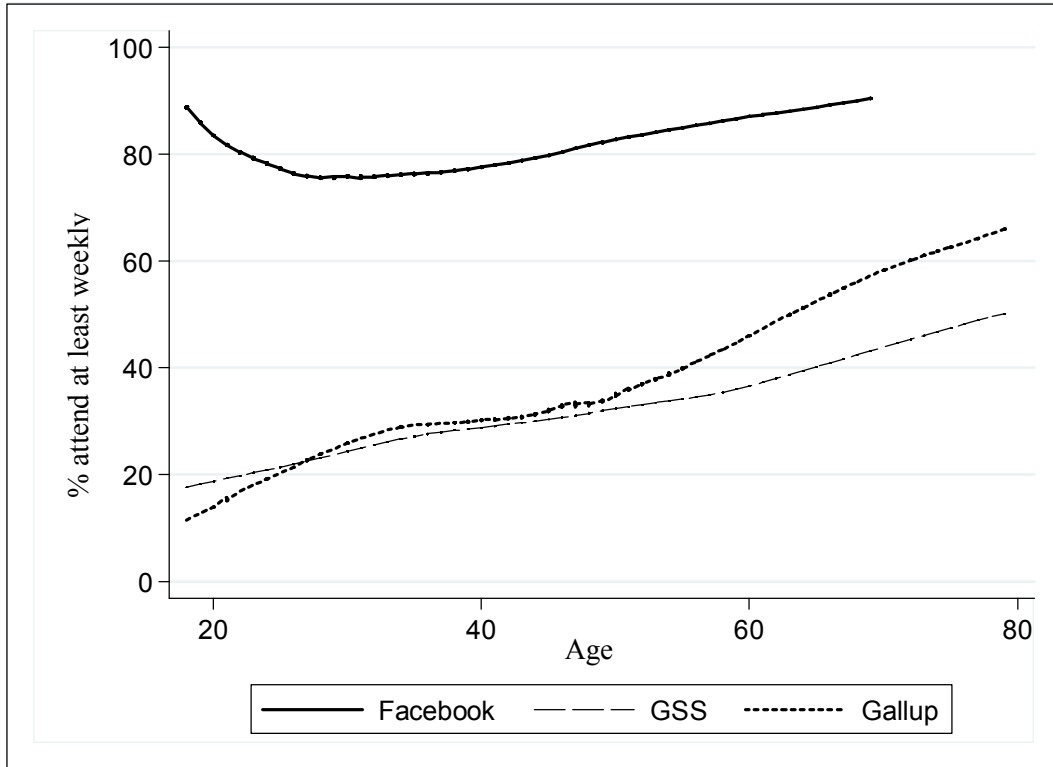
^aTo make the data comparable to the Gallup study, *Facebook* and GSS analysis restricted to those who currently identify as Catholic.

^bThe *Facebook* survey posed these questions somewhat differently than did Gallup. For each item, Gallup asked "Please tell me if you think a person can be a good Catholic without performing these actions or affirming these beliefs." Response categories are "yes," "no," "don't know," or "refuse." The *Facebook* sample asked respondents to indicate the extent to which they believe Catholics are obliged to engage in these acts or hold these beliefs.

^cIn the GSS, childhood salience is measured by frequency of Mass attendance at age 12; in the *Facebook* sample, it is measured by response to the question, "How important was religion in your family when you were growing up?" Studies show these measures of childhood religious salience correlate very highly.

^dChildhood salience, Catholic parents, and Catholic spouse were only asked in the 2008 GSS.

Figure 5. Relationship* between Weekly Mass Attendance and Age: *Facebook*, Gallup, and GSS Samples



*Smoothed means created in Stata using lowess