

# Using the Multi-Level Integrated Database Approach

Tom W. Smith

NORC/University of Chicago

Federal Committee on Statistical Methodology Research Conference

Arlington, VA

August, 2007

Non-response is a key component of total survey error (Couper and Groves, 1996; Groves, 2005; Groves and Couper, 1998; Martin, 2004; Smith, 2005; Stoop, 2005) and non-response has been increasing in magnitude over time (de Leeuw and de Heer, 2002; Smith, 1995; 2002a; Zukin, 2006). The two basic approaches to dealing with nonresponse are: 1) developing approaches to increase the response rate and 2) measuring non-response bias and using weighting and/or imputation to compensate for the detected bias (Dillman et al., 2002; Groves, 2005; Groves and Couper, 1998; Lessler and Kalsbeek, 1992; Zanutto and Zaslavsky, 2002). Both of these approaches have yielded only limited success. Despite extensive tests of new interviewing techniques involving a range of treatments (e.g. the use of incentives, tailoring by interviewers, improved interviewer training) and increases in the level of effort (e.g. longer field periods, more call backs, more use of converters), response rates have been falling. Procedures to measure and adjust for non-response bias have been hampered by the fact that often little is known about non-respondents and without information on non-respondents the possibility of both measuring non-response bias and compensating for it is severely limited.

We live in an information age in which considerable data exist in many forms and at many levels (e.g. individuals, households, neighborhoods, communities, etc.). These data sources can be tapped to aid in the collection, adjustment, and analysis of surveys. As Stoop (2004) has observed, by linking sampled cases “to a large number of other registers and administrative records, a large amount of data is available on nonrespondents and respondents with which nonresponse can be analyzed” and as Zanutto and Zaslavsky (2002) have noted, “Administrative records are a relatively inexpensive source of detailed information, especially as technology increases our ability to manipulate large datasets.”

The multi-level, integrated database approach (MIDA) described in this proposal can help to deal with the problems of both low response rates and the detection and adjustment for non-response bias. In addition, MIDA will also further substantive analysis by providing aggregate-level information for contextual analysis. Thus, MIDA will both advance survey-research methodology as well as enhance substantive research.

## Description of MIDA

The essence of MIDA is to use databases to collect as much information as practical about the target sample at both the case-level and at various aggregate levels

during the initial sampling stage. The following description of MIDA will use the example of national US samples of households based on addresses and as such is directly appropriate for postal and in-person samples. However, similar approaches can be applied to other modes and populations (e.g. national, RDD, telephone samples; panel studies; list-based samples; and local surveys).

The first step in MIDA is to extract all relevant, public information at both the case-level and aggregate levels from the sampling frame from which the sample addresses are drawn. In European samples based on population registers, there is often very useful information on such matters as gender, age, and household composition (Bethlehem, 2002; Stoop, 2004; van Goor, Jansma, and Veenstra, 2005; Voogt and Van Kempen, 2002) and list samples (e.g. of employees and HMO enrollees) often have a wealth of sampling frame information (Fowler et al., 2002; Groves, 2006; Kennickell, 2005; Lessler and Kalsbeek, 1992; Moore and Tarnai, 2002; Smith, 1999). But in the US, general population samples of addresses are typically nearly void of household-level information. However, US address samples are rich in aggregate-level information. Address/location of course is the one known attribute of all cases, whether respondents or non-respondents. Moreover, address-based sampling frames are typically based on the US Census and as such the appropriate Census data from blocks, tracts, place, etc. are part of the sampling frame and linked to each address. (That is, the local sample points are selected based on the Census and then addresses within those sample points are obtained from the United States Postal Service Delivery Sequence File and/or special field listings – the later especially typical for rural areas (O’Muircheartaigh, 2003).)

The second step is to augment the sampling frame by linking all cases in the sample to other databases. As Groves (2005) has noted, “Collecting auxiliary variables on respondents and nonrespondents to guide attempts to balance response rates across key subgroups is wise.”

At the case-level that means linking the addresses to such sources as telephone directories, credit records, property records, voter-registration lists, and other public sources (Berge et al., 2005; Brick et al., 2000; Cantor and Cunningham, 2002; Cox, 2006; Davern, 2006; Johnston et al., 2000; Marcus et al., 2006; Williams et al., 2006).<sup>1</sup> Examples of specific useful databases/providers are Accurant, Century List Services, Donnelley/infoUSA, Emerges.Com, Equifax, Experian, Fundrace, Info Quest, Peoplefinders, TARGUSinfo, Telematch, Transunion, the Ultimates, and the US Data Corporation. A number of special procedures have also been developed to use databases in ways not commonly expected and thereby extract much more information than available from more limited and superficial applications (Cantor and Cunningham, 2002; Smith, 2006a; Traub, Pilhuj, and Mallet, 2005; Williams et al., 2006).<sup>2</sup>

The information obtained would include first of all whether a match was or was not found (e.g. listed in telephone directory or not) and, if matched, whatever particular information is available (e.g. names, telephone numbers, credit reports, voter registration status).

---

<sup>1</sup> For a general discussion of record linkage involving surveys see Fair, 1996 and Jenkins et al., 2005. On linking surveys to administrative records see Obenski, 2006 and Davern, 2006.

<sup>2</sup> Survey researchers already have considerable experience in using databases, but further consultation will be carried out with other experts such as data librarians, geographical information systems specialists, cyber-information technicians and data miners, and records searchers such as paralegals and investigators.

At the aggregate level, this means merging information from sources other than those in the sampling frame.<sup>3</sup> Examples of aggregate-level data beyond that from the Census that could be appended are consumer information from such sources as Claritas PRIZM NE and Donnelley Marketing's FIND index, the National Address Server, voting information from national elections, and data on such other matters as vital statistics (Salvo and Lobo, 2003); crime rates (FBI, 2004), religion (Jones, 2002), public housing (HUD, 1998), HIV/STD rates (CDC), and public welfare utilization (Salvo and Lobo, 2003). An example of the extensive information that can be linked from databases to schools and school districts is illustrated by a recent NORC project (Hoffer, Ghadially, and Halverson, 2006).<sup>4</sup>

The linked data would include information from multiple-levels of aggregation. The multi-level analysis will start with household-based data and include neighborhood-level data from Census tract and zip code-based data sources, community-level data from the Census, election counts, crime rates, and other sources, and higher level aggregations (e.g. metropolitan areas and Census divisions).<sup>5</sup>

The third step in MIDA is to take information gained from the initial case-level linkages to secure additional information. For example, securing a name and telephone number from a telephone-directory search can lead to households being found in databases when a mere address was insufficient to allow a match. Also, once a respondent was identified, links to that person in addition to household-level matching could be carried out. Thus, the process of augmenting the sampling frame is iterative and continues during the data-collection phase.

The final step is to record, process, clean, and maintain a large amount of paradata for each case (Couper and Lyberg, 2005; Scheuren, 2000). This would include having interviewers systematically record information about the sample residence (e.g. dwelling type, condition of dwelling), contacts or call attempts, interactions with household members, and observations on the composition and demographics of the household (Bethlehem, 2002; Cantor and Cunningham, 2002; Gfroerer, Lessler, and Parsley, 1997; Groves, 2006; Kennickell, 2005; Lynn et al., 2002; Safir et al., 2002; Smith, 1983; Stoop, 2004).<sup>6</sup> As Cantor and Cunningham (2002) note surveys "should maintain the date and result of each contact or attempt to contact each subject (and each lead)... The reports

---

<sup>3</sup> When starting with addresses without prior Census information as part of the sampling frame, Census and other geographic-based information can be obtained by linking addresses to the geo units (e.g. Census tract, zip code, place/community, etc.) that they fall in. That is, the Census data are added as part of step two if they are not already available as part of the sampling frame. Address linkages to Census tract and higher geo units are possible for from 95-100% of cases (Geronimus, Bound, and Neidert, 1996; Groves and Couper, 1998; Kim, Smith, and Sokolowski, 2006).

<sup>4</sup> The most extensive example of linking all sample cases (i.e. both respondents and non-respondents) to databases is the matching of cases from six major government surveys to the 1990 Census (Gfroerer, Lessler, and Parsley, 1997; Groves and Couper, 1998). Unfortunately the Census can not be generally used for this purpose because of the Bureau of the Census' no access policy to household-level information. Although household-level linkage with the Census is not a viable option, the study demonstrates that 1) a very high level of matching can be achieved between surveys and other records using addresses (96% of non-respondents on the surveys and 97% of respondents were linked to the Census) and 2) the information on the characteristics of non-respondents was very useful in modeling and adjusting for non-response bias.

<sup>5</sup> For multi-level analysis see Bryk and Raudenbush, 1988; DiPrete and Forristal, 1994; and Raudenbush and Bryk, 2002.

<sup>6</sup> This is obviously not possible for postal surveys.

should provide cost and hit data for each method to help manage the data collection effort. In the end it helps to determine those methods that were the most and least cost effective for searching for the population of interest, and this knowledge can be used for planning future surveys.”

### The Utility of MIDA

Consider how the multi-level information in this greatly enriched, sampling frame can be used to advantage for data collection, non-response measurement and adjustment, interview validation, and substantive analysis.

#### Data Collection

First, more information on the target sample will make data collection both more efficient and more effective. For example, securing names and phone numbers can be very helpful in making contact with households and are particularly useful in the case of locked building, gated communities, and other hard to access residences. More information about households before the start of the data-collection phase can greatly ease making contact with households and thus allow efforts to be concentrated on gaining respondent cooperation. It is also very useful if a multiple-mode approach is used (e.g. data collection combining in-person + telephone).

Once contact is made, tailoring is very important in gaining cooperation (Couper and Groves, 1996; Groves and Couper, 1998; Smith, forthcoming). The more information that one has about the household (e.g. whether they have a listed phone number, home owner or renter, etc.), the better able one is to shape interviewers’ approaches and to provide and highlight information most salient to the sampled household (Groves, 2006; Groves, Singer, and Corning, 2000). It is not that well-run surveys do not already make some use of databases to assist interviewers, but what is not done is the careful evaluation of various databases and the retention of the information for other than data-collection efforts.

#### Non-Response Measurement and Adjustment

Second, while this added information will assist interviewers and decrease the overall non-response error, there will still remain a notable amount of non-response on even the better surveys. The information in the MIDA-augmented sampling frame will then be used to measure and adjust for non-response error.<sup>7</sup> Having a wide range of case-level and aggregate-level information is important both to test the representativeness of the achieved sample across as many variables as possible and because surveys covering different topics are likely to have different non-response profiles (e.g. non-voters under-represented in political surveys and the wealthy in the Survey of Consumer Finance – Kennickell, 1997; 2005). Having more relevant information on non-respondents allows

---

<sup>7</sup> It is likely that some information will be most valuable at the data-collection stage and other at the non-response adjustment stage. For example, name and telephone number would be most useful to aid the field work and having a listed/unlisted telephone number, mobility history, and housing tenure would likely be more valuable for non-response adjustments.

for better modeling of non-response bias and the creation of weights that more fully account for the biases and has the particular advantage of having augmented data for all sample cases (Groves, 2005). It also makes fresh, cross-sectional studies more like reinterview, panel studies where the bias from attrition can be well-modeled based on time 1 data (Lepkowski and Couper, 2002).

Research has shown that neighborhood, community, and higher level attributes of areas are correlates of non-response. For example, non-response is consistently and notably higher in large cities than in small towns (Groves and Couper, 1998; Smith, 1983; Smith, 1984; Steeh et al., 2001), in some regions and metropolitan areas vs. others (Groves and Couper, 1998; Johnson and Cho, 2004; Lepkowski and Couper, 2002; Montaquila and Brick, 1997; Murray et al., 2003; Smith, 1983); and related to other aggregate-level attributes such as density, crime rate/fear of crime, social disorganization, geographic mobility, and family structure (Couper and Groves, 1996; Groves, 2006; Groves and Couper, 1998; Goyder, Lock, and McNair, 1992; Gfroerer, Lessler, and Parsley, 1997; Johnson and Cho, 2004; Johnson et al., 2006; Kim, Smith, and Sokolowski, 2006; Kojetin, 1994; O'Hare, Ziniel, and Groves, 2005; van Goor, Jansma, and Veenstra, 2005; Voogt and van Kempen, 2002). Thus, aggregate-level variables are very useful for assessing, understanding, and adjusting for non-response bias (Brick and Broene, 1997; Johnson and Cho, 2004; Kalsbeek, Yang, and Agans, 2002; Kennedy and Bannister, 2005; Kennickell, 2005; Montaquila and Brick, 1997; Nolin et al., 2000; Turrell et al., 2003).

While MIDA is designed to address the matter of nonresponse bias in general, special attention will be focused on examining several prominent theories about the nature and source of nonresponse bias: social disorganization, social isolation, overextension, and structural barriers. MIDA will provide an opportunity to compare and evaluate these theories of non-response.

First, social disorganization theory holds that social structural conditions influence the social relations of people. Wirth (1938) notes that population size, density, and heterogeneity accompanying urbanization weaken individual, family, neighborhood, and social ties. Shaw and McKay (1969) show an association between certain structural conditions and the concentration of social ills such as delinquency. They attribute the higher prevalence of social ills in socially and economically disadvantaged areas to the differences in social organization in the community. Treating refusal rates in Primary Sampling Units (PSUs) "as a behavioral measure of interpersonal trust or helpfulness," House and Wolf (1978:1030) show a positive relationship between crime rate and refusal rate, and find that the total crime rate provides the strongest positive explanatory power on variation of refusal rates among different places. Groves and Couper (1998) show that, controlling for household characteristics, population density and the percentage of individuals under 20 years of age are positively related to survey cooperation. The individual and especially the aggregate level data collected here will provide multiple measures of social disorganization (e.g. crime level, concentration of poverty, residential instability).

Related to social disorganization theory is the concept of collective efficacy which holds areas vary in the willingness of people to "intervene on behalf of the common good" (Johnson et al., 2006; Sampson, Morenoff, and Earls, 1997). Collective efficacy is related to such neighborhood traits as low population turnover, higher

education, higher income, low density, fewer immigrants, and more intact families. Research has found that this propensity is related cooperation in surveys (Couper, Singer, and Kulka, 1998).

Second, social isolation theory argues that nonrespondents are likely to be poorly integrated members of society (Groves and Couper, 1998; Looseveldt and Carton, 2001; Stoop, 2005). According to this theory social isolates are likely to be non-respondents both because of personal misanthropy and because of social and civic disengagement. Personally, social isolates try to minimize inter-personal contacts with others and as such are disinclined to want to cooperate with and engage in an interview (i.e. a conversational interaction) with an interviewer (Converse and Schuman, 1974). Socially and civically, social isolates have little interest in general societal and community affairs and neither follow such matters nor are interested in discussing such topics in an interview. Thus, for these distinct, but associated, reasons social isolates are expected to be overrepresented among non-respondents. It will be possible to examine these expectations by both comparing households that are socially isolated (e.g. with no listed number, no members registered to vote nor belonging to large voluntary associations, etc.) to less isolated households and by comparing more engaged areas (e.g. higher voter turnout, more magazine/newspaper subscriptions) vs. less involved neighborhoods and communities.

Third, overextension theory argues that it is people leading busy lives that tend to be non-respondents (Campanelli, Sturgis, and Purdon, 1997; Groves and Couper, 1998; Lynn, 2002; Smith, 1984). This would include people working full time in general and especially those putting in over time, those with open-ended management responsibilities, and those whose work involves travel. It would naturally include people with multiple, major roles such as full-time employees and parents of small children or those providing in-home eldercare. Databases can often provide useful information on employment status and household composition that can be used to test this hypothesis.

Additionally, many structural factors such as gated communities, locked buildings, policies of gatekeepers, etc. influence contact rates and ultimately response rates and these can be observed and recorded by interviewers and examined by researchers. Including these structural impediments will better specify the overall non-response model.

## Interview Validation

Interviews are checked or validated through a combination of close supervision of field interviewers and the recontacting of respondents to verify that an interview had been conducted with the eligible respondent. Invalid interviews are a relatively small component of total survey error. MIDA will reduce it even further by allowing the information from the databases to be used along with recontacts to help corroborate that interviews were truly and correctly done.

## Substantive Analysis

Finally, for respondents the case-level and aggregate-level data in the augmented sampling frame can be utilized for crucial substantive analysis. While most case-level information would come from the interviews with the respondents, the added case-level

data would include both information uncovered in any particular survey and data that can be used to corroborate information reported by respondents.<sup>8</sup> Procedures for cross-checking information from different databases and between databases and surveys are discussed below.

Aggregate-level information is of great utility for research. Research has demonstrated that contextual, aggregate-level geographic effects in general and neighborhood characteristics in particular influence a wide range of attitudes and behaviors independent of the attributes of individuals. For example, research has shown that impacts exist on 1) political involvement (Bobo and Gilliam 1990; Cohen and Dawson 1993; Gilbert 1991), 2) residential and social mobility (Lee, Oropesa, and Kanan 1994; Massey and Eggers 1990; Massey et al. 1994; South, Baumer, and Lutz 2003), 3) the sexual and reproductive activities of youths and adults (Billy and Moore 1992; Brewster 1994a; Brooks-Gunn et al. 1993; Browning and Olinger-Wilbon 2003; Browning, Leventhal, and Brooks-Gunn, 2004; Cohen et al 2000; Crane 1991; South and Baumer 2001), 4) responses to poverty (Jencks and Mayer 1990; McLeod and Edwards 1995; Oreopoulos 2003), 5) racism and tolerance (Gibson 1995), 6) fear of and involvement in crime (Covington and Taylor 1991; Peeples and Loeber 1994; Sampson, Raudenbush, and Earls, 1997), 7) minorities politically (Cohen and Dawson 1993), economically (Lee et al. 1994; Massey and Eggers 1990), and in other ways (Brewster 1994b; Smith 1994a), 8) social capital and better health (Mellor and Milyo 2004), 9) group membership and economic improvement (Tolbert, Lyson, and Irwin 1998); 10) inequality and political trust (Rahn and Rudolph 2005); 11) religion and deviant behavior (Regnerus 2003), 12) drug use (Boardman et al. 2001; Ford and Beveridge, 2006; Galea, Ahern, and Vlahov, 2003; Snedker, Herting, and Walton, 2006), and 13) depression (Latkin and Curry, 2003).

Among the contextual effects that have been examined from the General Social Survey (GSS)(Davis, Smith, and Marsden, 2007) specifically are the following: 1) racial composition of the local population predicts levels of racial prejudice (Alesina and LaFerrara 2000; Charles 2003; Dixon and Rosenbaum 2004; Taylor 1998 and 2002) and class voting (Weakliem 1997), 2) higher collective levels of trust and civic engagement are associated with lower homicide rates (Rosenfeld et al. 1999 and 2001) and lower mortality in general (Kawachi et al. 1997b), 3) areas with greater aggregate happiness have lower mortality (Jencks 1999), 4) higher levels of anomia are related to higher local crime rates (Rosenfeld and Messner 1998), 5) community-level differences in attitudes on gender roles do not affect the demand for female labor (Cotter et al. 1998), 6) the prevalence of Fundamentalists reduces support for feminism (Moore 1999), 7) a higher level of people on welfare reduces support for welfare spending (Luttmer 1998), 8) living around gun owners increases one's likelihood of acquiring a gun (Glaeser and Glendon 1998), 9) lower income equality is associated with lower social trust and group membership (Kawachi et al. 1997a), 10) community heterogeneity influences civic engagement (Costa and Kahn 2002), 11) community norms shape attitudes toward capital punishment (Baumer, Messner, and Rosenfeld 2003), 12) state and regional differences may be declining over time (Weakliem and Biggert 1999), 13) voting and civic involvement vary by community as well as individual demographics (D'Urso 2003), 14)

---

<sup>8</sup>Examples of collaboration are the voter validations studies – Anderson and Silver, 1986; Burden, 2000; Silver, Anderson, and Abrahamson, 1986.

greater community acceptance of immigrants relates to more occupational achievement by immigrants (De Jong and Steinmetz 2004), 15) community religious beliefs and behaviors influence gender roles (Moore and Vanneman 2003), and 16) aggregate public opinion affects public policies on such as abortion laws, welfare payments, and AIDS-related funding (Brace et al. 2002).

The coding of a rich array of aggregate-level data from the sampling frame and a wide range of databases will facilitate such contextual analysis and make it a regular part of survey analysis rather than an occasional approach carried out only when special multi-level data are added, often after the fact, to standard surveys. In brief, the information in the augmented sampling frame that can be used to assist data collection and adjust for non-response bias can in turn be used for multi-level, contextual analysis.

### MIDA Expansion over Existing Practices

While all of the elements of MIDA have been used in some way or another in some existing surveys, the use of case-level and aggregate-level linkage to databases has not been used in an integrated, systematic manner. The use of databases has been quite limited in terms of both what sources are used and how the linked information is utilized, and the databases and the information from them have not been assessed and evaluated.

One of the limitations of existing approaches is that databases are not used in a systematic manner. For example, telephone directories are often used to try and find the name and number associated with a sampled address or to track a respondent in a panel who has moved. The telephone-directory searches are often quite helpful for these purposes, but their use is purely operational. The information gathered is used by interviewers to help locate respondents, but seldom, if ever, systematically analyzed, used for non-response adjustment, or retained as part of the final analysis file. Conversely, linkage data are sometimes collected for substantive purposes (e.g. to see if graduates of a job-training program end up on welfare), but this information is not used for field operations or non-response adjustment purposes.

A second limitation is that the use of different databases has apparently never been systematically assessed. Different practitioners use different data sources (e.g. telephone directories, credit records, various public, governmental files) based on their familiarity with data sets and/or the data providers and other general preferences. Apparently no rigorous comparisons of the ease-of-use, cost, and yield of various databases have been conducted and none have closely examined the cumulative gain from the use of multiple data sets (Smith, 2006a).

A third limitation has been that few databases have been typically utilized. Telephone directories are the only commonly used database. Other databases such as credit records, property records, and voter registration have been used only occasionally (and only for limited purposes when used at all). Many other potentially valuable databases have apparently never been used (e.g. political contribution lists, membership lists, subscription lists).

A final limitation is that the uses of databases have generally focused on only information obtained about respondents who are found in particular sources. Typically, searches in telephone directories are deemed useful when the target individual or household is located and as not useful when no match occurs (as is the case with the large



proportion of households with unlisted numbers plus those with no telephone). But being found or not found in a database is in itself a useful piece of information and should be recorded for comparing respondents and non-respondents. For example, those listed in the telephone directory are much more likely to be respondents than those not included (Brick et al., 2003; Brick, Montaquila, and Scheuren, 2002; Harvey et al., 2003; Minato and Luo, 2004; O'Hare, Ziniel, and Groves, 2005).

MIDA is designed to overcome each of these standard limitations by comparing and evaluating data sources, flagging both matched and unmatched records, and retaining data for use in all phases of research.

### MIDA Test and Application

To develop and test MIDA, we first propose to draw a sample of 400 addresses from the NORC national sampling frame (Davis, Smith, and Marsden, 2007) and link these to a large number of public databases at the household and/or aggregate levels. These will include, but will not be limited to 1) telephone directories, 2) voter-registration lists, and 3) credit records<sup>9</sup>. In addition, a wide range of innovative data sources will be explored such as group memberships, magazine subscriptions, political and charitable donations, property records, court records, etc. (The distinction is that the certain databases have been used enough to be of known value (e.g. telephone directories and credit records), while others are potentially very useful, but have not been utilized at all or at least not enough to know how valuable they are).

Attention will focus on general, national data sources rather than local, state, and otherwise restricted listings, but some of these will also be explored (e.g. Goyder, Lock, McNair, 1992; Harvey, 2003; Moore and Tarnai, 2002; Murray et al., 2003; Salvo and Lobo, 2003). The potential for acquiring useful linked information is great and often unappreciated. As Cantor and Cunningham (2002) note, "Proprietary databases available on the Internet and elsewhere contain detailed information on large numbers of people. Access to the databases is often restricted. However, these restrictions are often negotiable for limited searches for legitimate research purposes."

Careful account will be kept of the ease, cost, and linkage rates for each type of records obtained, on the consistency of results across different databases, and the cumulative results and total amount of information obtained across the multiple records. This will be used to identify the best databases for general and specific uses.

Second, another important step in the evaluation of MIDA will be checks on the quality of the data from the various databases (Brick et al., 2000; Groves, 2005; Zanutto and Zaslavsky, 2002). As Prewitt (2006) has advocated, rigorous tests of the quality of data from the databases will be conducted. Previous work at NORC gives reason for being optimistic. Work that NORC did in 2004 comparing a sample from a Donnelley database to that from a national survey of the elderly found that in terms of accuracy of information the database was "on par with the field sample" and concluded "our

---

<sup>9</sup> Credit records and other commercial databases include a lot more than the bottom line credit scores and related financial records, they typically include considerable information of mobility, home and vehicle ownership, various legal statuses, ages, household composition, etc. (Brick et al., 2000; Cantor and Cunningham, 2002).

preliminary work with the Donnelley list was encouraging and [we] recommend its use in future...”

Completeness, accuracy, and up-to-datedness will be assessed in two principal ways. First, for all sample cases (respondents and non-respondents), the data from alternative data sources will be compared to one another to get a measure of inter-record consistency. For example, there are three major credit records (Experian, Equifax, and Transunion) and a sample of cases will be linked to all three databases and the information contained within them compared. Other inter-record evaluations will also be conducted. Second, a sample of respondents from the GSS scheduled for reinterviews could be matched to databases and the information from the GSSs could be compared to information from the databases. These comparisons will assess all of the databases examined.

### Assessing and Adjusting for Non-response

The MIDA dataset will contain much more data about non-respondents than are usually available. The full dataset will have household and aggregate-level data for both respondents and non-respondents. Such a rich dataset is uncommon in nationally representative demographic/attitudinal surveys. It provides an opportunity to explore different approaches to estimating and adjusting for non-response bias. For the many variables for which the dataset contains values for both respondents and non-respondents, it will be possible to explore the hypothetical effects of non-response by comparing estimates from these variables for the full dataset with estimates on the respondent cases only. These analyses will suggest which estimates would be most vulnerable to non-response bias. This knowledge will then inform our understanding of the error implicit in estimates from the survey variables themselves, for which non-respondent data are not available. (Gelman and Carlin, 2002; Geronimus, Bound, and Neidert, 1996; Groves, 2005; Groves, 2006; Marker, Judkins, and Winglee, 2002; Meng, 2002; Zanutto and Zaslavsky, 2002).

In addition, the availability of data on non-respondents can improve weighting techniques. In recent rounds, the GSS has incorporated a non-response adjustment at the level of the Primary Sampling Unit (PSU) which are metro areas or non-metro counties. It assumes that the non-respondents in a given area are more like the respondents near them than other respondents. This assumption has been empirically verified and is probably the most common type of non-respondent adjustment used in national, in-person surveys. But the use of PSU to form non-response adjustment cells is limited in the improvement it can provide and is based primarily on a heuristic of availability rather than relying on specific theoretical connection with the study variables. The MIDA-enriched dataset, by providing data on both respondents and non-respondents on many variables, will allow for more discretion in creating non-response adjustment cells and for more sophisticated weighting adjustments (Bethlehem, 2001; Kalton and Kasprzyk, 1986).

Response-propensity weighting is a common method of adjusting for non-response. The theory behind this approach is that all cases, both responding and non-responding, have a non-zero propensity to respond which can be modeled with a logistic regression. The dependent variable is a dummy variable indicating response and the

independent variables are those that predict response: urbanicity, region, household size and composition, interest in the survey topic, etc. Responding cases are then weighted by the inverse of their response propensity to account for the non-responding cases, with low-propensity cases given more weight than high. Like the non-response weighting adjustment discussed above, this method often suffers from a lack of frame variables: the right hand side variables are usually those that are available for all cases rather than those that would be most appropriate. MIDA will permit more thoughtful choices in the independent variables and should improve the response-propensity weighting adjustment. (Ekholm and Laaksonen, 1991).

In addition to giving one a wider selection of variables with which to adjust the weights, MIDA will also provide data with which to compare and evaluate the adjustment methods. These results could greatly improve the weighting methods used for surveys in general.

Similarly, having more variables in the MIDA dataset will improve imputation techniques. Hot-deck imputation fills in values that are missing due to item-non-response by matching cases with missing data to cases without missing data. MIDA will allow better matches and should thus improve the imputation. Also, if the imputation technique chosen involves modeling (e.g., mean regression or multiple imputation), the MIDA dataset will allow better models to be formed with the additional variables. Either way, MIDA will improve the imputation techniques available to surveys in general (Marker et al., 2002).

## Conclusion

MIDA has the potential to advance social-science research in general by notably improving survey-research methodology. Moreover, it does so by drawing on one of the major societal changes in recent decades, the development of large-scale, computerized databases that hold extensive information about individuals, households, neighborhoods, and other societal units. As part of the testing of MIDA all databases utilized will be evaluated for usefulness and reliability.

Methodologically, it should help to increase response rates, allow for a much more comprehensive assessment of non-response bias, and facilitate the calculation of weights and imputations to adjust for the detected non-response bias. Besides providing for a general approach to deal with non-response, it will in particular permit the testing of several prominent theories and hypotheses explaining non-response: social disorganization theory, social isolation theory, overextension theory, structural impediments, etc. The auxiliary data from the databases will permit an examination of general, non-response models (Groves and Couper, 1998).

Substantively, MIDA will improve analysis by easily and automatically making multi-level, contextual variables as ready for analysis as data directly collected in surveys. As the list of examples cited above attest, geographic context has notable impacts on many aspects of people's lives. The contextual data from sampling frames and augmented from multiple databases will provide a rich, contextual array of data for analysis across scores of central substantive topics.

The incorporation of auxiliary data will become more common in surveys as researchers strive to improve contact strategies, increase response rates, and better adjust

for non-response. This is the right moment for a full, rigorous evaluation of the advantages and challenges in using databases, other metadata, and paradata to advance survey research.

## References

- Alesina, Alberto and Eliana LaFerrara, "Participation in Heterogeneous Communities," Quarterly Journal of Economics, 115 (2000), 847-904.
- Anderson, Barbara A. and Silver, Brian D., "Measurement and Mismeasurement of the Validity of Self-Reported Vote," American Journal of Political Science, 80 (1986), 771-785.
- Audit Bureau of Circulations, "Turning Information into Inspiration: Annual Report, 2005," (<http://www.accessabc.com>)
- Baumer, Eric P.; Messner, Steven F.; and Rosenfeld, Richard, "Explaining Spatial Variation in Support for Capital Punishment: A Multilevel Analysis," American Journal of Sociology, 108 (2003), 844-875.
- Berge, Keith H. et al., "Resource Utilization and Outcome in Gravely Ill Intensive Care Unit Patients with Predicted In-hospital Mortality Rates of 95% or Higher by APACHE III Scores," Mayo Clinic Proceedings, 80 (2005), 166-173.
- Bethlehem, Jelke G., "Weighting Nonresponse Adjustments Based on Auxiliary Information," in Survey Nonresponse, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- Billy, John O. G. and Moore, David E., "A Multilevel Analysis of Marital and Nonmarital Fertility in the U.S," Social Forces, 70 (1992), 977-1011.
- Boardman, Jason D.; Finch, Brian Karl; Ellison, Christopher G.; Williams, David R.; and Jackson, James S., "Neighborhood Disadvantage, Stress, and Drug Use Among Adults," Journal of Health and Social Behavior, 42 (2001), 151-165.
- Bobo, Lawrence and Gilliam, Franklin D., Jr., "Race, Sociopolitical Participation, and Black Empowerment," American Political Science Review, 84 (1990), 377-393.
- Bock, R. Darrell, Multilevel Analysis of Educational Data, New York: Academic Press, 1989.
- Brace, Paul; Sims-Butler, Kellie; Arceneaux, Kevin; and Johnson, Martin, "Public Opinion in the American States: New Perspectives Using National Survey Data," American Journal of Political Science, 46 (2002), 173-189.
- Brewster, Karin L., "Neighborhood Context and the Transition to Sexual Activity Among Young Black Women," Demography, 31 (1994a), 603-604.

- Brewster, Karin L., "Race Differences in Sexual Activity Among Adolescent Women: The Role of Neighborhood Characteristics," American Sociological Review, 59 (1994b), 408-424.
- Brick, J. Michael and Broene, Pam, "Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)," Working Paper No. 97-06. Washington, DC: National Center for Education Statistics, 1997.
- Brick, J. Michael et al., "Evaluating Secondary Data Sources for Random Digit Dialing Samples," Proceedings of the American Statistical Association, 2002. Washington, DC: ASA, 2000.
- Brick, J. Michael et al., "Increased Efforts in RDD Surveys," Paper presented to the American Association for Public Opinion Research, Nashville, May, 2003.
- Brick, J. Michael; Montaquila, Jill; and Scheuren, Fritz, "Estimating Residency Rates for Undetermined Telephone Numbers," Public Opinion Quarterly, 66 (2002), 18-39.
- Brooks-Gunn, J.; Duncan, Greg J.; and Klebanov, Sealand N., "Do Neighborhoods Influence Child and Adolescent Development?" American Journal of Sociology, 99 (1993), 353-395.
- Browning, Christopher R. and Olinger-Wilbon, Matisa, "Neighborhood Structure, Social Organization, and Number of Short-Term Sexual Partnerships," Journal of Marriage and the Family, 65 (2003), 730-745.
- Browning, Christopher R.; Leventhal, Tama; and Brooks-Gunn, Jeanne, "Neighborhood Context and Racial Differences in Early Adolescent Sexual Activity," Demography, 41 (2004), 697-702.
- Bryk, Anthony S. and Raudenbush, Stephan W., "Toward a More Appropriate Conceptualization of Research on School Effects: A Three-level Hierarchical Linear Model," American Journal of Education, 97 (1988), 65-108.
- Burden, Barry C., "Voter Turnout and the National Elections Studies," unpublished report, Harvard University, 2000.
- Campanelli, Pamela; Sturgis, Patrick; and Purdon, Susan, "Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates," SCPR, 1997.
- Cantor, David and Cunningham, Patricia, "Methods for Obtaining High Response Rates in Telephone Surveys," in Studies of Welfare Populations: Data Collection and Research Issues, edited by C. Citro et al.. Washington, DC: National Academy Press, 2002.

- Center for Disease Control (CDC), "Sexually Transmitted Disease Surveillance, 2004," Atlanta: Centers for Disease Control and Prevention, 2004.
- Charles, Camille Zubrinsky, "The Dynamics of Racial Residential Segregation," Annual Review of Sociology, 29 (2003), 167-207.
- Cohen, Cathy J. and Dawson, Michael C., "Neighborhood Poverty and African American Politics," American Political Science Review, 87 (1993), 286-302.
- Cohen, D.; Spear, S.; Scribner, R.; Kissinger, P.; Mason, K.; and Wildgen, J., "Broken Windows and the risk of gonorrhea," American Journal of Public Health, 90 (2000), 230-236.
- Converse, Jean M. and Schuman, Howard, Conversations at Random: Survey Research as Interviewers See It. New York: John Wiley & Sons, 1974.
- Costa, Dora L. and Kahn, Matthew E., "Civic Engagement and Community Heterogeneity: An Economist's Perspective," Paper presented to the Conference on Social Connectedness and Public Activism, Cambridge, 2002.
- Cotter, David A. et al., "The Demand for Female Labor," American Journal of Sociology, 103 (1998), 1673-1712.
- Couper, Mick P. and Groves, Robert M., "Social Environmental Impacts on Survey Cooperation," Quality & Quantity, 30 (1996), 173-188.
- Couper, Mick P. and Lyberg, Lars E., "The Use of Paradata in Survey Research," Proceedings of the International Statistical Institute, Sydney, April, 2005.
- Couper, Mick P.; Singer, Eleanor; and Kulka, Richard A., "Participation in the 1990 Decennial Census: Politics, Privacy, Pressures," American Politics Quarterly, 26 (1998), 59-80.
- Covington, J. and Taylor, R. B., "Fear of Crime in Urban Residential Neighborhoods: Implications of Between- and Within-Neighborhood Sources for Current Models," Sociological Quarterly, 32 (1991), 231-249.
- Cox, Christine S., "Integrating Data: Policy and Legal Issues," Report prepared for the Federal Committee on Statistical Methodology, November, 2006.
- Crane, Jonathan, "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing," American Journal of Sociology, 96 (1991), 1226-1259.
- D'Urso, Victoria T., "Internet Use and the Duration of Buying and Selling in the Residential Housing Market, Economic Incentives, and Voting," Unpublished

- Ph.D. Thesis, MIT, 2003.
- Davern, Michael. "Evaluating Linked Survey and Administrative Data for Policy Research," Paper presented to the Federal Committee on Statistical Methodology, Washington, DC, November, 2006.
- Davis, James A.; Smith, Tom W.; and Marsden, Peter V., General Social Survey, 1972-2006: Cumulative Codebook. Chicago: NORC, 2007.
- De Jong, Gordon F. and Steinmetz, Michele, "Receptivity Attitudes and the Occupational Attainment of Male and Female Immigrant Workers," Population Research and Policy Review, 23 (2004), 91-116.
- De Leeuw, Edith and De Heer, Wim, "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison," in Survey Nonresponse, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- Dillman, Don A. et al., "Survey Nonresponse in Design, Data Collection, and Analysis," in Survey Nonresponse, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- DiPrete, Thomas A. and Forristal, Jerry D., "Multivariate Models: Methods and Substance," Annual Review of Sociology, 20 (1994), 331-357.
- Dixon, Jeffrey C. and Rosenbaum, Michael S., "Nice to Know You? Testing Contact, Cultural, and Group Threat Theories of Anti-Black and Anti-Hispanic Stereotypes," Social Science Quarterly, 85 (2004), 257-280.
- Ekholm, Anders and Seppo Laaksonen. 1991. "Weighting via Response Modeling in the Finnish Household Budget Survey." *Journal of Official Statistics* 7(3): 325-337.
- Fair, Martha E., "Record Linkage in an Information Age Society," Proceedings of the Census Bureau's Conference and Technology Interchange. Washington, DC: Bureau of the Census, 1996.
- Federal Bureau of Investigation (FBI), "Uniform Crime Reporting Program Data: County-Level Detailed Arrest and Offense Data, 2004," Washington, DC: Federal Bureau of Investigation, 2004.
- Ford, Julie and Beveridge, Andrew A., "Neighborhood Crime Victimization, Drug Use, and Drug Sales," Paper presented to the American Sociological Association, Montreal, 2006.
- Fowler, Floyd J., Jr. et al., "Using Telephone Interviews to Reduce Nonresponse Bias to Mail Surveys of Health Plan Members," Medical Care, 40 (2002), 190-200.



- Galea, Sandro; Ahern, Jennifer; and Vlahov, David., "Contextual Determinants of Drug Use Risk Behavior: A Theoretic Framework," The Review of Economics and Statistics, 83 (2003), 257-268.
- Gelman, Andrew and Carlin, John B., "Poststratification and Weighting Adjustments," in Survey Nonresponse, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- Geronimus, Arline T.; Bound, John; and Neidert, Lisa J., "On the Validity of Using Census Geocode Characteristics to Proxy Individual Socioeconomic Characteristics," Journal of the American Statistical Association, 91 (1996), 529-537.
- Gfroerer, Joseph; Lessler, Judith; and Parsley, Teresa, "Studies of Nonresponse and Measurement Error in the National Household Survey of Drug Abuse," In L. Harrison & A. Hughes (Eds.), The validity of self-reported drug use: Improving the accuracy of survey estimates (NIH Publication No. 97-4147, NIDA Research Monograph 167, pp. 273-295). Rockville, MD: National Institute on Drug Abuse, 1997.
- Gibson, James L., "The Political Freedom of African Americans: A Contextual Analysis of Racial Attitudes, Political Tolerance, and Individual Liberty," Political Geography, 14 (1995), 571-599.
- Gilbert, Christopher P, "Religion, Neighborhood, Environments, and Partisan Behavior: A Contextual Analysis," Political Geography Quarterly, 10 (1991), 110-131.
- Glaeser, Edward L. and Glendon, Spender., "Who Owns Guns? Criminals, Victims, and the Culture of Violence," Paper presented to the American Economics Association, Chicago, 1998.
- Goyder, John; Lock, Jean; and McNair, Trish, "Urbanization Effects on Survey Nonresponse: A Test Within and Across Cities," Quality & Quantity, 26 (1992), 39-48.
- Groves, Robert M., "Designing Surveys Acknowledging Nonresponse," Unpublished report, National Academies, 2006.
- Groves, Robert M., "Nonresponse Rates and Nonresponse Bias in Household Surveys," Public Opinion Quarterly, 70 (2006), 646-675.
- Groves, Robert M., "Research Synthesis: Nonresponse Rates and Nonresponse Error in Household Surveys, Unpublished report, Institute for Social Research, 2005.

- Groves, Robert M., "Total Survey Error," Plenary Presentation to the American Association for Public Opinion Research, Miami Beach, May, 2005.
- Groves, Robert M. and Couper, Mick P., Nonresponse in Household Interview Surveys. New York: John Wiley & Sons, 1998.
- Groves, Robert M; Singer, Eleanor; and Corning, Amy, "Leverage-Saliency Theory of Survey Participation: Description and an Illustration," Public Opinion Quarterly, 64 (2000), 299-308.
- Harding, Jennifer et al., "The Do Not Call Registry: Friend or Foe? The Effect of Do Not Call Lists on Survey Response," Paper presented to the American Association for Public Opinion Research, Miami Beach, May, 2005.
- Harvey, Bart J. et al., "Using Publicly Available Directories to Trace Survey Nonresponders and Calculate Adjusted Response Rates," American Journal of Epidemiology, 158 (2003), 1007-1111.
- Hoffer, Thomas B.; Ghadialy, Rashna; and Halverson, Maries, "Collection of External Data Elements: Final Report and File Documentation," NORC report for the National Education Association, 2006.
- House, James S. and Sharon Wolf. 1978. "Effects of Urban Residence on Interpersonal Trust and Helping Behavior." *Journal of Personality and Social Psychology* 36:1029-1043.
- HUD (Department of Housing and Urban Development), "A Picture of Subsidized Households – 1998," Washington, DC: HUD, 1998.
- Hynes, Margaret M. and Mueller, Lloyd M., "A Report on Fatal and Non-Fatal Drug Overdoses in Connecticut, 2004," Connecticut Department of Public Health, 2004.
- Jencks, Christopher and Mayer, Susan E., "The Social Consequences of Growing Up in a Poor Neighborhood," in Inner-City Poverty in the United States, edited by L. E. Lynn, Washington, D.C.: National Academy Press, 1990.
- Jencks, Christopher, "What is the Optimal Level of Inequality?" Unpublished paper, Harvard University, 1999.
- Jenkins, Stephen P. et al., "Linking Household Survey and Administrative Record Data: What Should the Matching Variables Be?" Discussion Paper 489. German I Institute for Economic Research, 2005.
- Jenkins, Stephen P. et al., "Patterns of Consent: Evidence from a General Household Survey," Journal of the Royal Statistical Society A, 169(2006), 1-22.

- Johnson, Timothy and Cho, Young Ik, "Understanding Nonresponse Mechanisms in Telephone Surveys," Paper presented to the American Association for Public Opinion Research, Phoenix, May, 2004.
- Johnson, Timothy P.; Cho, Young Ik; Campbell, Richard T.; and Holbrook, Allyson L., "Using Community-Level Correlates to Evaluate Nonresponse Effects in a Telephone Survey," Public Opinion Quarterly, 70 (2006), 704-719.
- Johnston, S. Claiborne et al., "Endovascular and Surgical Treatment of Unruptured Cerebral Aneurysms: Comparison of Risks," Annals of Neurology, 48 (2000), 11-19.
- Jones, Dale E., Religious Congregations and Membership in the United States, 2000. Nashville: Glenmary Research Center, 2002.
- Kalsbeek, William D.; Yang, Juan; and Agans, Robert P., "Predictors of Nonresponse in a Longitudinal Survey of Adolescents," Proceedings of the American Statistical Association, 2002. Washington, DC: American Statistical Association, 2002.
- Kalton, Graham and Daniel Kasprzyk. 1986. "The Treatment of Missing Survey Data." *Survey Methodology* 12(1): 1-16.
- Kawachi, Ichiro; Kennedy, Bruce P.; and Lochner, Kimberly, "Long Live Community: Social Capital and Public Health," The American Prospect, 35 (1997a), 56-59.
- Kawachi, Ichiro; Kennedy, Bruce P.; Lochner, Kimberly; and Prothrow-Stith, D., "Social Capital, Income Inequality, and Morality," American Journal of Public Health, 87 (1997b), 1491-1498.
- Kennickell, Arthur B., "Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances," Paper presented to the American Statistical Association, Anaheim, August, 1997.
- Kennickell, Arthur B., "Darkness Made Visible: Field Management and Nonresponse in the 2004 SCF," Paper presented to the American Statistical Association, Minneapolis, August, 2005.
- Kim, Jibum, and Lauderdale, Diane, "The Role of Community Context in Immigrant Elderly Living Arrangements: Korean American Elderly," Research on Aging, 24 (2002), 630-653.
- Kim, Jibum; Smith, Tom W.; Kang, Jeong-han; and Sokolowski, John, "Community Context and Cooperation Rate" GSS Methodological Report No. 108, Chicago, NORC, 2006.

- Kim, Jibum; Smith, Tom W.; and Sokolowski, John, "Ecological Correlates of Cooperation Rate in the 2002 and 2004 General Social Survey," Paper presented to the American Association for Public Opinion Research, Montreal, May, 2006
- Kim, Jibum and Sokolowski, John, "Survey non-response in a national area probability sample as a dimension of survey quality; an analysis of community characteristics" Paper presented to the American Association for Public Opinion Research, Miami, May, 2005.
- Kojetin, Brian A., "Characteristics of Nonrespondents to the Current Population Survey (CPS) and Consumer Expenditure Interview Survey (CSES)," Proceedings of the American Statistical Association, 1994. Washington, DC: American Statistical Association, 1994.
- Latkin, Carl A. and Curry, Aaron D., "Stressful Neighborhoods and Depression: A Prospective Study of the Impact of Neighborhood Disorder," Journal of Health and Social Behavior, 44 (2003), 34-44.
- Lee, Barret; Oropesa, R.S.; and Kanan, James W., "Neighborhood Context and Residential Mobility," Demography, 31 (1994), 249-270.
- Lepkowski, James M. and Couper, Mick P., "Nonresponse in the Second Wave of Longitudinal Household Surveys," in Survey Nonresponse, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- Lessler, Judith T. and Kalsbeek, William D., Nonsampling Error in Surveys, New York: John Wiley & Sons, 1992.
- Link, Michael W. et al., "Has the National Do Not Call Registry Helped or Hurt State-Level Response Rates? A Time Series Analysis," Public Opinion Quarterly, 70 (2006), 794-809.
- Loosveldt, Geert and Carton, Ann, "Utilitarian Individualism and Panel Nonresponse," International Journal of Public Opinion Research, 12 (2002), 428-438.
- Luttmer, Erzo, "Understanding Income Redistribution: The Role of Interpersonal Preference, Information, and Mechanism Design," Unpublished Ph.D. Thesis, Harvard, 1998.
- Lynn, Peter, "PEDAKSI: Methodology for Collecting Data about Survey Non-respondents," Working Paper for the Institute for Social and Economic Research, 2002-05, University of Essex, 2002.
- Lynn, Peter et al., "The Effect of Extended Interviewer Efforts on Nonresponse Bias," in Survey Nonresponse, edited by Robert M. Groves et al., New York: John Wiley & Sons, 2002.

- Marcus, Pamela M et al., "Extended Lung Cancer Incidence Follow-up in the Mayo Lung Project and Overdiagnosis," *Journal of the National Cancer Institute*, 98 (2006), 748-756.
- Marker, David A.; Judkins, David R.; Wingless, Marianne, "Large-scale Imputation for Complex Surveys," in *Survey Nonresponse*, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- Martin, Elizabeth, "Unfinished Business," *Public Opinion Quarterly*, 68 (2004), 439-450.
- Massey, Douglas S. and Eggers, Mitchell L., "The Ecology of Inequality: Minorities and the Concentration of Poverty," *American Journal of Sociology*, 95 (1990), 1153-1188.
- Massey, Douglas S.; Gross, Andrew B.; and Shibuya, Kumiko, "Migration, Segregation, and the Geographic Concentration of Poverty," *American Sociological Review*, 59 (1994), 425-445.
- McLeod, Jane and Edwards, Kevan, "Contextual Determination of Children's Responses to Poverty," *Social Forces*, 73 (1995), 487-516.
- Mellor, Jennifer M. and Milyo, Jeffrey, "State Social Capital and Individual Health Status," Unpublished Paper, College of William and Mary, 2004.
- Meng, Xiao-Li, "A Congenial Overview and Investigation of Multiple Imputation Inferences under Uncongeniality," in *Survey Nonresponse*, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- Minato, Hiroaki and Luo, Lidan, "Towards a Better Estimate of Working Residential Number (WRN) Rate Among the Undetermined: An Application of Survival Analysis," Paper presented to the American Association for Public Opinion Research, Phoenix, May, 2004.
- Montaquila, Jill M. and Brick, J. Michael, "Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey," Working Paper 97-40, National Center for Education Statistics, 1997.
- Moore, Danna L. and Tarnai, John, "Evaluating Nonresponse Error in Mail Surveys," in *Survey Nonresponse*, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- Moore, Laura M., "A Multi-Level Analysis of Attitudes Regarding Women in Politics: Why is the United States South Different?" Unpublished Ph.D. Thesis, University of Maryland, 1999.

- Moore, Laura M. and Vanneman, Reeve, "Context Matters: Effects of the Proportion of Fundamentalists on Gender Attitudes," Social Forces 82 (2003), 115-139.
- Murray, Mary Cay, "Impact of Changes in the Telephone Environment on RDD Telephone Surveys," Paper presented to the American Association for Public Opinion Research, Nashville, May, 2003.
- Nolin, Mary Jo et al., National Household Education Survey of 1999: Methodology Report. Washington, DC: National Center for Education Statistics, 2000.
- Obenski, Sally, "Models and Applications of Administrative Records Research," Workshop on Data Linkages to Improve Health Outcomes, Washington, DC, September, 2006.
- O'Hare, Barbara C.; Ziniel, Sonja; and Groves, Robert M., "Modeling Components of Response Propensity in Centralized Telephone Surveys," Paper presented to the American Association for Public Opinion Research, Miami Beach, May, 2005.
- O'Muircheartaigh, Colm, "There and Back Again: Demographic Survey Sampling in the 21<sup>st</sup> Century," 2003 at [www.fcsn.gov/events,papers2003.html](http://www.fcsn.gov/events,papers2003.html)
- Oreopoulos, Philip, "The Long-Run Consequences of Living in a Poor Neighborhood," Quarterly Journal of Economics, 118 (2003), 1533-1575.
- Peebles, F. and Loeber, R., "Do Individual Factors and Neighborhood Context Explain Ethnic Differences in Juvenile Delinquency," Journal of Quantitative Criminology, 10 (1994), 141.
- Rahn, Wendy M. and Rudolph, Thomas J., "A Tale of Political Trust in American Cities," Public Opinion Quarterly, 69 (2005), 530-560.
- Raudenbush, Stephen Wand Bryk, Anthony S., Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, CA: Sage, 2002.
- Regnerus, Mark D., "Moral Communities and Adolescent Delinquency: Religious Contexts and Community Social Control," The Sociological Quarterly, 44 (2003), 523-554.
- Rosenfeld, Richard and Messner, Steven F., "Beyond the Criminal Justice System: Anomie, Institutional Vitality, and Crime in the United States," Paper presented to the American Sociological Association, San Francisco, 1998.
- Rosenfeld, Richard; Bray, Timothy M.; and Egley, Arlen, "Facilitating Violence: A Comparison of Gang-Motivated, Gang-Affiliated, and Nongang Youth Homicides," Journal of Quantitative Criminology, 15 (1999), 495-516.

- Rosenfeld, Richard; Messner, Steven F.; and Baumer, Eric P., "Social Capital and Homicide," Social Forces, 80 (2001), 283-309.
- Safir, Adam et al., "Effects on Survey Estimates from Reducing Nonresponse," Proceedings of the American Statistical Association, 2001. Washington, DC: ASA, 2001.
- Salvo, Joseph and Lobo, Arun Peter, "The Effect of Neighborhood Characteristics on Nonresponse in the Bronx Test Site of the American Community Survey," Proceedings of the American Statistical Association, 2003. Washington, DC: ASA, 2003.
- Sampson, Robert J.; Morenoff, Jeffrey D.; and Earls, Felton, "Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children," American Sociological Review, 64 (1999), 633-660.
- Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls, "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy," Science, 277 (1997), 918ff.
- Scheuren, Fritz, "Macro and Micro Paradata for Survey Assessment," unpublished report, Urban Institute, 2000.
- Shaw, Clifford R. and Henry D. McKay. 1969. *Juvenile Delinquency and Urban Areas*. Chicago: The University of Chicago Press.
- Silver, Brian D.; Anderson, Barbara A.; and Abramson, Paul R., "Who Overreports Voting," American Political Science Review, 80 (1986), 613-624.
- Smith, Tom W., Anti-Semitism in Contemporary America, New York: American Jewish Committee, 1994a.
- Smith, Tom W., "Developing Nonresponse Standards," in Survey Nonresponse, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002a.
- Smith, Tom W., "Estimating Nonresponse Bias with Temporary Refusals," Sociological Perspectives, 27 (1984), 473-489.
- Smith, Tom W., "Estimating the Status of Cases with Unknown Eligibility in Telephone Surveys," Paper presented to the Second International Conference on Telephone Survey Methodology, Miami, January, 2006a.
- Smith, Tom W., "A Generation of Data: The General Social Survey, 1972-2002," GSS Project Report No. 24. Chicago: NORC, 2002b.
- Smith, Tom W., "The Hidden 25%: An Analysis of Nonresponse on the General Social

- Survey,” Public Opinion Quarterly, 47 (1983), 386-404.
- Smith, Tom W., “A Methodological Analysis of HIV Risk Behavior from the 1988-1998 General Social Survey,” GSS Methodological Report No. 92. Chicago: NORC, 1999.
- Smith, Tom W., “A Methodological Analysis of HIV Risk Behavior from the 1988-2002 General Social Survey,” GSS Methodological Report No. 97. Chicago: NORC, 2003.
- Smith, Tom W., “Notes on John Brehm, The Phantom Respondent,” GSS Methodological Report No. 79. Chicago: NORC, 1993.
- Smith, Tom W., “Reporting Survey Nonresponse in Academic Journals,” International Journal of Public Opinion Research, 14 (Winter, 2002c), 469-474.O
- Smith, Tom W., “The Subsampling of Nonrespondents on the 2004 General Social Survey,” GSS Methodological Report No. 106. Chicago: NORC, 2006b.
- Smith, Tom W., “Survey Non-response in Cross-national Perspective: The 2005 ISSP Non-response Survey,” Survey Research Methods, forthcoming.
- Smith, Tom W., “Total Survey Error,” in Encyclopedia of Social Measurement, edited by Kimberf Kempf-Leonard. New York: Academic Press, 2005.
- Smith, Tom W., “Trends in Nonresponse Rates,” International Journal of Public Opinion Research, 7 (1995), 157-171.
- Snedker, Karen A.; Herting, Jerald R.; and Walton, Emily C., “Neighborhood Contextual Effects and Adolescent Substance Use: Exploring the Moderating Role of Neighborhoods,” Paper presented to the American Sociological Association, Montreal, 2006.
- South, Scott and Baumer, Eric P., “Community Effects on the Resolution of Adolescent Premarital Pregnancy,” Journal of Family Issues, 22 (2001), 1025-1042.
- South, Scott; Baumer, Eric P.; and Lutz, Amy, “Interpreting Community Effects on Youth Educational Attainment,” Youth and Society, 35 (2003), 3-36.
- Speizer, Howard and Baker, Reginald, “The Impact of the Federal Do Not Call Registry on Telephone Survey Research,” Paper presented to the American Association for Public Opinion Research, Miami Beach, May, 2005.
- Stapulonis, Rita A.; Kovac, Martha D.; and Fraker, Thomas M., “Surveying Current and Former TANF Recipients in Iowa,” Paper presented to the Association for Public Policy Analysis and Management, Washington, DC, November, 1999.



- Steeh, Charlotte et al., "Are They Really as Bad as They Seem? Nonresponse Rates at the End of the Twentieth Century," Journal of Official Statistics, 17 (2001), 227-247.
- Stoop, Ineke A.L., The Hunt for the Last Respondent: Nonresponse in Sample Surveys. The Hague: Social and Cultural Planning Office, 2005.
- Stoop, Ineke A.L., "Survey Nonrespondents," Field Methods, 16 (2004), 23-54.
- Taylor, Marylee C., "How White Attitudes Vary with the Racial Composition of Local Populations: Numbers Count," American Sociological Review, 63 (1998), 512-535.
- Taylor, Marylee C., "Fraternal Deprivation, Collective Threat, and Racial Resentment: Perspectives on White Racism," in Relative Deprivation: Specification, Development, and Integration, edited by Walker and Smith, New York: Cambridge University Press, 2002.
- Tolbert, Charles M.; Lyson, Thomas A.; and Irwin, Michael D., "Local Capitalism, Civic Engagement, and Socioeconomic Well-Being," Social Forces, 77 (1998), 401-427.
- Traub, Jane; Pilhuj, Kathy; and Mallett, Daniel T., "'You Don't Have to Accept Low Survey Response Rates' – How We Achieved the Highest Survey Cooperation Rates in Company History," Paper presented to the American Association for Public Opinion Research, Miami Beach, May, 2005.
- Turrell, Gavin et al. "The Socio-economic Patterning of Survey Participation and Non-response Error in a Multilevel Study of Food Purchasing Behaviour: Area- and Individual-level Characteristics," Public Health Nutrition, 6 (2003), 181-189.
- Van Goor, Henk; Jansma, Folbert; and Veenstra, Rene, "Differences in Undercoverage and Nonresponse between City Neighborhoods in a Telephone Survey," Psychological Reports, 96 (2005), 867-878.
- Varian, Hal; Wallenberg, Fredrik; and Woroch, Glenn, "The Demographics of the Do-Not-Call List," IEEE Security & Privacy, (Jan./Feb., 2005), 34-39.
- Voogt, Robert J.J. and van Kempen, Hetty, "Nonresponse Bias and Stimulus Effects in the Dutch National Election Study," Quality & Quantity, 36 (2002), 325-345.
- Weakliem, David L., "Race Versus Class? Racial Composition and Class Voting, 1936-1992," Social Forces, 75 (1997), 939-956.
- Weakliem, David L. and Biggert, Robert, "Region and Political Opinion in the Contemporary United States," Social Forces, 77 (1999), 863-886.

- Williams, Jason et al., "Enhancing the Validity of Foster Care Follow-up Studies through Multiple Alumni Locations Strategies," Child Welfare, 85 (2006), 499-521.
- Wirth, Louis. 1938. "Urbanism as a Way of Life." *American Journal of Sociology* 44:3-24.
- Young, Gay and Gately, Tamra, "Neighborhood Impoverishment and Child Maltreatment: An Analysis from the Ecological Perspective," Journal of Family Issues, 9 (1988), 240-254.
- Zanutto, Elaine and Zaslavsky, Alan, "Using Administrative Records to Impute Nonresponse," in Survey Nonresponse, edited by Robert M. Groves et al. New York: John Wiley & Sons, 2002.
- Zukin, Cliff, "The Future is Here! Where Are We Now? And How Do We Get There?" Public Opinion Quarterly, 70 (2006), 426-442.