

## ABSTRACT

Most researchers view statistical results based on non-random or low response-rate samples with skepticism at best and contempt at worst. However, the degree to which departures from random sampling and high response rates actually bias results is, in principle, an empirical question. An extensive, newly available archive of “mail panel” surveys (with response rates, as conventionally calculated, below 10 percent) allows us to examine this question. We compare the results from this archive with the results from similarly worded questions asked of a conventional high-response random sample, subjecting the mail panel data to validation tests of maximal stringency. The unmistakable conclusion is that this archive is a valuable source of reliable data for the social sciences, providing detailed evidence about many facets of social change over the last quarter century. On topics ranging from church attendance to world affairs to leisure time activities to feminism to financial concerns, the mail panel yields results essentially identical to those from high-quality random samples. When accompanied by careful diagnostics of the kind illustrated in this paper, non-probability samples can provide a legitimate source of survey information. Taken in conjunction with widespread trends toward lower response rates, our results suggest the timeliness of a broader exploration, both empirical and theoretical, about when and why low-response or non-random samples sometimes work.

**HOW IMPORTANT IS RESPONSE RATE?**  
**AN EVALUATION OF A “MAIL PANEL” SURVEY ARCHIVE**

**ROBERT D. PUTNAM**

Isabel and Peter Malkin Professor of Government and Public Policy  
John F. Kennedy School of Government, Harvard University  
E-mail: [robert\\_putnam@harvard.edu](mailto:robert_putnam@harvard.edu)

**STEVEN YONISH**

Department of Political Science, University of Wisconsin-Madison  
E-mail: [styonish@hbs.edu](mailto:styonish@hbs.edu)

It is an article of faith among social scientists that, to ensure the unquestionable validity of survey results, one should use random sampling with a high response rate to identify respondents. Indeed, the entire edifice of classical econometric analysis (significance levels, and so forth) is built on the assumption of randomness in sampling. But have we overstated the importance of high-response-rate random samples and underestimated the value of survey results achieved by other means? A newly available archive of “mail panel” surveys allows use to examine this question. While mail panel surveys – which are frequently used by commercial polling firms – use random sampling at some stages of the selection process, participants are essentially self-selected in the first stage. Such self-selection necessarily raises the distinct possibility that response bias will undermine the utility of the sample to the extent that the few people motivated to participate differ significantly from the many that were not motivated to respond. In this paper, we compare the results of several public-opinion questions contained in an extensive commercial data archive with the results achieved on similarly worded questions included in a conventional random probability sample. Three criteria establish the basis

for this inquiry: the distribution of responses on comparable questions, the trends in the variables over time, and the pattern of demographic correlates for the variables. We conclude with a discussion of the potential benefits of commercial mail panel data for scholarly research on public attitudes and behavior, and some general reflections on the circumstances under which non-probability samples may be of value in the social sciences.

## SUBSTANTIVE CONTEXT

The principal aim of this methodological inquiry is to provide a sound statistical basis with which to draw substantive inferences about issues of civic engagement in contemporary American society. Are Americans less connected with their communities than a generation ago or not? Putnam (1995a; 1995b) argued that civic engagement, social trust, and social capital have eroded over the last several decades. Others in both the popular and professional press have contradicted this claim, sometimes relying on anecdotal, but telling counter-examples (e.g., Lehmann 1996), sometimes citing statistical evidence to the contrary (e.g., Ladd 1996), and sometimes pointing to methodological pitfalls in Putnam's evidence (e.g., Smith 1997). This is not the place for a complete recital and evaluation of this debate<sup>1</sup>, but certain methodological maxims are relevant to sorting out the evidence:

*In assessing social change, two observations are better than one, and many are better than two.* No assessment of hypotheses about change is possible with data from a single point in time.<sup>2</sup> Data from two points in time offer some analytic leverage for

---

<sup>1</sup> For one useful collection, see Dionne (1997).

<sup>2</sup> Experts who know better have sometimes violated this elementary precept; see, for example, Kohut (1997).

testing claims about monotonic change, but are vulnerable to measurement inconsistency at either data point; a single measurement “error”—a subtle change in question order, for example—might lead to a mistaken judgment about the overall trend. Change measured at multiple points in time (particularly if the observed change is monotonic) is exponentially more reliable; if a given variable increases steadily from time 1 to time 4, say, three distinct “errors” in successive surveys or some steadily changing source of “error” would need to be posited to explain away the increase.

*In assessing social change, be attentive to possible changes in the accuracy of the measuring rod over time.* Some of Putnam’s original evidence (including the eponymous measure of league bowling) was vulnerable to the charge that the validity of the measure itself might have deteriorated over time. If bowling were gradually replaced by softball or soccer as the leisure sport of choice among Americans, then a decline in team bowling might simply have been offset by a rise in softball or soccer, both intrinsically team sports. More generally, if the keeping of membership records were itself the mark of organizational maturity, then any sample of such rosters over time might over-represent “senescent” organizations and under-represent emerging ones. The more generic the measure (e.g., membership in “organizations” rather than membership in the Elks club), the less its susceptibility to this critique, but if specific probes are used (as in the standard question about organizational membership in the General Social Survey), then it remains possible that newer *types* of organizations will be under-reported.<sup>3</sup>

Since social change often proceeds unevenly, *careful attention must be paid to matching measurement periods to hypotheses about patterns of change.* Claims about change over a period of several decades cannot accurately be tested by evidence from a

time span of a few years, and *vice versa*. Putnam's hypothesis posited change since the 1960s, whereas the most extensive critique (Ladd 1996) indiscriminately mixed data about "change" over a year or two and "change" over half a century. A fair test of the "Bowling Alone" hypothesis requires comparable data over as much of the last three decades as possible.

*In assessing social change, multiple indicators are better than a single indicator.*

Putnam's initial essay drew on evidence from organizational membership rosters, surveys of organizational membership, surveys of attitude change (especially social trust), and (more briefly) survey-based studies of political behavior and time diary studies. Since life is not lived in a single dimension, it is unlikely that all variables will move in the same direction at the same time, and discrepancies among multiple indicators may sometimes be as instructive as consistencies among them. We examine one illustration of this point below.

*To assess explanations for social change, data sets that are both broad (in topical coverage) and deep (in temporal coverage) are essential.* Aggregate behavioral change tends to be glacially slow and to be influenced by multiple factors. Individual-level correlates of social and political behavior at one point in time provide, at best, murky clues to patterns of change across time. Thus, there is a premium on large data sets with sufficient coverage to allow simultaneous testing of multiple hypotheses about sources of change within a single data set.

No single source of data exists that avoids all these methodological problems. However, long-term repeated surveys of national samples using multiple, constant, generic questionnaire items are distinctly more attractive than any alternative. For this

---

<sup>3</sup> For an early critique along these lines, see Baumgartner and Walker (1988).

good reason, much of the “Bowling Alone” debate has turned on responses to the General Social Survey, and (secondarily) the National Election Study series. However, as useful and rigorous as these two sources are, they are limited in their coverage of relevant facets of social behavior. Given its traditional focus on national election campaigns, the NES includes few measures of everyday civic involvement in its core of repeated questions. Coverage in the GSS is somewhat broader, but still limited. In addition to several standard questions on social trust, the most relevant question for our purposes is an item about membership in fifteen different types of organizations, from church groups to sports associations and from labor unions to hobby and garden clubs. Nevertheless, it is widely agreed that this standard question has distinct limitations as an all-purpose measure of civic engagement.

1. It measures only formal membership, not activity or involvement. Casual “check-book” or “mailing list” associations, which embody no real connections among individual members, are mixed in some unknown degree with intimate clubs that hold regular weekly meetings.<sup>4</sup>
2. It measures only the number of group *types* to which a respondent belongs, not the number of groups.<sup>5</sup>
3. It ignores the myriad of less formal types of involvement in social networks and community affairs.

---

<sup>4</sup> See Verba, Schlozman, and Brady (1995) on this distinction.

<sup>5</sup> For an early critique along these lines, see Baumgartner and Walker (1988).

In the absence of other large-scale longitudinal surveys, however, the debate about civic engagement has rested disproportionately on this single question in the General Social Survey.

## NEW SOURCE OF EVIDENCE

Recently, we have uncovered an important and hitherto unexploited source of annual survey evidence on civic and social activities extending back more than two decades: DDB Needham Life Style Panel Surveys (DDB). This newly available survey archive sheds much light on the “Bowling Alone” debate, in particular, by its inclusion of a wide range of social and community activities and attitudes. More generally, DDB data provide measures of many other aspects of social and political behavior, and thus they constitute an important new source of evidence that can be exploited by social scientists with more broadly defined substantive interests than ours. To the extent that it can be shown to be methodologically reliable, the DDB Needham Life Style archive constitutes one of the richest known sources of data on social change in America in the last quarter of the twentieth century.

By way of introduction, the DDB Needham advertising agency, each year since 1975, has commissioned Market Facts, a commercial polling firm, to question a national panel of American households about their consumer preferences and behavior.<sup>6</sup> Most of the roughly 20 page written questionnaire is taken up with inquiries about detergents and mutual funds, automobiles and stereos, and so on. However, every year a core set of

---

<sup>6</sup> We are grateful to Dhavan Shah, a former graduate student at the University of Minnesota, and his instructor, Professor William Wells, for alerting us to the existence of the DDB Needham-Market Facts surveys, to Doug Hughes and his colleagues at DDB Needham, particularly Chris Callahan, for generously making the data available for analysis, and to Sid Groeneman and his colleagues at Market Facts for

questions has been posed about “life style” issues, including such things as media usage, financial worries, social and political attitudes, self-esteem, and a wide range of social behavior, such as reading, travel, sports and other leisure activities, family life, and community involvement.

From the point of view of DDB Needham’s commercial clients, these “life style” questions are valuable for planning marketing strategies, defining market niches, and drafting advertising copy. Are churchgoers more likely to send Christmas cards, for example? Are fast food restaurants replacing the family dinner for two-career families? Are frequent moviegoers more liberal in their social attitudes? Are rock concert fans more likely than museum buffs to watch Monday Night Football?<sup>7</sup> From the point of view of social science, however, the DDB data provide an unparalleled source of information on trends in social behavior over the last two decades. Though our primary use of the data is to infer the substantive outlines of these trends, we first need to conduct a thorough review of potential methodological pitfalls. Precisely because of the potential utility of these data to other social scientists with substantive interests quite different from ours, our effort to ensure the validity of DDB data may be of general interest to the social science community. Moreover, in an era of declining response rates in most public opinion surveying<sup>8</sup>, it is perhaps useful to ask, as we do here, *what differences do low response rates make to substantive findings?*

---

gracious help in understanding the methodology used and its potential advantages and disadvantages. For a useful overview of the origins of “life style” market research, see Wells (1974 and 1975).

<sup>7</sup> The answer to all these questions is “yes.”

<sup>8</sup> For recent evidence on declining response rates, see Robert M. Groves and Mick P. Couper, *Nonresponse in Household Interview Surveys* (New York: Wiley, 1998), 155-187; John Goyder, *The Silent Minority: Nonrespondents on Sample Surveys* (Cambridge, U.K.: Polity Press, 1987), esp. 64; John Brehm, *The Phantom Respondents: Opinion Surveys and Political Representation* (Ann Arbor: University of Michigan Press, 1993), and Joop J. Hox and Edith D. de Leeuw, “A Comparison of Nonresponse in Mail, Telephone, and Face-to-Face Surveys,” *Quality & Quantity* 28 (November 1994): 329-344. For skepticism about the

## METHODOLOGICAL FAILINGS OF DDB NEEDHAM LIFE STYLE DATA

The DDB data come not from random population samples, but from a form of quota sampling called “mail panels.” The process begins when Market Facts acquires from commercial list brokers the names, addresses, and sometimes demographic characteristics of very large numbers of Americans—from drivers’ license bureaus, telephone directories, and many other sources. Large samples from these lists are then invited by mail to express willingness in advance to respond periodically to mail and phone inquiries about commercial products and services, as well as other current issues.<sup>9</sup> According to Market Facts officials, the rate of favorable response to such invitations varies considerably across different sectors of the population—from perhaps less than one percent among racial minorities and inner city residents to perhaps 5-10 percent among middle-aged, middle class, “middle Americans.” From this pre-recruited “mail panel” (numbering perhaps 500,000 at any one time) are then drawn demographically balanced samples<sup>10</sup> for the annual DDB Needham Life Style surveys (as well as hundreds of other commercial and other surveys throughout the year). Each Life Style respondent is mailed a long written questionnaire that he or she is asked to complete and return within several weeks. At this stage the response rate (c. 70-80 percent) is typically higher than for conventional random samples. So far as we have been able to ascertain, there

---

significance of trends in non-response, see Tom W. Smith, “Trends in Survey Non-Response,” *International Journal of Public Opinion Research* 7 (1995) 157-71.

<sup>9</sup> Respondents are occasionally offered a nominal gift—a packet of “Post-It’s” and a tiny tote bag, for example—for completing a particularly burdensome questionnaire.

<sup>10</sup> Groeneman (1994; emphasis in the original) reports that the sample is “drawn to approximate actual distributions of household income, population density, panel member’s age, and household size *within the 9 Census divisions.*” In an effort to achieve an appropriately balanced final sample, the starting sample is adjusted within each of the subcategories of race, gender, and marital status to compensate for expected differential returns. Weights are then applied to the actual respondents to match the demographic

has been no substantial change in these procedures over the last two decades, although less careful procedural records have been kept than would be characteristic of comparable academic archives, and in particular, systematic data on the rate of favorable responses to the initial mail invitations are lacking.

Compared to conventional random samples, the mail panel approach to consumer polling has a number of significant advantages, especially in terms of cost efficiency, but for scientific purposes it has several important potential drawbacks.

1. Because the initial recruitment is by mail, literacy in English is an essential requirement, and thus the bottom of the educational ladder is underrepresented, as are non-English speakers.
2. Effective response rates are much lower among racial minorities.<sup>11</sup>
3. Adults under 25 are slightly underrepresented, probably because their mobility makes them harder to track.

Social traits that are especially common in those sections of the population are thus also underrepresented in the DDB sample. An assessment of the overall representative quality of DDB data is presented in Table 1. Before turning to the table, however, we must note a particular sampling quirk in the DDB data. Until 1985 only married households were included in the sample. As such, Table 1 compares DDB and GSS data on a number of demographic categories as far back as 1986, the year in which DDB first

---

composition of the final sample to the target population. Questionnaires are mailed to roughly 5000 respondents; usable responses are generally received from an average of 3500-4000 respondents.

<sup>11</sup> This is also true for conventional sampling, but the disparity is greater for mail panels.

included both non-married men and women in annual samples.<sup>12</sup> In round numbers, it contains 10 percent too few people without a high school degree, 10 percent too few single respondents, 10 percent too many parents, and half as many racial minorities. Moreover, there is some indication that the DDB sample under-represents those in the highest and lowest categories of family income. In demographic terms, in short, these data reasonably represent the upper 80-90 percent of American society, but they do not well represent ethnic minorities, the very poor, the very rich, and the very transient.<sup>13</sup> They may also slightly over-represent the portion of the public that is most engaged with the mass media.<sup>14</sup>

Devoid of any substantive context, one is very hard pressed to make a definitive assessment of how such sample biases damage substantive claims drawn from these data. While studies of non-response succeed in drawing attention to target populations who are more likely to refuse participation in surveys (see, for instance, DeMaio, 1980, and Heberlein and Baumgartner, 1978), they fail to demonstrate how response bias undermines statistical results. The identification of compositional biases is a necessary

---

<sup>12</sup> For those interested in the substantive implications of this sampling quirk, we have found virtually no cases in which the observed trends between 1985 and 1998 differ significantly between married and single respondents, although in a number of cases there are modest but significant differences in the levels. Where the levels and/or trends in traits of interest vary by marital status, therefore, we have made some adjustment to track changes over the entire 1975-1998 period. For example, married people attend church more often than single people do, while singles attend club meetings more often than married people, but the trends in church- and club- going are essentially identical for both groups. In all cases where this sampling peculiarity poses potential problems of analysis, we analyzed the data separately by marital status.

<sup>13</sup> Though the questions are not exactly comparable, there is evidence that DDB data contains 10 percent too many homeowners, as compared to GSS data. There is also some evidence that the under-sampling of the less educated has been somewhat reduced in more recent years.

<sup>14</sup> Apart from the few exceptions noted in this paragraph, the general consistency between DDB and GSS samples is not very surprising in view of the DDB sample selection procedure. Starting samples are meticulously designed to be representative demographically. They balance on multiple factors simultaneously using a program that selects households in such a way that their marginal distributions on each of the balancing factors approximate the five respective target distributions. The program's algorithm is akin to that used in marginal weighting (or rim weighting). Lack of accurate representation appears

first step in the process of determining the utility of a particular sample, but it is not sufficient. Just as it would be reckless to proceed directly to data analysis, blithely sidestepping the potentially harmful consequences of identifiable sample biases, it is foolhardy to summarily reject a non-probability sample without sufficient knowledge of how such biases influence results in a particular substantive context.<sup>15</sup>

The classic example of the *Literary Digest* poll in the 1936 presidential election illustrates the necessity of analyzing response bias within a specific substantive context. The 1936 election was between Republican Alf Landon and Democratic incumbent Franklin D. Roosevelt. The *Literary Digest* sent out 10 million ballots to subscribers, people in the phone book, and people whose names were listed in auto registry records and voter registration records. The return rate was 23% and, due to the particular sample frame, the sample of respondents contained a disproportionate number of people of higher socioeconomic status. Despite uniformity in the respondent selection mechanism, the *Literary Digest* poll performed respectably in elections before 1936. However, the 1936 poll predicted that Roosevelt would garner 43% of the vote when, in reality, he won by a landslide with 62% of the vote. Clearly, the underprediction of the Democratic vote was directly attributable to the underrepresentation of working-class respondents in the sample. More to the point, the *Literary Digest* sample proved inadequate for election forecasting only in 1936 because the selection mechanism became highly correlated with the dependent variable. Specifically, Roosevelt's New Deal coalition in 1936 raised to an

---

problematic for race/ethnicity and the very youngest adult age group (18-24). This information is based on personal communication with Sidney Groeneman.

<sup>15</sup> Sophisticated statistical techniques to correct for non-response, or selection, bias are not applicable in our case because we lack background information on non-respondents. As such, we cannot employ advanced strategies for modeling the selection process, such as those used in Brehm (1993) and Dubin and Rivers (1989), and thus fall back on a decidedly less elegant approach to the determination of whether significant non-response bias exists.

unprecedented level the salience of social class in the presidential voting. Put simply, the poll results were biased because those who responded were significantly different (more Republican) than those who did not.

In a very similar fashion, the most important potential pitfall with the DDB Needham data archive comes in the possible introduction of substantive bias in the initial decision to participate in the mail panel. While the critical shortcoming of the *Literary Digest* poll in 1936 was the socioeconomic bias in the sample selection procedure, the validity of the DDB sample is doubly open to question; first, due to the compositional biases mentioned above and, second, because of the self-selection procedure for initial entry into the mail panel. The self-selection procedure raises the suspicion that, all else equal, those who choose to participate in a mail panel have a lower motivational threshold to engage more generally in other forms of participation. As such, there is a danger that the sample selection procedure is highly correlated with what we are trying to measure, namely decisions to engage in social and civic activities.<sup>16</sup> For instance, Goyder (1987) maintains that the degree of social responsibility felt by a prospective sample respondent may be affected by the degree of social cohesion and by the legitimacy of societal institutions. Similarly, an element of social validation is an integral part of the decision to participate in a survey to the extent that “one should be more willing to comply with a request to the degree that one believes similar others would comply with it” (Groves et al., 1992: 483). Yet even more relevant to our substantive framework is the notion that the act of responding to a survey conveys theoretically

---

<sup>16</sup> For a very limited range of relatively highbrow activities (going to the symphony, opera, ballet, art museums and other theater), Robinson (1989) found that a commercial Harris poll overreported participation in the arts due to disproportionate non-response from less-educated respondents. This is an

meaningful information. Some have argued that theories with predictive value for volunteering, voting, and other more mainstream activities should also apply to filling out questionnaires. Specifically, “the circumstances under which individuals come to the assistance or fail to help another should apply in returning a questionnaire” (Heberlein and Baumgartner, 1978: 461).

In short, there is abundant justification for skepticism in a priori assessments of the validity of DDB data for analysis of social and civic engagement. Relying on a highly self-selected non-probability sample, the data collection method departs from standard survey methodology in a way that seems to invalidate their utility for the study of mass behavior. This potential liability of DDB data looms even larger in light of the significant length of the DDB questionnaire, for we would expect the greater investment of time required for participation to have non-random consequences for the sample (Groves et al., 1992). In the absence of a full census of social behavior—something that not even the US Census Bureau believes in any more—the two key questions here are:

1. Do people who are willing to join a mail panel differ in relevant ways from people who are willing to respond to conventional surveys?
2. Has the degree of difference between the Life Style panel and conventional surveys changed over time, thus rendering judgments about trends suspect?

If the answer to question (1) is “yes,” then in some respects the DDB Needham Life Style data may misrepresent key parameters. Only if the answer to question (2) is also “yes,”

---

excellent example of how bias can undermine results when the behavior in question is “sensitive to relaxation in respondent selection procedures” (Robinson, 1989: 411).

however, will the trends in the Life Style data misrepresent trends as they would appear in a conventional random survey. A constant bias would be disconcerting, of course, but only a changing bias would affect our judgments about trends.

#### VALIDATION OF DDB NEEDHAM LIFE STYLE DATA

With respect to the quality of mail panel respondents, reassuring information is available from several studies that have directly compared results from mail panels and conventional samples. First, Groeneman (1994) reports that, apart from the demographic disparity described above (fewer young, poor, and racial minorities in the mail panel), there are surprisingly few differences between the two approaches, even on variables that might be thought to be especially sensitive to the difference in technique.<sup>17</sup> The samples do not differ in religious affiliation and religiosity, in public policy views (on tax policy, abortion, gun control), in their views about their own and the nation's economic circumstances, in their altruism (volunteering, philanthropy) or general "positivity," in their basic consumer orientations, purchasing habits, ownership or use of common products, in their health or fitness, or leisure time.<sup>18</sup> The only significant differences are (1) partisanship (mail panels are slightly less Democratic, probably because of the underrepresentation of racial minorities) and (2) media usage (mail panelists watch slightly more television and read slightly more newspapers).<sup>19</sup> The suggestion that low response rate may not bias substantive results can also be inferred from a recent study that compared results from "easy-to-reach" and "hard-to-reach" samples (Kohut, 1998).

---

<sup>17</sup> Groeneman (1994) compared panel and non-panel samples using data obtained from Market Facts mail panel and random-digit-dialing surveys. The panel component does not necessarily include respondents in the DDB data archive.

<sup>18</sup> Similar findings, though focused exclusively on product preferences, are reported in NFO research (n.d.).

Aside from clear differences on racial issues, the upshot of the comparison between respondents who initially agreed and those who at first refused is that there are no nontrivial differences on other issue stances, on quantity of media use, on engagement in daily activities, and on feelings about other people.

With respect to the reliability of mail panels, Visser et al. (1996) compare mail surveys conducted by a local newspaper in Columbus with data obtained from standard telephone interview procedures to predict the outcome of statewide elections in Ohio. To a remarkable degree, the mail panels have been consistently more accurate during the period from 1980 to 1994. In addition to the larger sample sizes, other factors contributing to the greater accuracy of mail panels include privacy, anonymity, and pacing advantages. Specifically, the self-administered mail questionnaires may be “less subject to social desirability biases” than interviews in which social interaction could threaten privacy and anonymity (Visser et al., 1996: 212). Moreover, the self-administered nature of filling out a mail questionnaire affords maximum convenience to a respondent, presumably reducing survey fatigue.

The demonstration of cross-sectional reliability of mail panel data provided by Visser et al. (1996) does not speak, however, to the legitimacy of using mail panels to track temporal changes. Fortunately, one of the two major national surveys of consumer confidence in the United States (from the University of Michigan) relies on conventional random sampling, whereas the other (from the Conference Board) has relied on a mail panel since its inception in the late 1960s. The specific questions used in the two studies also differ, as do the methods of index construction, so not all discrepancies between the

---

<sup>19</sup> The discrepancy in party identification, though statistically significant, is very slight. In 1996 the NES found 39 percent Democrats, 28 percent Republicans, and 33 percent Independents; in that same year the

two series could be attributable to the differences in sampling. Nevertheless, as Figure 1 shows, the long-run changes charted by the two methods have been very similar. (To minimize scaling differences, we have standardized the scores on each index. The semi-annual correlation between the two indices over more than three decades is  $R^2 = .55$ .) For fine-grained, month-to-month changes one or the other of these two surveys might be preferred, but the broad-gauged impressions of annual trends that one would glean from the two are quite similar.

To explore more fully the reliability of DDB data, we have taken advantage of the fact that this data set includes questions comparable to those asked mostly on a regular basis in the General Social Survey over roughly the same time span. With the lack of uniformity in question wording and temporal coverage between the two data sources, however, those questions with nearly identical wording and greater temporal coverage in both surveys, provide a superior basis for validation tests. As our approach is a criterion comparison, we recognize all too well the fact that any discrepancies we uncover are not due solely to non-response (or self-selection), but may come from other sources such as item unreliability, question wording, and interviewer error, just to name a few (Smith, 1983). In our case, the question wording differences figure most prominently so we will provide additional information, when necessary, on any differences in wording that may affect our comparison of comparable questions. These question items include attitudes toward the legalization of marijuana and abortion, feminism, fear of the Soviet Union,

---

Life Style sample reported 37 percent Democrats, 31 percent Republicans, and 32 percent Independents.

financial worries, military service, post-materialism, church attendance, and leisure activities.<sup>20</sup> For each of these items, we ask three questions:

1. Do the *levels* of response on these variables differ between the two samples, taking into account obvious differences in question wording?
2. Do the *trends* that one would infer about the underlying trait differ between the two samples?
3. Do the underlying patterns of *demographic correlates* of these variables differ between the two samples?

To anticipate our conclusions, in every case the answer is “no.” For purposes of describing and explaining this range of attitudes and behavior, at least, the two surveys are virtually indistinguishable, despite marked differences in both the sampling (random vs. quota) and questioning procedure (personal interviews vs. mail questionnaires).

Consider, first, support for the legalization of *marijuana*, for which the two surveys happen to use virtually identical question wording.<sup>21</sup> Figure 2 shows that the overall level of support revealed in the two surveys is remarkably similar.<sup>22</sup> Even more

---

<sup>20</sup> We also examined somewhat comparable questions on social trust, union power, and gun ownership. The results were consistent with those reported here. Both the rates and trends of social trust are very similar in the two studies. Although the questions on gun ownership in the two surveys were quite different, and the question on union power has been posed only three times in the GSS, the basic trends are quite comparable in the two surveys. Put simply, we have found no paired questions in the two surveys that would call into question the essential comparability of the two data sets. That is, we have not singled out for presentation here comparisons that support our conclusion.

<sup>21</sup> The DDB question is a six-point agree-disagree scale based on the following statement: Marijuana use should be legalized. The GSS question asks the following: “Do you think the use of marijuana should be made legal or not?” We collapse the DDB responses into a binary agree-disagree variable to match the GSS format.

<sup>22</sup> Because marital status is not strongly correlated with support for legalization of marijuana in either survey, we do not here present the data disaggregated by that factor, in order to avoid unnecessary graph clutter. Doing so has no effect on our conclusions.

striking, the two surveys describe exactly the same pattern over time—declining support for legalization during the early 1980s, a plateau from 1987 to 1990-91, and sharp increases thereafter.<sup>23</sup> Additional support for the face validity of DDB data is contained in Table 2 where we present levels of support for specific demographic categories.<sup>24</sup> Looking at the figures in the first two columns, one notices the reasonable similarity in demographic patterns and is struck by the absence of any glaring discrepancies. A more systematic evaluation (see Table 3) shows that when a large number of demographic characteristics (age, year, gender, education, region of the country, marital status, parental status, income, employment status, race, , homeownership, and type of dwelling) are used to predict support for the legalization of marijuana, the marginal effect of each independent variable, *controlling simultaneously for all the others*, is essentially the same in the two data sets.<sup>25</sup> Specifically, in the multivariate analysis, the regression coefficients for virtually all of the independent variables are of the same sign and relative magnitude in the two data sets. The only exception, in terms of the demographic correlates, is for education, which has a stronger impact in the GSS sample, yet the sign of the DDB coefficient is the same. Similarly, the GSS coefficient for the year of survey suggests that the temporal decline in support for legalization, all else equal, is greater than that indicated by the DDB coefficient, yet again this amounts to a difference in degree, not in kind, as the sign of the coefficients match. Moreover, these favorable

---

<sup>23</sup> Interpolating the missing years in the GSS, the annual correlation between the two series is  $R^2 = .78$ .

<sup>24</sup> These figures represent means within each demographic category across the full time span for which comparable data are available (1981-1996).

<sup>25</sup> The coefficients in Table 3 are estimated using ordinary least squares for the entire time span using pairwise deletion in order to maximize the number of cases included in the analysis. In light of the large number of cases, particularly for the DDB estimation, the conventional level of significance ( $p < 0.05$ ) may establish an overly generous standard for hypothesis testing. That is, the greater efficiency of the estimation procedure should necessitate an increase in the statistical standard used to interpret the results

results are achieved using a very conservative approach, as we do not control for known compositional irregularities in the Life Style surveys (specifically race and income), and thus our approach is biased against our finding of few, if any, differences.

The pattern of demographic correlates in DDB and GSS for marijuana legalization thus appears remarkably similar over a two-decade period. Still, it would surely be a damaging blow to the reliability of DDB data to discover that the annual estimates of the demographic correlates fluctuate wildly. To examine this issue, we estimate the model in Table 3 separately for each year and measured the temporal variability of each demographic measure in DDB and GSS. In *every* case, the temporal variability of the DDB estimate is smaller than that of the GSS estimate, and generally half the size.<sup>26</sup> All told, our comparison of the pattern of demographic correlates and their temporal variability establishes an extremely stringent test of comparability, particularly for data sets produced by different organizations, procedures and for different purposes. It implies that the deep structure of conditional relations among all these variables is quite similar in the two data sets. For example, according to the GSS, the probability in 1990 that a 35-year-old single white mother, who has two years of college education and was working part-time and renting an apartment in a middle-sized New England city, favored marijuana legalization is 35 percent, whereas the comparable probability for DDB data was 38 percent, a difference well within sampling error.

Another basis for comparison is available through questions asked in each survey about support for legalized *abortion*. The GSS asks separately about support for legal

---

(Raftery, 1996). As such, consistent with Raftery (1996), we use a t-value of 3.09 to evaluate the significance of any coefficient difference between GSS and DDB.

abortion under seven different circumstances (fetal defects, rape, and so on), culminating with a question (used here as our standard of comparison) about abortion “if the woman wants it for any reason.” The Life Style survey asked respondents to express their views of the proposition “I support legalized abortion” on a six-point agree-disagree scale. It seems reasonable to compare support for the “abortion on demand” position posed in the GSS with the more intense levels of agreement to the less demanding Life Style question. As Figure 3 shows, the full level of pro-choice support registered in the Life Style surveys is consistently 12-14 percentage points higher than that of GSS, a difference largely attributable to those respondents who most carefully qualify their agreement to the Life Style question and would probably have dissented from the GSS “abortion on demand” question. That predictable difference apart, however, the annual ups and downs of pro-choice sentiment are remarkably parallel in the two data sets.<sup>27</sup> Moreover, a look at some basic demographic patterns of support for legalized abortion (see Table 2) reveals no alarming discrepancies, with the sole exception being the greater level of support for legalized abortion among DDB seniors. In the multivariate analysis, once again, the deep structure of demographic correlates in the two surveys is also essentially identical, with only a few minor exceptions.<sup>28</sup> In short, the history and demography of attitudes toward abortion reflected in these two surveys is very similar.

---

<sup>26</sup> We were skeptical at first of this result, suspecting that the smaller temporal variability of DDB estimates may be due to the presence of significant carryover in the DDB mail panel. However, these suspicions were dismissed upon learning from DDB sources that the panel is refreshed annually.

<sup>27</sup> Interpolating the missing years in the GSS, the annual correlation between the two series is  $R^2 = .79$ .

<sup>28</sup> The two largest exceptions are (1) controlling for all the other 22 demographic variables, having school-aged children is associated with 5 percent less support for legal abortion in the GSS survey, compared with a 10 percent reduction in the Life Style survey, and (2) controlling for all the other 22 demographic variables, living in a single family dwelling is associated with a 3 percent reduction in pro-choice sentiment in the Life Style survey, but makes no significant difference in the GSS. In no case is the sign of the relationship between pro-choice sentiment and any given independent variable different in the two samples.

For measuring *feminism*, two somewhat different questions have been used in the two surveys. The Life Style poll asks simply for agreement or disagreement that “a woman's place is in the home,” whereas the GSS poses the view that “women should take care of running their homes and leave running the country up to men.” The more assertive GSS stimulus consistently triggers about 10-15 percent more disagreement than the more prosaic Life Style question. Once again, however, the trend lines over time are parallel in the two surveys, as shown in Figure 4, and once again, the demographic correlates are quite similar in the two surveys.<sup>29</sup>

For measuring attitudes to *communism and Russia*, the respective questions in the two surveys are formulated quite differently, so that direct comparisons of the absolute levels of antipathy are impossible. In the GSS, respondents were asked to rate how much they “liked” or “disliked” Russia on a ten-point scale. In the Life Style surveys, respondents were asked to respond to the following proposition, using a six-point agree-disagree scale: “Communism is the greatest peril in the world today.” Despite these substantial differences in wording, however, as Figure 5 shows, the relative trends in the underlying sympathy or antipathy toward communist ideology and its earthly embodiment over these two decades from Brezhnev to Andropov to Gorbachev to Yeltsin appear highly similar in the two surveys.<sup>30</sup>

For tapping *financial worries*, unfortunately, the measures available in the two surveys are quite different. In the GSS respondents are simply asked, “We are interested

---

<sup>29</sup> Interpolating the missing years in the GSS, the annual correlation between the two series is  $R^2 = .88$ .

<sup>30</sup> Interpolating the missing years in the GSS, the annual correlation between the two series is  $R^2 = .81$ . Because this pair of dependent variables is scored so differently, it is not possible to match demographic correlates in detail. However, the strongest correlates are virtually identical in the GSS and Life Style surveys: fear of communism and dislike of Russia are greatest in the earlier years of the surveys, among older cohorts and less educated respondents, in the South, and outside the largest cities.

in how people are getting along financially these days. So far as you and your family are concerned, would you say that you are pretty well satisfied with your present financial situation, more or less satisfied, or not satisfied at all.”<sup>31</sup> In the Life Style surveys, we combined responses to four closely related agree-disagree items into a single factor score “Financial Worries Index” (reversing the scoring on items 3 and 4):

1. Our family is too heavily in debt.
2. No matter how fast our income goes up we never seem to get ahead.
3. We have more to spend on extras than most of our neighbors do.
4. Our family income is high enough to satisfy nearly all our important desires.

The absolute scores on these two quite different measures are, of course, impossible to compare directly. However, in Figure 6 we present evidence on trends between 1975 and 1997 separately for married and single respondents in the two surveys. In both surveys, single respondents express significantly more financial worries than married respondents, and the trends over time for the two groups are broadly similar in both surveys. Both surveys show sharply increasing financial concerns from the late 1970s until the bottom of the recession in 1982, both show some increase in financial optimism during the boom of the middle 1980s<sup>32</sup>, both show a resurgence of financial concern during the Bush years, and both show a slightly widening gap in financial worries between single and

---

<sup>31</sup> For statistical purposes, we coded “not at all” as 2, “more or less” as 1, and “pretty well” as 0.

<sup>32</sup> For reasons that are not clear, the Life Style data suggest a deeper level of financial worry among married people during the stagflation of the 1970s than is shown in the GSS data. In addition, the Life Style data imply that the resurgence of financial worries among married people began in 1986, whereas it did not show up in the GSS until 1988. These are the only significant discrepancies in the trend lines for the two series. Interpolating the missing years in the GSS, the annual correlation between the two series is  $R^2 = .55$  for single respondents and  $R^2 = .17$  for married respondents.

married respondents after 1990. Despite the substantial differences in the underlying questions, these parallel lines appear striking confirmation that the two samples respond in comparable ways to national economic developments.<sup>33</sup>

A rather different sort of cross-validation comes from answers to a directly comparable question about past military service, asked in the General Social Survey fourteen times between 1974 and 1994 and asked annually in the Life Style survey between 1983 and 1988. Because military service typically occurs only at a well-defined stage in the life cycle, we can aggregate across all male respondents in each of the two survey series, and estimate the fraction of any given birth cohort (e.g., men born in 1945) who served in the armed forces. Among men born in the first seven decades of this century, the fraction of veterans has varied from more than 80 percent of men born in 1924-26 (that is, those who turned 18 in 1942-44) to less than 10 percent of those born in 1966-68 (and who thus turned 18 in 1984-86). Figure 7 shows the astonishingly parallel annual estimates of this fraction from the General Social Survey and the Life Style surveys.<sup>34</sup> Once again, it is essentially impossible to distinguish the results of the two polls.

Validation in a quite different substantive domain comes from questions from Ronald Inglehart's long-standing project on value change in "post-materialist" societies.

---

<sup>33</sup> Because this pair of dependent variables is scored so differently, it is not possible to match demographic correlates in detail. However, the strongest correlates are virtually identical in the GSS and Life Style surveys: Financial worries are greatest among those with low real income (unsurprisingly!), tenants (instead of homeowners), younger cohorts, parents with children at home, African-Americans, mobile home dwellers, residents of major metropolitan areas, and people in the labor force (as opposed to retirees and housewives). The only significant contrast between the two data sets is that, with all else controlled, the GSS finds the lowest financial worries in the South, whereas the Life Style survey finds the lowest financial worries in the north-central states, but in neither case are these regional differences as great as the correlations listed previously.

<sup>34</sup> The correlation of the two estimates of veterans' status across the 74 annual cohorts of men born between 1895 and 1968 is  $r=.92$ .

Inglehart's standard battery of questions invites respondents to choose among various values or "national priorities."

- Maintain order in the nation
- Give people more say in government decisions
- Fight rising prices
- Protect freedom of speech
- A stable economy
- Progress toward a less impersonal and more humane society
- Progress toward a society in which ideas count more than money
- The fight against crime

Figure 8 compares responses to these values in the 1997 and 1998 DDB Needham Life Style surveys with responses to similar items in the 1990 and 1995 World Values Survey. Not only are the relative priorities very similar, but the absolute values are remarkably close, especially given the likelihood of genuine change during these years. For example, the successive surveys show a gradual decline in the priority given to inflation control during the 1990s, but virtually the same path is traced in the DDB and WVS surveys (results not shown). In other words, despite the differences in the two types of sample, essentially the same profile of social values appears in both surveys.

#### THE SPECIAL CASE OF CHURCH ATTENDANCE

Church attendance is of special interest as a measure of our validation of the DDB data, since it is itself one measure of community involvement. In this case, the question

wording in the two surveys is essentially identical: “How often do you attend religious services?” in the GSS and “How often during the past twelve months have you attended church [or other place of worship]?” in the Life Style surveys.<sup>35</sup> Unfortunately, the ranked response options are somewhat different in the two surveys, so in order to allow more precise comparisons, we have converted each rank to an interval number representing the estimated annual church attendance, as shown in Table 4. Reasonable observers might differ over exactly what “several times a year,” for example, means in quantitative terms, but our basic results are not sensitive to exactly what integers are assigned to the various ranks.<sup>36</sup>

Figure 9 then charts the average annual church attendance reported in the two surveys across these two decades. Once again, the absolute level of church attendance suggested in the two surveys is similar, as are the trends over time.<sup>37</sup> Both the Life Style and GSS samples report average church attendance of about twice a month (22-25 times-per-year), and each exhibits a modest, steady decline of roughly 10-15 percent over these two decades. Looking at some basic demographic patterns for all years combined, we see in Table 5 that the results are strikingly similar, with no substantial departures for any age, educational, or regional categories. Once again—and here we present the detailed systematic evidence in Table 6, given the substantive centrality of this measure of community involvement—the patterns of multivariate demographic prediction are very

---

<sup>35</sup> The words “or other place of worship” were added to the standard question from 1996 onwards.

<sup>36</sup> Across entire GSS sample, our interval scores correlate  $r=.993$  with those generated by an entirely independent algorithm that we discovered after defining ours (Hout and Greeley, 1998, footnote 10), though our algorithm generates mean scores 4 percent higher than theirs, primarily because we assumed that “several times a year” meant 6, on average, whereas they assume it meant 3.

<sup>37</sup> Recall that Groeneman (1994), using identically phrased questions, found identical levels of church attendance in a mail panel sample and a random sample. In both the GSS and the Life Style surveys the time trend is statistically highly significant, although the rate of decline is greater in the Life Style data. On

similar in these two surveys. This table may be read as follows: controlling for the more than two dozen variables represented in the table, the GSS data suggest that women attend church exactly 5.3 more times-per-year than men, whereas the DDB data suggest that the difference between the sexes in church going is 4.8 times-per-year. To take another example, the GSS data suggest that parents of school-aged kids attend church 2.6 more times-per-year than non-parents, whereas the DDB estimate is 3.2 times-per-year. In other words, these two data sets provide very similar estimates of even highly specific features of American religious behavior.

Along lines similar to our earlier analysis of marijuana legalization, a more rigorous comparison of the pattern of demographic correlates reveals few exceptions to the overall degree of conformity. As before, the significant exceptions are largely differences of degree and do not result from any changes in the direction of impact. For example, the coefficients for year of birth suggest, in keeping with conventional wisdom, that successively younger generations are generally disinclined to attend religious services. Though the extent of the decline is significantly stronger in the DDB sample, both archives concur that a real, substantial, and generationally-based decline has occurred. The same logic applies to the significant differences in the impact of education and home ownership. Slightly less benign are the comparative results for income and homemaker. In these cases, the demographic measure is significant in DDB but not in GSS, implying substantively that DDB homemakers, all else equal, are more church-going and that higher income DDB respondents, all else equal, are less so.

---

the other hand, the annual changes in the two series are not highly correlated. Interpolating the missing years in the GSS, the annual correlation between the two series is  $R^2 = .11$ .

Such a state of affairs would be particularly injurious to the reliability of DDB data if these effects demonstrated temporal volatility. In fact, as with the temporal stability analysis for marijuana legalization, the DDB estimates exhibit a remarkable degree of stability. The sole exception, in results not shown, is the higher instability of the DDB coefficient for education. Whereas all other DDB variables display greater temporal stability, and many with standard deviations less than half of those for GSS, the temporal volatility of the DDB estimate for education is only 4% higher than that of GSS. In all, the pattern of demographic correlates in DDB reveals two slight anomalies, for income and homemakers, but the greater temporal stability of DDB estimates suggests that inferences about DDB trends are not threatened by a source of bias that intrudes intermittently. Despite the differences in polling organization, sampling and polling procedures, and response alternatives, one would draw the same general conclusions about the level, trend, and demographic distribution of religious observance from these two surveys.<sup>38</sup> Taken as a whole, this evidence on religious observance provides added confirmation for our basic methodological assessment—the image of social change represented in the Life Style surveys is for the most part very similar to the image that emerges from the General Social Survey.

#### A LIMITED COMPARISON OF LEISURE ACTIVITIES

Another topic of considerable relevance to community involvement is participation in leisure activities. While the Life Style surveys provide comprehensive, annual data on a

---

<sup>38</sup> One minor, but revealing exception involves race. *Ceteris paribus*, Black Americans attend church 4.2 more times-per-year than whites according to the GSS, but only 2.8 more times-per-year according to the Life Style surveys. This modest discrepancy is intelligible, since, in effect, Blacks who volunteer for the Life Style surveys are economically and culturally less distinctive than Blacks in the more representative GSS. This is a useful reminder that we should be cautious about generalizing to racial minorities from the Life Style surveys.

wide range of leisure activities since 1975, GSS coverage is confined to a narrower range of activities for select years. Table 7 summarizes a final cross-validation of the Life Style surveys from a series of questions posed in the General Social Survey only in 1993, and a subset of those in 1998, regarding a series of “leisure or recreational activities ... done in the past twelve months.” Most of the activities asked about in this series of questions were also posed in a similar form in the 1993 and 1998 Life Style survey.<sup>39</sup> The incidence of these activities as reported in the two surveys was astonishingly similar, well within the limits of sampling error. How many Americans went to the movies last year? GSS says 70 percent, DDB says 72 percent. How about hunting and fishing? GSS says 37 percent, DDB says 37 percent. How about classical concert-going? GSS says 16 percent, DDB says 17 percent. In other words, the profiles of leisure activities represented in the mail panel of the Life Style survey and in the random sample of the GSS were essentially identical.<sup>40</sup>

---

<sup>39</sup> In a few cases, the GSS question combined activities—for example, hunting and fishing—that were listed separately in the Life Style survey and in such cases we have adjusted the Life Style responses appropriately. In a few other cases, the equivalent question was posed in the Life Style survey not in 1993, but in some adjacent year, as indicated in the footnotes to Table 4. Separately, we have compared trends over time in the GSS and the Life Style surveys on hunting. The question wording differs slightly, and thus the absolute levels do as well, but the trends are virtually identical. GSS: “Do you go hunting?” Life Style: “How many times did you go hunting last year?” Similarly, the Life Style time trend with respect to “There should be a gun in every home” tracks very closely the time trends on two related (but not identical items in the GSS on attitudes and behavior with respect to guns: “Do you happen to have in your home any guns or revolvers?” and “Would you favor or oppose a law which would require a person to obtain a police permit before he or she could buy a gun?” Moreover, the pattern of demographic predictors of attitudes to gun ownership in the two surveys are very similar—pro-gun attitudes are strongest among men, in small towns and rural areas, in the southern and mountain states, married homeowners, infrequent church-goers, and older and less educated people, controlling in all cases for these and other demographic characteristics.

<sup>40</sup> Additionally, we have compared Life Style results in 1982 and 1984 with a simultaneous, roughly comparable set of responses from monthly Roper surveys (aggregated annually) on the frequency of activities such as having dinner at a restaurant, attending a movie, and attending a sporting event. The Roper questions are phrased in terms of “did you happen to engage in this activity this last week?” whereas the Life Style questions are phrased in terms of “how often last year did you engage in this activity?” When we convert the Roper “last week” responses to “times per year” (simply by multiplying by 52), the results are virtually identical to the Life Style responses: dinner at restaurant: 19 times per year for each; movies: 5 times per year for each; sports event: 4 times per year in the Roper surveys, 5 times per year in the Life Style surveys.

Additional validation evidence on leisure activities is provided in Figures 10 and 11. The extent of home video usage from 1984-1997 is tracked in Figure 10. This graph compares the aggregate trend for the DDB question on number of times per year one rents or purchases a video tape with the number of hours the average person spends watching home videos.<sup>41</sup> Though the metrics of the two measures are quite different, the level and trend since 1984 exhibit a stunningly precise fit. Only slightly less impressive is the comparison in smoking rates between DDB and GSS shown in Figure 11. Again, despite minor variation in the exact question wording, the levels of smoking indicated by each data source are very similar. Moreover, both trends show a modest overall decline since 1984, the first year in which DDB included the smoking question in the Life Style survey.

## DISCUSSION

The primary motivation for this methodological foray is a concern with the temporal reliability of the DDB data archive. The virtues of their considerable time span and topical coverage would seem to offer limitless promise for research in the social sciences. More particularly, they shed enormous light on many features of social and civic engagement at the heart of the social capital debate. Specifically, the DDB data archive contains a vast array of social interaction measures, ranging from working on community projects to entertaining guests at home. Moreover, apart from measures occupying the central core of any analysis of social capital, the archive includes measures of family life, work life, personal finances, health and political views. If and when these data enter the

---

<sup>41</sup> Data on home video usage are taken from the *Statistical Abstract of the United States* (Table 884).

public domain, they will permit substantively rich, methodologically sophisticated longitudinal research on topics rarely covered elsewhere.

The general principle guiding our analysis in this paper was to subject the DDB data to validation tests of maximal stringency. Much more work along these lines was conducted but not reported here, but the unmistakable conclusion is that the DDB archive is an extremely valuable and valid source of data for the social sciences. As additional researchers explore these newly available data, more subtle biases will no doubt be uncovered, as is virtually always true for novel data sets. Because the DDB data set includes standard questions on a number of social and economic topics, we are confident that other scholars, more industrious than we, will find additional opportunities for cross-validation of the DDB data with other scientifically reputable time-series surveys.<sup>42</sup>

The pitfalls of using these data, as we have demonstrated, are few and far between. Without exception, the distribution of responses on comparable questions in DDB and GSS match very closely. And with few minor exceptions, the pattern of demographic correlates for comparable questions is alike. Thus, with scant substantively significant exceptions, the conclusions that one could draw from the Life Style survey are virtually identical to those that one could draw from the more familiar and scientifically rigorous General Social Survey, at least where we can find roughly comparable measures. That reliability fortifies our confidence about the validity of Life Style evidence on aspects of social behavior for which we lack any other systematic evidence about change over the last several decades and for which the Life Style surveys provide uniquely deep coverage. That is, the substantive and temporal coverage of available random national

samples is insufficient to address our specific concerns about *trends* in community involvement. Moreover, our substantive agenda does not pivot on accurate *snapshot* estimates of social and political behaviors, a task for which random samples with high response rates are ideally suited. Under these specific and rather common circumstances, when particular issues of social, political and economic processes or changes are under study, a researcher should be cautiously optimistic about the use of sub-optimal non-random samples.

Why we found only negligible differences between the Life Styles surveys and the General Social Survey remains an open question. Our earlier discussion provided ample reason to expect a low degree of commonality between the two data sources. Rather, the similarities were quite unmistakable. In the main, the pattern of results suggests that any response bias present in the Life Style survey does not correlate with any of a wide range of substantive measures that we analyzed. In the absence of a rich literature on why this may be the case, we can only speculate on possible features of the Life Styles survey that may have mitigated the potential for bias. First, we know that the Life Style samples were selected from a pre-recruited panel, yet with an attempt made to carefully balance each annual sample using multiple balancing targets. Second, the overriding purpose of the Life Style survey may have minimized the opportunity for respondents to self-select on issues of social and political participation. Of much greater salience in the Life Style survey are consumer attitudes and behavior, and these attributes may not be associated one way or another with propensities to engage in community activities. Moreover, with such a lengthy questionnaire covering a multitude of issues,

---

<sup>42</sup> In February 1997 the Life Style survey added a question about the legalization of same sex marriages; 74 percent disapproved. In a comparable question posed by the Gallup Poll in November 1996 72 percent of

the choice to participate may not depend on how strongly you feel about a single subject. In this case, multiple sources of self-selection may be offsetting one another. We hope these speculations will provide some impetus for a more general theoretical inquiry on why low-response, or non-random, samples sometime seem to work. The development of such a theoretical framework could go some way towards making errors from low-response more predictable.

In short, we wish to place on the agenda for methodological research this central question: *Under what circumstances and with respect to what attributes does response rate actually affect substantive conclusions?* Our results suggest that non-response error may not be as uniformly threatening to valid inference as often supposed. At minimum, the striking similarity in substantive inferences drawn from the Life Styles surveys and GSS demonstrates that a non-random selection process will not always produce fatal bias. The real issue, of course, is not “high-response, random samples: yes or no?” The real issue is when response rate bias is likely to be a serious threat to substantive inference.

At one extreme, for example, when seeking a point estimate of the incidence of some hitherto unstudied attitude or behavior, it is surely prudent to rely on conventional random samples with the maximum possible response rate. On the other hand, when tracking secular trends in familiar domains, where standard benchmarks from random samples are readily available (e.g., voting turnout or social trust), mail panel data may be much more reliable than scholars have previously assumed, and the substantial cost savings from this technique might sensibly be invested elsewhere. We also emphasize the need to look more carefully at how particular sources of low response rate might interact with particular variables. For example, we suspect that the DDB Needham Life Style

---

those respondents expressing an opinion disapproved. See Gallup (1998).

survey archive might be unreliable on racial attitudes, given its severe under-sampling of the black underclass, although the archive itself contains no relevant measures that would enable us to test this hypothesis. In many cases, however, among all sources of survey error, low response rate or non-random respondent selection may be relatively benign compared to other design features.

With scant scholarly attention so far to how various sources of survey error interact to bias substantive inferences, scientific judgment on our speculation awaits further theoretical development and empirical evidence on circumstances under which response rates have a crucial impact. Nonetheless, when accompanied by careful diagnostics of the kind illustrated in this paper, non-probability samples can be a legitimate source of information.

## Bibliography

- Baumgartner, Frank and Jack L. Walker. 1988. "Survey Research and Membership in Voluntary Associations." *American Journal of Political Science* 32:908-27.
- Brehm, John. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.
- DeMaio, Theresa. 1980. "Refusals: Who, Where and Why." *Public Opinion Quarterly* 44:223-233.
- Dionne, E.J. 1997. (Ed.) "Civil Society: What is it? Why is everybody talking about it?" *The Brookings Review* (Fall):2-41.
- Dubin, Jeffrey A., and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods and Research* 18:360-390.
- Gallup Organization. 1997. "From the Editor: Gallup Guide to Contemporary American Sexual Norms: Homosexual Relations." December.  
[http://www.gallup.com/POLL\\_ARCHIVES/9712edc.htm#hmmor](http://www.gallup.com/POLL_ARCHIVES/9712edc.htm#hmmor)
- Goyder, John. 1987. *The Silent Minority: Nonrespondents on Sample Surveys*. Boulder, CO: Westview.
- Groeneman, Sid. 1994. "Multi-purpose household panels and general samples: How similar and how different?" Paper presented at the annual meeting of the American Association for Public Opinion Research, Danvers, Massachusetts.
- Groves, Robert M., Robert Cialdini, and Mick Couper. 1992. "Understanding the Decision to Participate in a Survey." *Public Opinion Quarterly* 56:475-495.
- Hadaway, C. Kirk, Penny Long Marler and Mark Chaves. 1993. "What the Polls Don't Show: A Closer Look at U.S. Church Attendance." *American Sociological Review* 58:741-752.
- Heberlein, Thomas, and Robert Baumgartner. 1978. "Factors Affecting Response Rate to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review* 43:447-462.
- Kohut, Andrew. 1998. "Conservative Opinions not Underestimated, but Racial Hostility Missed." Washington, D.C.: Pew Research Center on the People and the Press.
- Kohut, Andrew. 1997. "Trust and Citizen Engagement in Metropolitan Philadelphia: A Case Study." Washington, D.C.: Pew Research Center on the People and the Press.
- Ladd, Everett C. 1996. "The Data Just Don't Show Erosion of America's 'Social Capital.'" *The Public Perspective* 7 (June/July):5-22.

- Lehmann, Nicholas. 1996 "Kicking in Groups: Alleged decline of America's communal capital," *Atlantic Monthly* (April):22-27.
- NFO Research. n.d. (c. 1990). "The Comparability of Consumer Panel and Random Sampling Procedures." Toledo, Ohio.
- Putnam, Robert D. 1995a. "Bowling Alone: America's Declining Social Capital," *Journal of Democracy* 6 (January):65-78.
- Putnam, Robert D. 1995b. "Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America," *P.S.: Political Science and Politics* 28 (December):1-20.
- Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." In *Sociological Methodology 1995*, ed. C. C. Clogg. Oxford: Basil Blackwell.
- Robinson, John P. 1989. "Survey Organization Differences in Estimating Public Participation in the Arts." *Public Opinion Quarterly* 53:397-414.
- Smith, Tom W. 1997. "Factors Relating to Misanthropy in Contemporary American Society," *Social Science Research*. 26:170-196.
- Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.
- Visser, Penny, Jon Krosnick, Jesse Marquette, and Michael Curtin. 1996. "Mail Surveying for Election Forecasting: An Evaluation of the *Columbus Dispatch* Poll." *Public Opinion Quarterly* 60:181-227.
- Wells, William D. 1974. (ed.) *Life Style and Psychographics*. American Marketing Association.
- Wells, William D. 1975. "Psychographics: A Critical Review." *Journal of Marketing Research* 12:196-213.

Table 1

**Demographic Profile of Life Style and GSS samples, 1986-1997**

	DDB	GSS
AGE (mean)	46.5	44.1
<b>MARITAL STATUS</b>		
Married	72.1%	60.6%
Divorced or Separated	9.5	11.8
Never married	10.9	20.4
Widowed	7.5	7.2
<b>FAMILY INCOME</b>		
Under \$10,000	8.8%	12.9%
\$10,000 – 24,999	27.0	26.8
\$25,000 – 39,999	25.4	22.3
\$40,000 – 59,999	21.8	17.6
Over \$60,000	16.9	20.4
<b>EDUCATIONAL ATTAINMENT</b>		
Less than High School	3.0%	8.1%
Attended High School but no degree	7.2	11.9
High School degree	35.2	34.0
Some College but no 4-year degree	28.5	25.8
College Degree (4 year)	13.2	14.1
Advanced Degree	13.0	6.1
<b>MINORITY STATUS</b>		
African-American	6.5%	12.0%
<b>GENDER</b>		
Female	54.1%	54.9%
<b>PARENTAL STATUS</b>		
Children at Home	50.2%	40.4%
<b>WORK STATUS</b>		
Employed Full-time	57.1%	51.7%
Employed Part-time	8.9	11.6
Retired	16.0	11.9
Keeping House	12.1%	15.0
Sources: General Social Survey, 1986-1997; DDB Needham Life Style consumer surveys, 1986-1997		

**TABLE 4**  
**Quantifying Life Style and GSS church attendance responses**

GSS RESPONSE ALTERNATIVES	IMPUTED SCORE	PERCENT 1972-96	LIFE STYLE RESPONSE ALTERNATIVES	IMPUTED SCORE	PERCENT 1975-97
Never	0	14	None in the past year	0	24
Less than once a year	0.5	8	1-4 times	2	15
Once a year	1	14	5-8 times	6	6
Several times a year	6	13	9-11 times	10	5
Once a month	12	7	12-24 times	18	7
2-3 times a month	30	9	25-51 times	38	16
Nearly every week	40	6	52+ times	54	28
Every week	52	22			
More than once a week	60	8			
		100			100

**TABLE 6**  
**Predicting average annual church attendance**

	GSS		DDB		Difference (Life Style - GSS)	Std. Error Of Difference	t-value for difference
	B	Std. Error	B	Std. Error			
(Constant)	255.29	55.16	429.36	29.24	174.08	62.43	2.79
Year of survey	0.05	0.03	0.04	0.02	-0.01	0.04	-0.25
Year of birth	-0.18	0.02	-0.26	0.01	-0.08	0.02	-4.16
Sex (women)	5.31	0.47	4.79	0.21	-0.52	0.51	-1.03
Married? (Y/N)	1.19	0.66	2.90	0.41	1.71	0.78	2.20
Divorced/Separated? (Y/N)	-3.40	0.77	-4.50	0.52	-1.10	0.94	-1.17
Widowed? (Y/N)	-0.07	0.98	1.15	0.59	1.22	1.15	1.06
Typically work full time	0.67	0.73	1.31	0.41	0.63	0.84	0.76
Typically work part time	3.10	0.92	5.00	0.49	1.90	1.04	1.83
Retired	1.40	1.00	1.30	0.49	-0.10	1.11	-0.09
Homemaker	0.96	0.86	4.02	0.45	3.06	0.98	3.14
Children at home (0=N)(1=Y)	2.57	0.49	3.18	0.21	0.62	0.53	1.16
Education (6-level index)	1.46	0.18	2.36	0.08	0.90	0.20	4.60
Family income (quintiles)	0.01	0.19	-0.71	0.08	-0.72	0.20	-3.58
African Americans	4.17	0.62	2.75	0.42	-1.42	0.75	-1.91
New England	6.30	1.06	4.34	0.45	-1.96	1.15	-1.70
Mid Atlantic	4.78	0.76	5.80	0.33	1.02	0.83	1.23
East North Central	6.58	0.72	8.15	0.32	1.56	0.79	1.98
West North Central	9.26	0.91	9.69	0.40	0.43	0.99	0.43
South Atlantic	6.08	0.74	6.59	0.33	0.50	0.81	0.62
East South Central	8.55	0.96	10.62	0.43	2.07	1.05	1.96
West South Central	7.79	0.88	7.84	0.37	0.05	0.96	0.05
Mountain	5.69	1.03	4.88	0.46	-0.81	1.13	-0.71
Owns residence	3.79	0.56	1.55	0.29	-2.24	0.63	-3.57
Dwelling Apartment	0.60	2.01	-1.72	0.43	-2.32	2.06	-1.13
Dwelling mobile home	-3.12	2.12	-2.24	0.49	0.88	2.18	0.41
Dwelling Single family house	0.97	1.98	1.68	0.35	0.72	2.01	0.36
Adjusted R-squared	0.083		0.096				
Standard Error	20.54		21.64				
N	10518		59430				

For dummy variables, the omitted categories are as follows:

- Work status: Student, unemployed, and other
- Race: Whites and other
- Region: Pacific Coast
- Size of place: Small-town or rural area
- Dwelling: Condominium or other

Sources: General Social Survey (1973-1996); DDB Needham Life Style consumer surveys (1975-1997)

**TABLE 7**  
**Leisure activities during preceding 12 months (1993, 1998)**

GSS wording / <i>Life Style wording</i>	1993		1998	
	GSS	Life Style	GSS	Life Style
Went out to see a movie in a theater / <i>Went to the movies</i>	72%	70%	67%	73%
Recorded a TV program so you could watch it later / <i>Videotaped a TV program on a VCR</i> <sup>43</sup>	63%	70%		
Grew vegetables, flowers, or shrubs in garden / <i>Worked in garden</i>	62%	68%		
Participated in any sports activity, such as softball, basketball, swimming, golf, bowling, skiing, or tennis / <i>Played softball and/or went swimming and/or played golf and/or went bowling and/or went skiing and/or played tennis</i> <sup>44</sup>	59%	69%		
Attended an amateur or professional sports event / <i>Attended a sporting event</i>	56%	56%		
Went camping, hiking, or canoeing / <i>Went camping and/or went hiking</i> <sup>45</sup>	44%	44%		
Visited art museum or gallery / <i>Visited an art gallery or museum</i>	41%	47%	37%	44%
Made art or craft objects, such as pottery, woodworking, quilts, or paintings / <i>Worked on a crafts project (needlework, etc.)</i> <sup>46</sup>	41%	48%	38%	58%
Went hunting or fishing / <i>Went hunting and/or went fishing</i>	37%	37%		
Played a musical instrument like a piano, guitar, or violin / <i>Played a musical instrument</i>	24%	23%	24%	20%
Went to classical music or opera performance / <i>Went to classical concert</i>	16%	17%	17%	15%
Attended auto, stock car, or motorcycle race / <i>Went to auto race</i> <sup>47</sup>	16%	9%		

Sources: General Social Survey, 1993,1998; DDB Needham Life Style consumer surveys, 1992-1994,1998

<sup>43</sup> Life Style data are available for 1988-1991 only. Figure here is for 1991.

<sup>44</sup> Since the Life Style questionnaire asked about each of these sports separately, in effect six separate probes were employed. This difference almost certainly inflated the Life Style results, relative to the single GSS question.

<sup>45</sup> Hiking was included in Life Style surveys in 1975-1984 and 1996-97; figure for 1993 was interpolated. Canoeing was never included in Life Style surveys.

<sup>46</sup> Life Style data are available for 1994-1997 only. Figure here is projected for 1993.

<sup>47</sup> Attendance at auto races was included in Life Style surveys only in 1997, and the figure for that year is given in Table 4.

**TABLE 3**  
**Predicting average annual support for marijuana legalization**

	GSS		DDB		Difference (Life Style - GSS)	Std. Error Of Difference	t-value for difference
	<b>B</b>	Std. Error	<b>B</b>	Std. Error			
(Constant)	5.46	1.28	-2.23	0.59	-7.69	1.41	-5.45
Year of survey	-0.007	0.001	-0.002	0.000	0.005	0.001	4.53
Year of birth	0.005	0.000	0.004	0.000	-0.001	0.000	-1.92
Sex (women)	-0.08	0.01	-0.04	0.00	0.03	0.01	2.86
Married? (Y/N)	-0.07	0.02	-0.08	0.01	-0.02	0.02	-0.96
Divorced/Separated? (Y/N)	0.03	0.02	-0.01	0.01	-0.04	0.02	-1.69
Widowed? (Y/N)	-0.003	0.02	-0.05	0.01	-0.05	0.03	-1.80
Typically work full time	-0.04	0.02	-0.04	0.01	0.01	0.02	0.24
Typically work part time	-0.03	0.02	-0.05	0.01	-0.02	0.02	-0.83
Retired	-0.03	0.02	-0.04	0.01	-0.01	0.03	-0.55
Homemaker	-0.03	0.02	-0.05	0.01	-0.03	0.02	-1.12
Children at home (0=N)(1=Y)	-0.07	0.01	-0.04	0.00	0.03	0.01	2.22
Education (6-level index)	0.03	0.00	0.01	0.00	-0.02	0.00	-5.17
Family income (quintiles)	-0.002	0.004	0.005	0.002	0.01	0.00	1.68
African Americans	0.01	0.01	0.02	0.01	0.01	0.02	0.59
New England	-0.06	0.03	-0.03	0.01	0.03	0.03	1.16
Mid Atlantic	-0.09	0.02	-0.05	0.01	0.04	0.02	1.95
East North Central	-0.11	0.02	-0.08	0.01	0.03	0.02	1.36
West North Central	-0.14	0.02	-0.10	0.01	0.03	0.02	1.42
South Atlantic	-0.13	0.02	-0.05	0.01	0.08	0.02	4.20
East South Central	-0.14	0.02	-0.09	0.01	0.05	0.02	2.21
West South Central	-0.14	0.02	-0.08	0.01	0.06	0.02	2.57
Mountain	-0.09	0.02	-0.05	0.01	0.03	0.03	1.31
Owns residence	-0.01	0.01	-0.03	0.01	-0.02	0.01	-1.11
Dwelling Apartment	0.03	0.05	-0.02	0.01	-0.05	0.05	-0.97
Dwelling mobile home	0.05	0.05	-0.02	0.01	-0.07	0.05	-1.38
Dwelling Single family house	0.03	0.05	-0.03	0.01	-0.06	0.05	-1.24
Adjusted R-squared	0.085		0.042				
Standard Error	0.40		0.41				
N	7315		52834				

For dummy variables, the omitted categories are as follows:

- Work status: Student, unemployed, and other
- Race: Whites and other
- Region: Pacific Coast
- Size of place: Small-town or rural area
- Dwelling: Condominium or other

Sources: General Social Survey (1973-1996); DDB Needham Life Style consumer surveys (1975-1997)

Table 2

**Simple demographic breakdown of support for legalization of marijuana and legal abortions**

	<i>Support for legalization of marijuana</i>		<i>Support for legal abortions</i>	
	DDB	GSS	DDB	GSS
<b>EDUCATIONAL ATTAINMENT</b>				
Less than High School Degree	20	16	28	25
High School degree	20	19	36	37
Some College but no 4-year degree	24	24	44	47
College Degree (4 year) or higher	26	26	51	57
<b>AGE CATEGORY</b>				
18-29	31	28	38	43
30-44	27	25	42	46
45-59	18	18	41	40
60+	15	12	41	32
<b>REGION</b>				
East	24	21	45	48
Midwest	19	19	35	37
South	22	18	37	34
West	28	29	50	51

Sources: General Social Survey, 1982-1996 and DDB Needham Life Style consumer surveys, 1981-1996.

Table 5

**Simple demographic breakdown of church attendance**

	<i>Number of times attended church</i>	
	DDB	GSS
<b>EDUCATIONAL ATTAINMENT</b>		
Less than High School Degree	20.3	22.4
High School degree	22.6	22.8
Some College but no 4-year degree	23.2	22.1
College Degree (4 year) or higher	25.5	23.8
<b>AGE CATEGORY</b>		
18-29	17.7	17.5
30-44	21.9	21.4
45-59	24.4	24.5
60+	27.6	27.9
<b>REGION</b>		
East	22.6	21.0
Midwest	25.8	23.7
South	24.6	25.4
West	18.2	18.3
Sources: General Social Survey, 1972-1996; DDB Needham Life Style consumer surveys, 1975-1997.		