The ANNALS of the American Academy of Political and Social Science
http://ann.sagepub.com/

An Assessment of the Multi-level Integrated Database Approach

Tom W. Smith and Jibum Kim

The ANNALS of the American Academy of Political and Social Science 2013 645: 185
DOI: 10.1177/0002716212463340

An Assessment of the Multi-level Integrated Database Approach

By
TOM W. SMITH
and
JIBUM KIM

The Multi-level Integrated Database Approach (MIDA) is an innovative procedure to improve survey research in general and to assess and adjust for nonresponse bias in particular. Its utility for data collection, nonresponse measurement, estimate adjustment, interviewer validation, and analysis are considered. Also discussed is how MIDA extends survey methodology beyond existing practices. A demonstration of a MIDA-enhanced sample frame for U.S. households is carried out. Finally, further steps for the use and testing of MIDA are covered.

Keywords: aggregate-level data; paradata; nonresponse; auxiliary data; database linkage

The Multi-level Integrated Database Approach (MIDA) is an innovative procedure to improve survey research in general and to assess and adjust for nonresponse bias in particular. In this article, after describing MIDA, we outline its utility for data collection, nonresponse measurement, estimate adjustment, interviewer validation, and analysis. We then discuss how MIDA extends survey methodology beyond existing practices and demonstrate the construction of a MIDA-enhanced sample frame for U.S. households. We conclude by indicating further steps for the use and testing of MIDA.

MIDA

The essence of the MIDA approach is the use of multiple databases to collect as much

Tom W. Smith specializes in the study of societal change and survey methodology. He is principal investigator of the National Data Program for the Social Sciences and director of its General Social Survey at the National Opinion Research Center (NORC) at the University of Chicago. He is cofounder and was secretary general of the International Social Survey Program.

Jibum Kim is a senior research scientist at NORC at the University of Chicago and an associate professor in the Department of Sociology at Sungkyunkwan University, Seoul, Korea.

DOI: 10.1177/0002716212463340

information as is practical about the target sample at both the household level and at various aggregate levels during the initial sampling stage. The first step in MIDA is to extract all relevant public information at both the household and aggregate levels pertaining to the sampling frame from which respondent addresses are drawn. In Europe, samples based on population registers often contain very useful information on such matters as gender, age, and household composition (Bethlehem 2001; Stoop 2004; van Goor, Jansma, and Veenstra 2005; Voogt and van Kempen 2002). Likewise, list-based sampling frames (e.g., of employees and HMO enrollees) often have a wealth of information (Fowler et al. 2002; Groves 2005, 2006; Kennickell 2005; Lessler and Kalsbeek 1992; D. Moore and Tarnai 2002; Smith 1999).

In the U.S. population, however, samples of addresses typically are nearly devoid of household-level information, though U.S. address files are often rich in aggregate-level data. The address or location of the sampled unit is, of course, the one known attribute of all cases, whether respondent or nonrespondent. Moreover, address frames are typically based on the U.S. census, and as such, data on blocks, tracts, places, and so on may be linked to each address. To be more specific, local sample points are generally selected based on Census Enumeration Lists, and then addresses within those sample points are obtained from the United States Postal Service Delivery Sequence File or special field listings, the latter of which are used especially 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 (2006, 669) has noted, "Collecting auxiliary variables on respondents and nonrespondents to guide attempts to balance response rates across key subgroups is wise." At the household level, this operation generally means linking addresses to such sources as telephone directories, credit records, property records, voter-registration lists, and other publicly available sources (Berge et al. 2005; Brick et al. 2000; Cantor and Cunningham 2002; Cox 2006; Davern 2006; Johnston et al. 2000; Marcus at al. 2006; Williams et al. 2006).[1] Special procedures have been developed to use databases in uncommon ways and thereby extract much more information than is available from more limited and superficial applications (Cantor and Cunningham 2002; Smith 2006; Traub, Pilhuj, and Mallet 2005; Williams et al. 2006).[2]

The information obtained from linking to other databases would include, first, whether a match was or was not found (e.g., listed in the telephone directory or not; registered to vote or not) and, if matched, whatever particular information was available from that source (e.g., names, telephone numbers, voter registration status). At the aggregate level, linkage involves merging information from sources other than those in the sampling frame.[3] Examples of aggregate-level information beyond what is available from the census include consumer data compiled from proprietary databases, such as the PRIZM NE file of Claritas Corporation, magazine subscriptions (Audit Bureau of Circulations 2005), religious congregations (Jones 2002), voter and vital statistics registries (Salvo and Lobo 2003), crime reports (FBI 2004), public housing records (U.S. Department of Housing and Urban Development 1998), HIV/STD prevalence rates (Centers for Disease Control and Prevention 2004), and welfare lists (Salvo and Lobo 2003). The linked dataset would thus include information from multiple levels of aggregation, beginning with household-based data but also neighborhood-level data from census tract– and zipcode-based sources; community-level data from the census, election counts, crime rates, and other sources; and higher-level data for aggregations such as metropolitan areas and census divisions.[4]

The third step in MIDA is to take data gained from initial household-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 where an address was insufficient to allow a match. Once a respondent has been identified, moreover, links to that person in addition to household-level matching can be carried out. The process of augmenting the sampling frame is thus iterative and can continue even during the data collection phase.

The final step in MIDA is to record, process, clean, update, and maintain a large amount of paradata for each case (Couper and Lyberg 2005; Scheuren 2000). This operation includes having interviewers systematically record information about the sample residence (e.g., dwelling type, condition of dwelling), compiling information on contacts or call attempts, recording information about interactions with household members, and making observations on the composition and demographics of the household (Bethlehem 2001; 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).[5] As Cantor and Cunningham (2002, 59) note, surveys "should maintain the date and result of each contact or attempt to contact each subject (and each lead). . . . The reports 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 in searching for the population of interest, and this knowledge can be used for planning future surveys." Such information exists for both the nonrespondents and respondents and thus can be readily utilized to examine nonresponse bias.

The Utility of MIDA

We now turn to a consideration of how multilevel information gathered and compiled into a greatly enriched sampling frame is advantageous for a variety of purposes, including data collection, nonresponse measurement, adjustment for unit and item nonresponse, interview validation, and substantive analysis.

Data collection

Compiling more information on members of the target sample will generally 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 these items are particularly useful in the case of locked buildings, gated communities, and other hard-to-access residences. More information about households obtained 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., undertaking data collection through a combination of in-person and telephone interviews).

Once contact is made, tailoring is very important in gaining cooperation (Couper and Groves 1996; Groves and Couper 1998; Smith 2007). The more information one has about the household (e.g., whether there is a listed phone number, home owner or renter, etc.), the better able one is to shape interviewers' approaches to potential respondents and to provide and highlight salient information about the sampled household (Groves 2006; Groves, Singer, and Corning 2000). Well-run surveys today do already make some use of databases to assist interviewers, but what typically is not done is the careful evaluation of various databases and the retention of the information for activities other than data collection efforts.[6]

Nonresponse measurement and adjustment

Although additional information will assist interviewers and decrease the overall nonresponse error, there will still remain a notable degree of nonresponse on even the better surveys. The information in the MIDA-augmented sampling frame can then be used to measure and adjust for nonresponse error.[7] Having a wide range of household-level and aggregate-level data 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 nonresponse profiles (e.g., nonvoters will be underrepresented in political surveys and the wealthy in consumer surveys; see Kennickell 1997, 2005). Access to more relevant information on nonrespondents allows for better modeling of nonresponse 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 and community attributes correlate strongly with rates of nonresponse. Nonresponse is consistently and notably higher in large cities than in small towns, for example (Groves and Couper 1998; Smith 1983, 1984; Steeh et al. 2001), and higher in some regions and metropolitan areas than in others (Groves and Couper 1998, 2001; Johnson and Cho 2004; Lepkowski and Couper 2002; Montaquila and Brick 1997; Murray et al. 2003; Smith 1983). Nonresponse is also related to other aggregate-level factors, 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, 2001; 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; Smith 2002; van Goor, Jansma, and Veenstra 2005; Voogt and van Kempen 2002). As a result, aggregate-level variables are very useful for assessing, understanding, and adjusting for nonresponse bias (Brick and Broene 1997; Johnson and Cho 2004; Kalsbeek, Yang, and Agans 2002; Kennickell 2005; Montaquila and Brick 1997; Nolin et al. 2000; Turrell et al. 2003).

While MIDA is designed to address the issue of nonresponse bias in general, special attention can be focused on several prominent theories about the nature and source of nonresponse bias: social disorganization, social isolation, overextension, and structural barriers. The theory of social disorganization holds that contextual circumstances influence social relations among people. Wirth (1938) noted early on that greater population size, density, and heterogeneity weaken individual, family, neighborhood, and social ties. Shaw and McKay (1969) found an association between certain structural conditions and the concentration of social ills such as delinquency, and they attributed the higher prevalence of social ills in socially and economically disadvantaged areas to 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, after controlling for household characteristics, population density and the percentage of individuals under 20 years of age positively predict survey cooperation.

The collection of individual- and, especially, aggregate-level data thus provides 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 that areas vary in the willingness of people to intervene on behalf of the common good (Johnson et al. 2006; Sampson, Morenoff, and Earls 1999). High levels of collective efficacy follow from neighborhood traits such as low population turnover, high education, high income, low density, few immigrants, and more intact families. Research has found that the propensity toward collective efficacy is significantly related to cooperation in surveys (Couper, Singer, and Kulka 1998).

Social isolation theory argues that nonrespondents are likely to be poorly integrated members of society (Groves and Couper 1998; Looseveldt and Carton 2002; Stoop 2005). According to this theory, nonrespondents are likely to be social isolates both because of personal misanthropy and because of social and civic disengagement. Individually, social isolates try to minimize interpersonal contact with others and are disinclined to cooperate with and engage in survey interviews (i.e., a conversational interaction; see Converse and Schuman 1974).

Social isolates have little interest in societal and community affairs and neither follow such matters nor are interested in discussing them in an interview. Thus, for these distinct, but associated, reasons, social isolates are expected to be over-represented among nonrespondents. Under MIDA, it is possible to examine these expectations, both by comparing households that are socially isolated (e.g., with no listed number, no member registered to vote, no member belonging to large voluntary associations, etc.) to less isolated households and by comparing engaged areas (e.g., higher voter turnout, more magazine/newspaper subscriptions) to less involved neighborhoods and communities.

Overextension theory argues that people leading busy lives tend to be nonrespondents (Campanelli, Sturgis, and Purdon 1997; Groves and Couper 1998, 2001; Lynn 2002; Smith 1984). Busy people include those working full time, especially those putting in overtime; those with open-ended management responsibilities; and those whose work involves travel. The category naturally also includes people with multiple, major roles such as full-time employees and parents of small children, or those providing in-home eldercare. MIDA 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, the policies of gatekeepers, and so on influence contact rates and ultimately response rates. Such structural impediments can be observed and recorded by interviewers to better specify the overall nonresponse model.

Interview validation

Interviews are checked or validated through a combination of close supervision of field interviewers, recontacting respondents to verify that an interview has been conducted, and the use of computer audio-recorded interviewing (Smith and Sokolowski 2011). Although invalid interviews are a relatively small component of total survey error (Smith 2005; Groves and Lyberg 2010), MIDA can reduce the number of invalid interviews even further by allowing information from databases to be used along with recontact information to help to corroborate that interviews were truly and correctly done.

Substantive analysis

Finally, information on respondents contained in household- and aggregate-level data files in the augmented sampling frame can be used to undertake crucial substantive analyses. Whereas most household-level information would come from the interviews with the respondents, these household-level data could be supplemented with additional information from the augmented sample frame. Data from the database-augmented sample frame can be used to add information that the survey does not cover, supply missing data for variables that are covered by the survey, and corroborate information reported by respondents.[8]

Research has demonstrated that contextual, aggregate-level geographic variables in general and neighborhood characteristics in particular influence a wide range of attitudes and behaviors independent of the attributes of individuals. Studies have documented contextual effects on political involvement (Bobo and Gilliam 1990; C. Cohen and Dawson 1993; Gilbert 1991); residential and social mobility (Lee, Oropesa, and Kanan 1994; Massey and Eggers 1990; Massey, Gross, and Shibuya 1994; South, Baumer, and Lutz 2003); the sexual and reproductive behaviors of young people and adults (Billy and Moore 1992; Brewster 1994a; Brooks-Gunn et al. 1993; Browning and Olinger-Wilbon 2003; Browning, Leventhal, and Brooks-Gunn 2004; D. Cohen et al. 2000; Crane 1991; South and Baumer 2001); responses to poverty (Jencks and Mayer 1990; McLeod and Edwards 1995; Oreopoulos 2003); racism and tolerance (Gibson 1995); fear of and involvement in crime (Covington and Taylor 1991; Peeples and Loeber 1994; Sampson, Raudenbush, and Earls 1997); minorities politically (C. Cohen and Dawson 1993), economically (Lee, Oropesa, and Kanan 1994; Massey and Eggers 1990), and in other ways (Brewster 1994b; Smith 1994); social capital and health (Mellor and Milyo 2005); group membership and economic improvement (Tolbert, Lyson, and Irwin 1998); inequality and political trust (Rahn and Rudolph 2005); religion and deviant behavior (Regnerus 2003); drug use (Boardman et al. 2001; Ford and Beveridge 2006; Galea, Ahern, and Vlahov 2003; Snedker, Herting, and Walton 2009); and depression (Latkin and Curry 2003).

Among contextual effects that have been documented using data from the General Social Survey (Davis, Smith, and Marsden 2009) are the racial composition of the local population predicts levels of racial prejudice (Alesina and LaFerrara 2000; Charles 2003; Dixon and Rosenbaum 2004; Taylor 1998, 2002) and class voting (Weakliem 1997); higher collective levels of trust and civic engagement are associated with lower homicide rates (Rosenfeld, Bray, and Egley 1999; Rosenfeld, Messner, and Baumer 2001) and lower mortality in general (Kawachi et al. 1997); areas with greater aggregate happiness have lower mortality (Jencks 1999); higher levels of anomie are related to higher local crime rates (Rosenfeld and Messner 1998); community-level differences in attitudes on gender roles do not affect the demand for female labor (Cotter et al. 1998); the prevalence of fundamentalists reduces support for feminism (Moore 1999); a higher amount of people on welfare reduces support for welfare spending (Luttmer 1998); living around gun owners increases one's likelihood of acquiring a gun (Glaeser and Glendon 1998); lower income equality is associated with lower social trust and group membership (Kawachi, Kennedy, and Lochner 1997); community heterogeneity influences civic engagement (Costa and Kahn 2003); community norms shape attitudes toward capital punishment (Baumer, Messner, and Rosenfeld 2003); state and regional differences may be declining over time (Weakliem and Biggert 1999); voting and civic involvement vary by community as well as individual demographics (D'Urso 2003); greater community acceptance of immigrants relates to more occupational achievement by immigrants (De Jong and Steinmetz 2004); community religious beliefs and behaviors influence gender roles (Moore and Vanneman 2003); and aggregate public opinion affects public policies, 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 can facilitate and extend such contextual analyses and make them a regular part of survey analysis rather than an occasional approach carried out only when special multilevel data are added, often after the fact, to standard surveys. Rather than involving an extensive, extra, post hoc effort, the contextual data would be precollected for the entire frame and thus automatically available for contextual analysis. In brief, the information in the augmented sampling frame that can be used to assist data collection and adjust for nonresponse bias can in turn be used for multilevel, contextual analysis.[9]

MIDA expansion over existing practices

While elements of MIDA have been used in one way or another in existing surveys, household- and aggregate-level linkage to databases has not yet been accomplished in an integrated, systematic manner. For example, telephone directories are often used to try to find the name and number associated with a sampled address or to track a panel respondent who has moved. Although such searches are often quite helpful, their use is purely operational, and information gathered by interviewers to locate respondents is seldom, if ever, systematically analyzed, used for nonresponse adjustment, or retained as part of the final data file. Conversely, although linkage data are sometimes collected for substantive purposes (e.g., to see if graduates of a job-training program end up on welfare), this information is generally not used for field operations or nonresponse adjustment purposes.

A second limitation is that the use of multiple databases has apparently never been systematically assessed. Different practitioners use different data sources (e.g., telephone directories, credit records, various public and governmental files) based on their familiarity with datasets or data providers and their general preferences, but no rigorous comparison of the ease of use, cost, and yield of various databases has been conducted, and no study has closely examined the cumulative gain from the use of multiple datasets (Smith 2006).

A third limitation has been that relatively few databases are currently in use for survey research. 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 not useful when no match occurs (as is the case with the large proportion of households with unlisted numbers plus those with no telephone). Being found or not found in a database, however, is in itself a useful piece of information and should be recorded for comparing respondents and nonrespondents. 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; Kennedy, Keeter, and Dimock 2008; 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.

Constructing a MIDA-Enhanced Sample Frame

Sample

To test MIDA in a real-world application, the National Opinion Research Center (NORC) sampling statisticians drew a sample of four hundred addresses clustered in forty segments that was (1) representative of addresses from NORC's national sample frame and (2) similar to the type of sample used in NORC's General Social Survey and other national surveys that it conducts. These forty segments included one or more addresses from fifty-four zip codes. NORC's sample frame utilizes a multistage, area probability design in which national frame areas (NFAs) are selected, segments are selected within the NFAs, and addresses are selected within the segments (Davis, Smith, and Marsden 2009). For sampling strata representing urban areas covered by city-style addresses (i.e., with street name and number), addresses were selected from the Delivery Sequence File (DSF) compiled and maintained by the United States Postal Service (USPS). Access to the DSF is provided to users via USPS Certified DSF[2] Licensees. In NORC's case, they obtained the sample from ADVO (now Valassis Communications, a private vendor). For strata with rural dwellers and less complete coverage (fewer city-style addresses), NORC sent out field enumerators to compile their own address and location lists. The strata covered by the DSF list represented about 85 percent of the U.S. population, and the NORC-listed areas covered the other 15 percent of the population (Davis, Smith, and Marsden 2009; Harter et al. 2010).

Augmenting the sample frame

As noted above, the first step in MIDA is to extract all useful information from the sampling frame. In this case the frame is constructed from two sources, the U.S. Census and the DSF as augmented by NORC's own address listings. The Census Bureau provides a wide range of demographic data from block groups, tracts, places, counties, and metro areas. What is available depends on the geographic unit, with more detailed information being released for larger units. For census tracts, hundreds of demographic breakdowns are available by variables such as age, race, ethnicity, gender, marital status, household size, income, labor-force status, and education (Gatewood 2001).

From the DSF provided by ADVO there are forty-one variables for each address. About sixteen deal with the addresses themselves (e.g., zip code, street number, walk sequence, etc.). Many other variables deal with the status of the address (e.g., seasonal delivery, vacancy, "do not deliver," and business versus residential). Still other variables cover how the mail is actually delivered to the address (e.g., Record Type, Delivery Point Type, and OWGM Indicator [Only Way to Get Mail]). Besides defining the location of the address, these auxiliary variables provide important information about the nature of the unit at the address, such as whether it is a vacant household unit, a seasonal household unit, or a throwback address in which mail is delivered to a PO Box instead of the street address. These and other characteristics are likely to correlate with and predict final disposition statuses of sample addresses used in an actual survey.

Aggregate-level data

Since the address and geographic location of the sampled cases are known, all can be linked to aggregate-level data that are tied to location. Table 1 lists the aggregate-level sources that cases were linked to. While they represent a wide range of sources and variables, they are not exhaustive, but rather are illustrative of the type of information that can be compiled. What can be linked and at what geo-level depends on how the aggregate sources are geographically organized and coded. The sources may have information coded according to exact location in terms of either longitude and latitude or address, or aggregated into various units such as zip code, census categories such as block group and tract, and political units such as place/community and county.

Global Information Systems (GIS)

One way to link aggregate-level information to addresses is via GIS. The longitude and latitude (L/L) of all sample addresses is known, and this enables the addresses to be linked to any other data source that has been L/L coded. A large and growing amount of information is available through GIS databases. It is possible to code Euclidean distances between sampled addresses and various targets (e.g., nearest school or Superfund site) and to categorize these into discrete categories (e.g., within a mile, 1–9 miles, 10+ miles). Alternatively, instead of using Euclidean distance, travel times via the road network can be calculated and used as either continuous or categorized variables.

Among the L/L linked facilities are hospitals, trauma centers, schools, colleges and universities, places of worship, government offices, cemeteries, golf courses, cultural centers such as museums and zoos, major retail centers, transportation

TABLE 1
Aggregate-Level and Geographic Sources

Source	Geographic Unit	Variables
Energy Information Administration	GIS/Distance	Location of power plants
EPA/Superfund Sites	GIS/Distance	Location/type of Superfund sites; 26 variables in table plus follow-up site narratives
Federal Bureau of Prisons	GIS/Distance	Location of federal prisons
HUD/Subsidized Households	GIS/Distance	Location and characteristics of government-subsidized HUs and tenants; 64 variables
StreetPro/MapInfo Professional	GIS/Distance	Location of hospitals, schools/universities, places of worship, government facilities, cemeteries, golf courses, recreational facilities (e.g., zoos, museums), major retail centers, transportation hubs, airports, etc.
Trauma Centers	GIS/Distance	Location and type of trauma centers
PrisonerLife.com	Address	Location of 1,507 correctional facilities
FUNDRACE (huffingtonpost.com)	Address/Zip code	Amount and recipient of donations; 4 variables on amount and party of campaign contributions by zip code
National Center for Charitable Statistics/IRS	Address/Zip code	Not-for-profits except churches; about 42 variables covering types of organization, assets, and various administrative matters
Claritas	Block Group to County	Demographics, housing/property, automobiles, financial, telephone, purchases, outdoors, insurance, audio/video, contributions, medical, interest, high-tech/computers, misc; about 1,000 aggregate variables
U.S. Census, Decennial Census	Block Group to County	Variables vary by geo level. For census tract, hundreds of demographic combinations by such variables as age, race, ethnicity, gender, marital status, household size, income, labor force status, education, etc.
Dun & Bradstreet Businesses	Zip code	40 mostly economic variables on 10 million businesses
EPA	Zip code	Various databases: Envirofacts, Toxics Release Inventory, Facility Registry System, Enforcement and Compliance History
Internal Revenue Service	Zip code	38 variables on individual taxes and income
U.S. Economic Census	Zip code, County	Number, type, size of employers
Association of Religion Data Archives	County	466 variables mostly about number of adherents in specific denominations
Audit Bureau of Circulations	County	Circulation levels for hundreds of periodicals
County Characteristics, 2000–2007 (ICPSR)	County	470 variables, mostly census and governmental variables
FBI Arrest and Offense Figures	County	63 measures of arrests and crimes reported to police
Presidential Election Returns	County/Community/Precinct	Votes for presidential candidates, other offices; sub-county votes are not centrally available

hubs, airports, prisons, military installations, parks and recreational areas, Superfund sites, public housing units, power plants, and rivers or lakes (see Table 1). Examples of studies doing this include Branas et al. (2005) on trauma centers, Downey (2006) and Holmes (1999) on employers, and Salvo and Lobo (2003) on various governmental measures.

Of course GIS-based mapping programs such as Google Maps and MapQuest have a wide range of commercial and noncommercial sites that can be linked to given addresses and stratified by distance. For example, tests of dermatologists, drug stores, churches, synagogues, tattoo, firearms, pizzas, schools, and nails all produced reliable results. (It is noteworthy that "nails" located nail salons [as intended] and not hardware stores.) However, there is no known way to use GIS-based mapping programs in batch mode to search across all target addresses for a given type of site. Nor does one get estimated travel times or distances unless one does a follow-up search on each initial hit.

Addresses

Other information is identified by address, but no GIS data are included. As long as the addresses are in a city-style format, they can be converted to L/L using ArcGIS or a similar routine and then handled as other GIS-based data. Examples include national lists of correctional facilities, political contributions listed under federal election law, and the not-for-profits list maintained by the IRS and provided to users by the National Center for Charitable Statistics (see Table 1).

Small-level census categories

As discussed above, the U.S. Census was one of the original sources of information in the construction of sampling frames, and thus all public census data are potentially available for use in MIDA.[10] What is released publicly depends on geographic level, with less detail being available at lower levels of aggregation, such as blocks, and progressively more for larger geographic units, such as tracts, zip codes, places, counties, metropolitan areas, states, and regions.

Zip codes

Many databases provide information aggregated at the zip code level. These include online resources such as Zip-codes.com,[11] Melissa Data,[12] and Zip Codes To Go.[13] Most of these vendors provide only basic location information and a limited range of census-based demographics. Some extra data are appended to a few databases (e.g., data on income tax refunds and returns in www.zipcodestogo.com). In addition to these zip code–centric databases, many other sources also aggregate by zip code (often along with other geo-units such as block groups, census tracts, community areas, places, and counties). The U.S. Census Bureau itself has regrouped 2000 population data into zip codes[14] and also organized the 2002 economic census by zip code.[15] The IRS likewise makes income tax information available by zip code as well as higher geographic aggregations.[16] The Environmental Protection Agency also offers some zip code–level data and more for larger geounits.[17] Political contributions are aggregated by zip code by number of Democratic and Republican contributors and total amount donated.[18]

Private-sector products are also available from a variety of sources. Claritas Corporation,[19] for example, offers a wide range of census-based demographic data and commercial statistics aggregated by zip codes and other geographical units. MarketPlace by Dun and Bradstreet[20] codes more than 10 million U.S. business establishments with respect to more than forty variables by zip code and county (Powell et al. 2006).

Zip codes are widely used in epidemiological and medical research (Grubesic and Matisziw 2006; Krieger et al. 2002a, 2002b), most commonly to correlate the distribution of diseases and medical diagnoses with socioeconomic factors (Pappas et al. 1997; Thomas et al. 2006) and population statistics (Peel et al. 2005). When zip code data are based on matches to the U.S. Decennial Census statistics, the assignment of zip code values depends on the matching of block-group data to the larger zip code areas. While reliable procedures for such assignment have been developed, there is an element of estimating and approximating in aggregating census data by zip codes, or what the Census Bureau designates as ZIP Code Tabulation Areas (ZCTAs).[21] When the original data are collected by zip code, street address, or GIS indicators that can be definitively assigned to zip codes, the assignment errors that occur while retrofitting census data to zip codes do not arise.

Counties and metropolitan areas

A wide range of information is available by county and metropolitan area. Of course any data collected at a more detailed level, such as census tracts or zip codes, can be aggregated at the county or metropolitan level as well. Examples include information from both the decennial population and economic censuses. Additional data available at the county and metropolitan levels include denominational adherents; circulation levels for periodicals; reported crimes and arrests; election returns; and other variables relating to governing, such as variables about taxes, public expenditures, Medicare enrollment, and building permits (see Table 1).

With respect to election returns, turnout at the county level is easily assessable, but voting rates for communities, precincts, and other smaller units are much less available (Committee on State Voter Registration Databases 2008, 2009), though the situation varies greatly from state to state. In Indiana, the Board of Elections maintains subcounty results only on hard copy and nothing is available online. Pennsylvania has detailed figures online, but one needs to go separately to each county's website to gather the results. Iowa has precinct-level results by county at a centralized site. At this point, however, it appears that no national compilation of community- or precinct-level votes has been attempted. Even when available, subcounty data are challenging to work with. Although it is relatively easy to match addresses at the community level with voting results, identifying what precinct an address is in is much more difficult.

Geo-demographic segmentation

Geo-demographic segmentation, sometimes referred to as geo-demography, is a multivariate classification procedure for dividing areas into distinctive sociodemographic types. It is used especially in marketing to define lifestyle groups that would help in the targeting of particular types of consumers. The most widely used systems are PRIZM NE by Claritas, Mosaic by Experian, and Community Tapestry by ESRI.com. PRIZM NE has identified sixty-six lifestyle cluster types, whereas Community Tapestry has sixty-five and Mosaic sixty. By way of example, PRIZM NE assigns types to individual households based on a combination of census data compiled at the block-group level as well as unspecified other noncensus information.

Address-/household-level data

Although aggregate-level data are generally available for all cases, the availability of information for particular addresses depends on whether they are linked to various databases. Household-level linkages are only possible when dwellings have city-style addresses (unit number and street name), which are traditionally absent in rural areas. Fortunately, to ease the tasks of first responders, as part of enhanced post-9/11 efforts, city-style addresses are being assigned to more and more dwellings (DiSogra, Callegaro, and Hendarwan 2009). The effort to do so is being coordinated by the LACSLink system of the USPS (Key and Miracle 2009). In the NORC sample of four hundred addresses, 95.8 percent had searchable street-style addresses. It is anticipated that this percentage will be even greater in the future.

It is also noteworthy that an additional 3 percent of addresses could not initially be linked. All these addresses were from one rural segment that had a place name associated with it that differed from that used in other sources. Even though it was easy to determine an alternative place name for these cases to render them linkable, this cautions that alternative or changed names may be used across different sources, and one needs to be sensitive and flexible to such variation when matching addresses.

General record-linkage issues

Searches of and linkages across databases depend in part on which search algorithms are employed. These are rarely, if ever, explicitly specified, and one learns about some of their properties only from usage. The most limiting features of algorithms are those that require an exact match across all elements of search terms. For example, a restrictive system would be one that used the full capitalized word "Street" in its records and would not recognize "street" or "St." as matching terms to it. Other examples are systems that would not match the search for "234 Maple" when their records contained "234 Maple St." Alternative spellings or misspellings are also problematic. Many systems require that place names be spelled exactly the same so that "Olde Boalsburg Road" would not match "Old Boalsburg Road" or "Pittsburg" would not be linked to "Pittsburgh." Other systems, however, locate phonetic equivalents and would match the two previous examples. Of course, if a misspelling is too divergent even algorithms allowing for phonetic equivalence or similarity will not find matches. Additionally, very permissive algorithms will increase false positive matches (e.g., confusing "234 Maple Lane" with "234 Maple Drive"). However, given the procedures actually used, false negatives appear much more common than false positives.

When a match is made, the next issue is whether information on the current resident or residents can be identified. Multiple listings per address are the norm, and they may or may not include the current residents. Dates on which a named person is associated with an address are frequently available, but they often end short of the current date, are often incomplete, and are sometimes contradictory. In most cases the likely current residents can be identified. For additional cases two or more possible current residents are suggested. For a small number of cases, no likely current residents appear (i.e., at least one name is associated with the address, but there is no evidence that that person currently resides there).

For the current residents, the issue becomes what additional information is available. For the reverse-directory searches, the situation is simple; a found address generates a name and phone number. For many other databases, a wide range of potential information may be available. Some variables, such as gender, may exist for almost all linked names, but others, such as age or date of birth, may be available for substantially fewer people, and still others like social security numbers may be relatively rare.

Finally, even when matched information is available, the issue arises as to whether it is accurate. The legal disclaimer by one database illustrates this point, "Important: The Public Records and commercially available data sources used in this system have errors. Data is sometimes entered poorly, processed incorrectly, and is generally not free from defect. This system should not be relied upon as definitively accurate. Before relying on any data this system supplies, it should be independently verified. For Secretary of State Documents, the following data is for information purposes only and not an official record. Certified copies may be obtained from that individual state's Department of State." Errors can consist of information that is simply wrong (e.g., transposed numbers, misreports in original records used by the databases) or out of date (e.g., employment information that refers to a previous job, marital status before a recent marriage/divorce).

Multiunit addresses

Multiunit addresses include everything from duplexes to very large apartment complexes and represent all situations when two or more household units have the same street number for their address. (For brevity's sake, these various types of multiunits will be referred to as "apartments.") Addresses with multiple housing units at the same street address present special challenges. First, some such units, especially smaller ones such as duplexes, do not have official or regular unit designations such as numbers, letters, or number-letter combinations. They may have no designation or only informal and irregular designations (e.g., "rear," "basement"). Second, many databases do not recognize apartment numbers as searchable fields, and thus it is often not possible to search for specific apartments.

Third, even when regular unit designations exist and as such are a field in a particular database, the full information may not exist for units or people associated with a particular street number. Some may have an apartment number listed, and others have no specific listing. When an apartment is searched for in various databases, one will often get a list of all people associated with that street address regardless of their apartment number, or one name will be reported following some default heuristic. For example, Accurint (which manages the LexisNexis database) reports only the name of the person with the surname that comes first in the alphabet. In NORC's sample of 383 searchable addresses, 24.2 percent involved multiunit addresses.

Using record linkage outcomes

The initial result of any database search is that the target address is either found or not found. Except for addresses in unsearchable, non-city-style addresses, a result of "not found" generally means that the particular address is not in the database (e.g., has no listed telephone number or no registered voter). Of course, errors or variations in the listing of addresses in either the sampled addresses or the target database may also cause some nonmatches, or false negatives, and conceivably even some false positives. The outcome of each search (found or not found) should be recorded and used as a measure for studying nonresponse bias. Since a found/not found code will be generated for all sample addresses and all linked databases, one can construct an inter-database measure of how often an address appears across records. This aggregate measure may prove to be a good predictor of nonresponse in general and may be especially good for identifying households that are socially disengaged and "flying below the social radar." The more databases an address is found in, the more engaged that household and its residents may be considered to be.

While some addresses in the NORC study could not be linked to particular databases, this does not mean that no nonresponse bias assessment is possible. Instead, the unlinked cases become a category for analysis. For example, after linking all addresses with city-style addresses to a reverse directory database, one can classify all cases as having a listed phone number versus not having a listed phone number and examine response outcomes with respect to this dichotomy. After linking to a database with information on the gender of adults, there would be the unlinked households and those for whom gender information is known (e.g., only males, only females, an opposite gender couple, other, and both gender combinations), and the response rate for each of these five categories could be examined. Thus, all cases in the sample are covered and retained even when no linkage occurs.

Databases

There are a large and expanding number of data sources that can be linked to addresses either on the household or aggregate level. Data sources come in complex and ever changing types. While the major types are separately discussed below, there is considerable variation across databases within categories and considerable overlap across databases within different categories.

Reverse directories

There are various types of "reverse directories" (e.g., phone number to name, phone number to address). Our interest is in reverse directories from address to name and/or phone number. From the master national directory of listed phone numbers maintained by the phone companies, a number of providers offer online address look-ups. Principal sources include 411.com, address.com, infoUSA411.com, phonenumber.com, whitepages.com, and yb.com. All these and several others were tested by NORC. Given that they all depend on national phone directories as a primary source, it is not surprising that they yield similar, though not identical, results. A comparison across various reverse directories at times produced virtually identical results, indicating that they were using the same editions of the telephone directories and had apparently not enhanced them in any way. Across other reverse directories, more differences appeared. For example, addresses that had yielded no hits in 411.com were run against a combination of yb.com, infoUSA411.com, and whitepages.com. For 27 percent, a hit was found; for 7 percent, a link at the apartment address without a specific apartment number was found; and for 3 percent, a possible alternative description of the address was suggested. It is unclear whether these additional links came because different editions of the telephone directories were being used, because the lists were enhanced from other sources,[22] because the address matching algorithms differed, or because of a combination of factors. The results do indicate that multiple sources should be consulted to maximize the linkages.

One common limitation is that reverse directories do not accept apartment number as a predefined, searchable field, and this complicates searches involving apartments. However, it was discovered that some would recognize apartment numbers if inputted as part of the street address field. Moreover, in the output from the search, the apartment number of found people was often included and thus could be used to identify the correct unit among those at the same street address. While it might be possible to develop a way to search the output automatically for the apartment matches, at present this must be done manually.

List providers

List providers offer both general lists and many types of specialized lists (from boaters to voters). In general, each list is a standalone, and typically there are no or limited attempts to compile a dossier on individuals or addresses by merging across lists, but even within this category there are exceptions. Among the more prominent list providers are Century List Services, e-merges.com, InfoUSA.com, and U.S. Data Corporation. Catalist, e-merges.com, and registeredvoterslists.com cover voter registration lists specifically.[23] These differ from some other lists in that they more frequently have data from outside the public registration information appended to them in what is commonly called "data enhancement." Lists may also be "groomed," updated, verified, or pruned of "deadwood." This is done in various ways by the list providers, including incorporating address changes, cross-checking with other lists, and checking death records.

General address-searchable databases

A wide variety of sources allow searches by addresses. In general these sources include more information than the reverse-directory databases but less credit and financial information than the credit reports. Often one can obtain minimal information about an address for a lower price and needs to pay a premium for more complete information. The exact information available and its format vary considerably across providers. Some produce a list of only what appears to be people currently/recently associated with the address, while others are more inclusive and list even people associated with an address many years ago. They do not document their criteria for inclusion. Some allow individual addresses to be searched by investigators, and others require batch processing of addresses by the data provider itself. The sources consulted include Accurint.com, atxp.choicepoint.com (AutoTrackXP), Donnelleymarketing.com/infoUSA.com, govdmvrecords.com, government-records.com/public-records-search.com, infoUSA411.com, Intelius.com, Peachtreedata.com, Peoplefinders.com, peoplesearchnow.com, and Targusinfo.com. As an example, information about using Accurint is presented.

Accurint

Link searches can be done either in batch mode, in which a list of addresses is processed by Accurint, or in individual mode, in which the investigators enter addresses one at a time. To describe the capabilities of Accurint, examples using individual mode are presented. One starts with the Person Search template and since one has only address information to start with, one fills in the location fields and leaves fields for name, social security number, telephone number, date of birth, and age range blank. One can then select Person Search or Advanced Person Search,[24] which produces a list of people associated with the queried address. The list generally includes any names associated with the address as well as some quite old and outdated linkages. It often includes multiple listings of the same person under slight variants of his or her name (e.g., full name, nickname, middle initial, middle name, etc.). Generally, the most recent linked person is listed first and probable current resident is often indicated. For each linked person, it will report name and as much of the following information as it has for each named person: gender, date or year of birth and age, age at death, telephone number, address, dates at residence, and social security number (SSN).

After this initial search, several possibilities exist. Especially if it is unclear who is likely to be the current resident, one may want to do a second Person Search for one or more of the linked names. In particular, one can enter the name and see if that person seems to reside at the sampled address or lives elsewhere. This can help to clarify if the named person is still at the target address. Then, either directly from the initial linkages or after the intermediate search to clarify who is the current resident, one can download various extended reports for each linked individual. One such report is the Summary Report, which indicates matters such as state in which and the approximate time the SSN was issued; "others associated with SSN"; and a yes/no indication for bankruptcy, property, and corporate affiliations.

Another report is the Comprehensive Report, which includes information on who owns the property, probable current and past residents, land use/zoning, names and other information on neighbors, listings in numerous public records such as hunting/fishing permits, professional licenses, aircraft registration, concealed weapons permits, property assessments, utility hookups, voter registrations, place of employment and position there, possible relatives, possible associates, and possible former residences. As part of the Comprehensive Report, but also available separately, People at Work reports on place of employment and position. Some of this information can also be requested in a diagram format, Relavint, that "will help you visualize the relationships between people and their possible relatives and associates, vehicles, property, and even businesses."

Credit reports

There are three major providers of credit reports (Equifax, Experian, and TransUnion). These companies are oriented toward providing credit scores and relating financial information about individuals to lenders and businesses in general. While focusing on providing this information for named individuals, they all allow searches based on more limited information such as address alone.

Credit reports are generally not accessible by researchers with new samples, but under some circumstances these could be utilized in panel studies to follow up with respondents.

Specialized change databases

Several types of databases specialize in people changing statuses such as moving or dying. Using these specialized databases can help to clarify who is the current resident at an address, and of course they are especially useful in panel studies. New movers databases are one change type. The two main sources for such information are the change of address listings from the USPS known as NCOA and telephone number changes and connects (e.g., New Connect/New Movers from Telematch.com and New Movers from infoUSA.com; Hotline List of New Telephone Listings from newmoverslist.com). In some cases these are supplemented by listings from utilities, governmental property records, magazine subscriber move data, and other "proprietary sources" (e.g., Hotline PLUS from newmoverlists.com, New Home Owners Lists from directmail.com, New Movers from infoUSA.com, New Movers from cas-online.com).

Another specialized source for changes comes from various death records. While used more commonly in longitudinal panel studies and epidemiological research, death records can be consulted to see if names associated with an address are still living, which can be especially useful when the dates associated with the persons linked to an address are several years old or the identified person is elderly. Although several databases do indicate if a person matched to a particular address is deceased and will often report a date or year of death, such notations are not universal and may not be up to date.

The best source is the National Death Index maintained by the National Center for Health Statistics.[25] According to the website, this source is "available to investigators solely for statistical purposes in medical and health research." Another source is the Social Security Death Master File (see Hill and Rosenwaike 2001–2002). For people 65 and over, it covers 93 to 96 percent of deaths in the National Death Index. The Social Security list is accessible via various commercial sites.[26] Other sources include state death records.[27]

Extended searches

If initial address searches come up with names of residents or phone numbers for the address, follow-up searches can be made using these identifiers. Such follow-up searches can serve several purposes. They can help to determine current residents when multiple people are associated with a particular address because of either people from different apartment numbers not being distinguished or turnover in tenancy, to confirm that the located person does currently reside at the sampled address (i.e., has not moved away), and to add additional information about the located person.

Most of the databases used in the initial, address-only search will permit further searches with the added information on name and telephone number (see Accurint example above). In addition, there are numerous other databases that do not allow address-only searches but can be utilized once the additional identifiers have been obtained. These include Claritas.com, in-foquest.com, statewidegovrecords.com, telematch.com, and theultimates.com.

These extended searches can fill in information that is missing from the initial linkages. For example, an initial source may turn up a name, but no demographic information. A follow-up search using name may then add certain demographic information such as age. Or it might turn up names, but no clear evidence on who is the current resident. Searching on the listed names can often clarify who is likely to be currently at the address in question. Similarly, an initial search may turn up a phone number but no name, and that phone number could lead to a name and other information emerging from subsequent searches. If a follow-up search yields new information, that information could lead to another search that in turn could yield more leads and further fruitful searches. Pooling information across databases and from a combination of initial and extended sources will collect more complete information on the residents of sampled addresses and lead to a better assessment of nonresponse bias.

An illustration of database search results

To demonstrate MIDA's use of multiple, address-level sources, results from searching five databases were merged together. These included three general, address-based sources (Accurint, Peachtree, and infoUSA), one reverse directory (411.com), and one list of registered voters (Catalist). Of the 383 cases with street-style address, some linkage to databases was found for all but 9 addresses. These unlinked addresses were searched for in various other maps and address-based databases (e.g., Google Maps, MapQuest, American Fact Finder). Two were found to represent an area that had changed both its place name and zip code, and one also had a misspelled street name. One could not be located in any source, which may indicate some problem with the address such as an error in the post office list or a change in place name. Whether this represents an errant address for a residence that actually exists or a nonexistent residence is unclear.

Four were located and appeared to designate actual residences. These could represent long-term vacancies, very recent construction, or residences occupied by people who have managed to avoid inclusion in a wide range of databases. Two were recognized as legitimate addresses but showed either a vacant lot or open country at the point associated with the address. Of course, since the mapping programs show only approximate locations, this may reflect their limitations rather than the absence of a residence linked to the address. These handful of uncertain cases would be resolved if the sampled addresses were utilized as part of an actual survey rather than merely at the sample frame augmenting stage as in this preliminary study.

When the unlinked and non-street-style addresses are considered together, only 6.5 percent of addresses had no information at the household level outside the sample frame. This is much lower than the 26 percent unlinked by Raghunathan and Van Hoewyk (2005) in a similar study. This difference reflects mostly their use of a single database and may also reflect changes from 2005 to 2009 and more thorough search strategies. A last name was obtained for 97.4 percent of the city-style addresses, age for 93.7 percent, social security number for 88.8 percent, phone number for 83.1 percent, gender for 78.6 percent (about 95.1 percent when first names were used to infer gender), race/ethnicity for 78.1 percent, income for 66.1 percent, occupation for 56.0 percent, and education for 37.0 percent. Using multiple databases of course notably increases the proportion of addresses with information. For example, in individual sources age was available for 24.5–78.9 percent of addresses and from at least one source for 93.7 percent of addresses.

Of course, when particular information comes from a specific, public record, one needs to use a database that accesses that source rather than consult multiple databases. For example, from the voter-registration database (Catalist) and the federal, political-contributions database (FUNDRACE), one can determine if the sampled address has one or more registered voters and whether any federally regulated political contributions were made. One or more voters were present at 67.3 percent of addresses. 44.6 percent had a voter in the 2004 general election, and 43.2 percent had a voter in the 2008 general election. Voting status was also related to other variables such as resident type. For example, 75.3 percent of the nonapartment addresses and 43.2 percent of the apartments had at least one active voter. For those addresses with a registered voter, their party is of course known: 46.4 percent had a Democrat, 28.1 percent a Republican, and others included independents and third-party registrants. For the 2008 campaign cycle, none of the addresses had a confirmed federal political contribution.

Aggregate- and household-level data can be used separately, or they can be joined together and analyzed. A few examples will illustrate the type of comparisons that can be made: using data from the census on track-level racial composition and from the Postal Service's DSF on housing type (apartment/not apartment building) showed no association between these variables. Linking the Superfund database with information on having a phone number also showed little association. In areas with no Superfund sites, 39.2 percent had a listed phone number; and in areas with one or more sites, 37.9 percent had a phone number; the aggregate data on public housing units showed an association with having a listed phone number. The association was 40.2 percent in areas with few units and 33.3 percent in areas with higher public housing density.

Limitations

A number of factors do limit or complicate the use of databases for MIDA. First, there are legal restrictions. Even when there is a database that permits address searches and has content of interest, use may be restricted. For example, the various providers of national voting and registration lists are limited by state laws as to both what they can provide to users and what type of users may access the records for that state. Thus, to access the Illinois records, one needs a letter from the State Board of Elections or Illinois attorney general authorizing the search. For California, the records may be used only for a political purpose. Information from driver's license records are restricted by the provisions of the Driver's Privacy Protection Act, which generally excludes any research use. The Gramm-Leach-Bliley Act regulates access to other types of records.

Second, several general factors complicate the extraction of information from databases. Most databases are limited in the documentation that they supply about their data. The original source of information is often not indicated, nor its recency. Likewise, quality-control procedures are never detailed. Definitions and data-procurement procedures are often not indicated. Even obtaining limited information usually involves considerable digging or special requests. However, database providers will usually clarify matters and often supply additional documentation upon request. Also, additional information comes from using the databases and becoming familiar with their features through application and comparison. The impediment was usually not that information was being withheld to cover up poor procedures or serious flaws but that the information and documentation had never been compiled, had not been prepared for dissemination, or were restricted to protect proprietary interests. In addition, most databases collect specific information from various sources and include it when found and omit it otherwise. There are therefore many gaps in the data matrix. Thus, a given source will typically include certain information (e.g., last name) while omitting other data (e.g., race or marital status).

Third, much information is obtained from state-level, public records. The content, form, and availability of this information (e.g., voter registration lists, property records, professional licenses) vary from state to state, and thus the information is often not uniform across states. Databases, however, do harmonize data to reduce this problem.

Fourth, changes across products and firms are fairly frequent. The nature of databases and who maintains them often varies over time, and one needs to continually update information about databases and their providers. Fifth, many databases are frequently updated. This is generally a positive attribute since it means that new information is being added; but it also means that using the same database for the same addresses just a few months apart can produce appreciable differences in results. Sixth, information in the databases may be out of date. While accurate when compiled, it may not reflect more recent changes in statuses. This may involve alterations in a person's personal characteristics (e.g., a marriage or divorce, or a job change) or changes in official information such as a new place name or zip code for an address. While it does occur, out-of-date data does not seem to be a major problem.

Seventh, errors in either the original source record or in the databases compiled from original sources occur. Some errors were detected in all records. Even the Postal Service's master DSF wrongly classified a few apartment buildings as nonapartments. However, by cross-checking across multiple records, the accurate status was usually determinable. Consulting multiple sources notably reduced the level of error.

Eighth, information about addresses is about the households that reside there and the individuals who reside in the households, not about a specific respondent. The household-level information is one level of aggregation above that of an individual household resident selected as a respondent. As such, it is very useful in determining the attributes of households that yield respondents versus nonrespondents, but it is not a direct comparison of respondents and nonrespondents. In most cases, however, it is relatively easy to match an individual from a household-level database to a respondent selected from the household members. Based on the 2002–2008 General Social Surveys, 18 percent of respondents live in a household with no other adults and 53 percent live in a household with two adults of opposite gender. Thus, for 71 percent of respondents, the identity of the respondent can be readily matched between the residents listed in the database and the respondent selected in the survey. For other households, some matching is possible, but it becomes more difficult and less certain.

Ninth, apartments are especially challenging since most databases do not allow explicit searches for units. The extra complication of apartments is illustrated by looking at the number of names listed from the reverse-directory search. For nonapartments, there were 2.9 residents on average listed per address. For apartments, an average of 19.0 people per address were listed. Most of the large excess results from the fact that the reverse-directory database, like most, did not distinguish among individual units in an apartment building. That is, the names represent people associated with the street address of the building and not the individual unit in the sample. In addition, the greater turnover in apartment renters versus homeowners also contributes somewhat to the higher number of people associated with apartment addresses.

When dealing with apartments, one needs to (1) determine if there is a way to get units recognized, (2) try alternative input formats both within and across databases, (3) examine the output to see if unit is indicated at this stage, and (4) do follow-up searches across databases to clarify who resides in a particular unit. By taking these steps, links can be made for most apartment units.

Tenth, even among street-style addresses, a few will not link to residents even across multiple databases (9 of 383). The existence and nature of addresses in general and unlinked addresses in particular can be examined by running addresses in various mapping programs such as Google Maps and MapQuest. These could often clarify whether an address exists; indicate whether it is a residence, business, or other; and show if it is a single-family residence or a multiunit dwelling. Of course, these sources are also invaluable for interviewers in the field who need to locate the address to conduct an interview.

Finally, there are several challenges linked with using the names found in various databases and associated with a particular address. Many databases will list multiple names associated with an address.[28] This often includes many names associated with an address at some point in the past. Many databases do indicate dates associated with a particular person and often designate the most recent resident as the likely current occupant. As noted above, follow-up searches of names associated with an address can be done to figure out who the current residents are.

In addition, many databases list the same individual multiple times under slight variations. While it is usually obvious that these are the same person, that is not always the case. In some cases, the records are intentionally unclear. For example, it used to be common for women living alone to use only their initials or even a late husband's first name to disguise that theirs was a female-headed household. Another problem is that shared family names often make it possible to infer which named individuals are members of the same households, but the same family names are not used by some married couples, most cohabitating couples, and members of some mixed or blended families. Finally, some databases such as telephone directories list only a single name for a household and thus ignore other household members. In addition, most databases do not list the names of minors in the household. Careful attention to details and comparison across sources is needed to clarify the household composition of many households.

While it is of course desirable that all information is complete and accurate, data can be useful without being perfect. When the databases are being used to identify a respondent at the target address or to follow up with a mover in a panel study, complete accuracy is needed. One needs to identify the respondent or figure out where the respondent now resides. But when nonresponse bias is being examined, data with some error can still be useful. For example, owing to moves and other factors, voter registration information will never be completely up to date. Voter information for an address may not reflect recent registrations or moves. But most voter information for sampled addresses will be accurate (i.e., reflecting the current registration status of the present residents).

In addition, the information can be seen as reflecting the situation on the address at the most recent time the voting records were updated, and that should be useful in accessing nonresponse bias even when individual changes have occurred for some addresses. For one thing, it is likely, but not established by this research, that addresses tend to be occupied over time by people with similar sociodemographics. For another, the address or housing unit is an element in the sample, above that of the respondent but below the other aggregate-level units (e.g., block group, census tract, zip code). As long as the address-level information is accurate, it can be used to identify the sample and to discriminate between addresses yielding respondents versus nonrespondents, even when the occupancy of some household units has changed since the records were compiled. When the address-level information is complete, accurate, and up to date, one can compare the characteristics of respondents and nonrespondents. When the address-level information is less perfect (but still generally accurate), one can still compare the characteristics of addresses yielding respondents to those with nonrespondents just like one can look at areas (e.g., tracks, zip codes, etc.) with higher and lower response rates.

Further MIDA Tests and Full Application

Having shown that MIDA can be used to augment a national sample frame, the next step is to test its utility in an actual survey. Specific application of MIDA regarding nonresponse bias detection and correction, and substantive analysis are described below.

Assessing and adjusting for nonresponse

A dataset constructed using MIDA will contain much more data about nonrespondents than are usually available. The full dataset will have household- and aggregate-level data for both respondents and nonrespondents. Such a rich dataset is uncommon in nationally representative demographic or attitudinal surveys. It provides an opportunity to explore different approaches to estimating and adjusting for nonresponse bias. For the many variables for which the dataset contains values for both respondents and nonrespondents, it will be possible to explore the effects of nonresponse 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 nonresponse bias. This knowledge will then inform our understanding of the error implicit in estimates from the survey variables themselves, for which nonrespondent data are not available (Gelman and Carlin 2002; Geronimus, Bound, and Neidert 1996; Groves 2005, 2006; Marker, Judkins, and Winglee 2002; Meng 2002; Zanutto and Zaslavsky 2002).

In addition, the availability of data on nonrespondents can improve weighting techniques. For example, in recent rounds, NORC's General Social Survey has incorporated a nonresponse adjustment at the level of the PSU, which may be metropolitan areas or nonmetropolitan counties. It assumes that the nonrespondents 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 nonrespondent adjustment used in national, in-person surveys. But the use of PSUs to form nonresponse 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 connections with the study variables. The MIDA-enriched dataset, by providing data on both respondents and nonrespondents on many variables, will allow for more discretion in creating nonresponse adjustment cells and for more sophisticated weighting adjustments (Bethlehem 2001; Kalton and Kasprzyk 1986).

Response-propensity weighting is a common method of adjusting for nonresponse. The theory behind this approach is that all cases, both responding and nonresponding, have a nonzero propensity to respond that can be estimated. 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, and so on. Responding cases are then weighted by the inverse of their response propensity to account for the nonresponding cases, with low-propensity cases given more weight than high ones. Like the nonresponse weighting adjustment discussed above, this method often suffers from a lack of frame variables: these variables are usually those that are available for all cases rather than those that would be most appropriate. MIDA permits 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 will greatly improve the weighting methods used on surveys in general. Having more variables in the MIDA dataset will also improve imputation techniques. Hot-deck imputation fills in values that are missing due to item nonresponse 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, Judkins, and Winglee 2002).

Substantive analysis

Auxiliary data from MIDA can of course also be used for substantive analysis. For example, the wide range of content in the GSS would be particularly well suited for examining the utility of multilevel, aggregate data. Contextual analysis has already been shown to be very valuable in analysis using the GSS (see references cited above), and the wide content will provide a broad test of the value of such contextual data. In addition, the greater use of the GSS will ensure that the contextualized data will also be widely utilized by researchers. The public data file will be constructed to ensure that no deductive disclosure of respondents will be possible. Files with more detailed information, but not personal identifiers or information readily allowing deductive disclosure, will be made accessible to researchers following standard, limited-access protocols to ensure confidentiality.

Conclusion

MIDA has the potential to advance social science research by notably improving survey-research methodology. It does so by drawing on a major societal change of recent decades: the development of large-scale, computerized databases that hold extensive information about individuals, households, neighborhoods, and other societal units. Methodologically, it should help to increase response rates, allow for a much more comprehensive assessment of nonresponse bias, and facilitate the calculation of weights and imputations to adjust for the detected nonresponse bias. Besides providing for a general approach to deal with nonresponse, it will, in particular, permit the testing of several prominent theories and hypotheses explaining nonresponse: social disorganization theory, social isolation theory, overextension theory, structural impediments, and so on. In addition, the auxiliary data from the databases will permit an examination of general, nonresponse models (Groves and Couper 1998, 2001). Substantively, MIDA will improve analysis by easily and automatically making multilevel, contextual variables as ready for analysis as are data directly collected in surveys. As the list of examples cited above attests, 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 partial pilot study described here demonstrates the richness of both aggregate- and household-level data sources and the practicality of linking both to a national sample of addresses. However, it also shows that using multiple databases and linking them both to each other and to a national sample of addresses is a complex and challenging task that must be carried out carefully. Nonetheless, when available databases are rigorously utilized to augment sample frames with aggregate- and household-level data, survey research clearly benefits in several ways: data collection efforts are enhanced, nonresponse bias may be measured, nonresponse bias may be corrected using improved weights, and contextual data may routinely be added to survey data files to enhance substantive analyses.

Notes

1. 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).

2. Survey researchers already have considerable experience in using databases. Other experts include data librarians, geographical information systems specialists, cyber-information technicians and data miners, and records searchers such as paralegals and investigators.

3. 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 to 100 percent of cases (Geronimus, Bound, and Neidert 1996; Groves and Couper 1998; Kim, Smith, and Sokolowski 2006).

4. For a multilevel analysis, see Bryk and Raudenbush (1988), DiPrete and Forristal (1994), and Raudenbush and Bryk (2002).

5. This is obviously not possible for postal surveys.

6. While databases have been used for some time to assist surveys, their use has been informal and underdocumented. For example, many random-digit-dialing (RDD) surveys routinely run telephone numbers by business lists of phone numbers, but this is often not mentioned, and providing details on the removal rate is even rarer (Merkle et al. 2009). On the General Social Survey, hard-to-contact households are routinely searched for to get a name or phone number to aid in making contact with and obtaining an interview from the cases. However, no systematic record of the searches or their outcomes has been maintained.

7. It is likely that some types of information will be most valuable at the data collection stage and others at the nonresponse adjustment stage. For example, name and telephone number would be most useful

to aid the fieldwork, and having a listed/unlisted telephone number, mobility history, and housing tenure would likely be more valuable for nonresponse adjustments.

8. Examples of collaboration are the voter validation studies. See Anderson and Silver (1986); Burden (2000); Silver, Anderson, and Abramson (1986).

9. What is important information will depend in part on the stage at which it is being used and how it is being used. At the initial interviewing stage, finding the name of the likely resident and his or her phone number can be very useful in making contact and gaining cooperation. In a panel survey, finding a current address for a mover or the person's place of employment is valuable for locating him or her. For assessing nonresponse bias, name and actual telephone number are of little use, while information on age, gender, voting status, and so on can be very useful to check the representativeness of the interviewed sample.

10. This includes the American Community Survey as well as the decennial census.

11. See www.zip-codes.com.

12. See www.melissadata.com.

13. See www.zipcodestogo.com.

14. See censtats.census.gov.

15. See censtats.census.gov; www.census.gov/geo/ZCTA/zcta.html; www.census.gov/epcd/www/zipstats.html.

16. See www.irs.gov.

17. See www.epa.gov.

18. See www.fundrace.huffingtonpost.com.

19. See www.claritas.com.

20. See www.sales-tools.com/MP/mpprod.htm.

21. For one such matching and the error associated with it, see www.census.gov/geo/ZCTA/zcta.html as well as www.census.gov/epcd/www/zipstats.html (Grubesic and Matisziw 2006; Krieger at al. 2002a, 2002b; Thomas et al. 2006).

22. Some reverse directories offer links to allied databases where follow-up searches for unfound addresses can be conducted. For example, whitepages.com links to peoplefinder.com. In other cases, additional information on a found address can be accessed within the reverse directory itself. It also appears that some reverse directories augment the information they receive from the telephone directories.

23. On state voter databases and efforts to improve them, see Committee on State Voter Registration Databases (2008, 2009).

24. When starting with addresses only, these two search routines are essentially the same.

25. See www.cdc.gov/nchs/ndi.htm.

26. See, for example, ssdi.rootsweb.ancestry.com; www.genealogybank.com.

27. See, for example, death-records.net.

28. A large number of names per address are most often associated with it being a multiunit dwelling. Among single-unit dwellings, more names are associated with larger household size (i.e., more adults; not minor children since they are not listed); more turnover, which would be associated with its being a rental property rather than owner-occupied; and an older dwelling, for which there could be more former residents associated.

References

Alesina, Alberto, and Eliana LaFerrara. 2000. Participation in heterogeneous communities. *Quarterly Journal of Economics* 115 (3): 847–904.

Anderson, Barbara A., and Brian D. Silver. 1986. Measurement and mismeasurement of the validity of self-reported vote. *American Journal of Political Science* 80 (4): 771–85.

Audit Bureau of Circulations. 2005. *Turning information into inspiration: Annual report 2005*. Arlington Heights, IL: Audit Bureau of Circulations. Available from http://www.accessabc.com (accessed 12 June 2012).

Baumer, Eric P., Steven F. Messner, and Richard Rosenfeld. 2003. Explaining spatial variation in support for capital punishment: A multilevel analysis. *American Journal of Sociology* 108 (4): 844–75.

Berge, Keith H., Deborah R. Maiers, David P. Schreiner, Stephen M. Jewell, Perry S. Bechtle, Darrel R. Schroeder, Susan R. Stevens, and William L. Lanier. 2005. Resource utilization and outcome in gravely ill intensive care unit patients with predicted in-hospital mortality rates of 95% or higher by APACHE IIIs scores. *Mayo Clinic Proceedings* 80 (2): 166–73.

Bethlehem, Jelke G. 2001. Weighting nonresponse adjustments based on auxiliary information. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, Roderick J. A. Little, and John L. Eltinge, 275–88. New York, NY: John Wiley.

Billy, John O. G., and David E. Moore. 1992. A multilevel analysis of marital and nonmarital fertility in the U.S. *Social Forces* 70 (4): 977–1011.

Boardman, Jason D., Brian K. Finch, Christopher G. Ellison, David R. Williams, and James S. Jackson. 2001. Neighborhood disadvantage, stress, and drug use among adults. *Journal of Health and Social Behavior* 42 (2): 151–65.

Bobo, Lawrence, and Franklin D. Gilliam Jr. 1990. Race, sociopolitical participation, and black empowerment. *American Political Science Review* 84 (2): 377–93.

Brace, Paul, Kellie Sims-Butler, Kevin Arceneaux, and Martin Johnson. 2002. Public opinion in the American states: New perspectives using national survey data. *American Journal of Political Science* 46 (1): 173–89.

Branas, Charles C., Ellen J. MacKenzie, Justin C. Williams, William Schwab, Harry M. Teter, Marie C. Flanigan, Alan J. Blatt, and Charles S. ReVelle. 2005. Access to trauma centers in the United States. *Journal of the American Medical Association* 293 (21): 2626–33.

Brewster, Karin L. 1994a. Neighborhood context and the transition to sexual activity among young black women. *Demography* 31 (4): 603–4.

Brewster, Karin L. 1994b. Race differences in sexual activity among adolescent women: The role of neighborhood characteristics. *American Sociological Review* 59 (3): 408–24.

Brick, J. Michael, and Pam Broene. 1997. Unit and item response, weighting, and imputation procedures in the 1995 National Household Education Survey (NHES:95). National Center for Education Statistics Working Paper No. 97-06, Washington, DC.

Brick, J. Michael, David Judkins, Jill Montaquila, David Morganstein, and Gary Shapiro. 2000. Evaluating secondary data sources for random digit dialing samples. In *JSM proceedings*, Survey Research Methods Section, 7–14. Alexandria, VA: American Statistical Association.

Brick, J. Michael, David Martin, Patricia Warren, and J. Wivagg. 2003. Increased efforts in RDD surveys. In *JSM proceedings*, Survey Research Methods Section, 26–31. Alexandria, VA: American Statistical Association.

Brick, J. Michael, Jill Montaquila, and Fritz Scheuren. 2002. Estimating residency rates for undetermined telephone numbers. *Public Opinion Quarterly* 66 (1): 18–39.

Brooks-Gunn, Jeanne, Greg J. Duncan, Pamela Klebanov, and Naomi Sealand. 1993. Do neighborhoods influence child and adolescent development? *American Journal of Sociology* 99 (2): 353–95.

Browning, Christopher R., Tama Leventhal, and Jeanne Brooks-Gunn. 2004. Neighborhood context and racial differences in early adolescent sexual activity. *Demography* 41 (4): 697–702.

Browning, Christopher R., and Matisa Olinger-Wilbon. 2003. Neighborhood structure, social organization, and number of short-term sexual partnerships. *Journal of Marriage and the Family* 65 (3): 730–45.

Bryk, Anthony S., and Stephen W. Raudenbush. 1988. Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education* 97 (1): 65–108.

Burden, Barry C. 2000. Voter turnout and the National Election Studies. *Political Analysis* 8 (4): 389–98.

Campanelli, Pamela, Patrick Sturgis, and Susan Purdon. 1997. *Can you hear me knocking? An investigation into the impact of interviewers on survey response rates*. London: Survey Methods Centre, Social and Community Planning Research, University of London.

Cantor, David, and Patricia Cunningham. 2002. Methods for obtaining high response rates in telephone surveys. In *Studies of welfare populations: Data collection and research issues*, eds. Michele Ver Ploeg, Robert A. Moffitt, and Constance F. Citro, 55–85. Washington, DC: National Academies Press.

Centers for Disease Control and Prevention. 2004. *Sexually transmitted disease surveillance 2004*. Atlanta, GA: Centers for Disease Control and Prevention. Available from http://www.cdc.gov/std/stats04/toc2004.htm (accessed 22 June 2012).

Charles, Camille Zubrinsky. 2003. The dynamics of racial residential segregation. *Annual Review of Sociology* 29:167–207.

Cohen, Cathy J., and Michael C. Dawson. 1993. Neighborhood poverty and African American politics. *American Political Science Review* 87 (2): 286–302.

Cohen, Deborah, Suzanne Spear, Richard Scribner, Patty Kissinger, Karen Mason, and John Wildgen. 2000. Broken windows and the risk of gonorrhea. *American Journal of Public Health* 90 (2): 230–36.

Committee on State Voter Registration Databases. 2008. *State voter registration databases*. Washington, DC: National Academies Press.

Committee on State Voter Registration Databases. 2009. *Improving state voter registration databases: Final report*. Washington, DC: National Academies Press.

Converse, Jean M., and Howard Schuman. 1974. *Conversations at random: Survey research as interviewers see it*. New York, NY: John Wiley.

Costa, Dora L., and Matthew E. Kahn. 2003. Civic engagement and community heterogeneity: An economist's perspective. *Perspectives on Politics* 1 (1): 103–11.

Cotter, David A., JoAnn DeFiore, Joan M. Hermsen, Brenda Marsteller Kowalewski, and Reeve Vanneman. 1998. The demand for female labor. *American Journal of Sociology* 103 (6): 1673–1712.

Couper, Mick P., and Robert M. Groves. 1996. Social environmental impacts on survey cooperation. *Quality & Quantity* 30 (2): 173–88.

Couper, Mick P., and Lars Lyberg. 2005. The use of paradata in survey research. Paper presented at the 55th Session of the International Statistical Institute, Sydney, Australia.

Couper, Mick P., Eleanor Singer, and Richard A. Kulka. 1998. Participation in the 1990 decennial census: Politics, privacy, pressures. *American Politics Quarterly* 26 (1): 59–80.

Covington, J., and R. B. Taylor. 1991. Fear of crime in urban residential neighborhoods: Implications of between- and within-neighborhood sources for current models. *Sociological Quarterly* 32 (2): 231–49.

Cox, Christine S. 2006. *Integrating data: Policy and legal issues*. Report prepared for the Federal Committee on Statistical Methodology, November, Washington, DC.

Crane, Jonathan. 1991. The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology* 96 (6): 1226–59.

Davern, Michael. 2006. Evaluating linked survey and administrative data for policy research. Paper presented to the Federal Committee on Statistical Methodology, November, Washington, DC.

Davis, James A., Tom W. Smith, and Peter V. Marsden. 2009. *General Social Survey cumulative codebook, 1972–2008*. Chicago, IL: NORC.

De Jong, Gordon F., and Michele Steinmetz. 2004. Receptivity attitudes and the occupational attainment of male and female immigrant workers. *Population Research and Policy Review* 23 (2): 91–116.

DiPrete, Thomas A., and Jerry D. Forristal. 1994. Multivariate models: Methods and substance. *Annual Review of Sociology* 20:331–57.

DiSogra, Charles, Mario Callegaro, and Erlina Hendarwan. 2009. Recruiting probability-based web panel members using an address-based sample frame: Results from a pilot study conducted by knowledge networks. Paper presented to the American Statistical Association, August, Washington, DC. Available from http://www.knowledgenetworks.com/ganp/docs/jsm2009/abs_jsm_2009_submitted.pdf (accessed 22 June 2012).

Dixon, Jeffrey C., and Michael S. Rosenbaum. 2004. Nice to know you? Testing contact, cultural, and group threat theories of anti-black and anti-Hispanic stereotypes. *Social Science Quarterly* 85 (2): 257–80.

Downey, Liam. 2006. Using geographic information systems to reconceptualize spatial relationships and ecological context. *American Journal of Sociology* 112 (2): 567–612.

D'Urso, Victoria T. 2003. Internet use and the duration of buying and selling in the residential housing market, economic incentives, and voting. PhD diss., Department of Economics, Massachusetts Institute of Technology, Cambridge, MA.

Ekholm, Anders, and Seppo Laaksonen. 1991. Weighting via response modeling in the Finnish Household Budget Survey. *Journal of Official Statistics* 7 (3): 325–37.

Fair, Martha E. 1996. Record linkage in an information age society. In *Proceedings of the Census Bureau's Conference and Technology Interchange*. Washington, DC: U.S. Census Bureau. Available from http://www.census.gov/prod/2/gen/96arc/xia-fair.pdf (accessed 22 June 2012).

FBI. 2004. *Uniform crime reporting program data: County-level detailed arrest and offense data, 2004*. Washington, DC: FBI.

Ford, Julie, and Andrew A. Beveridge. 2006. Neighborhood crime victimization, drug use, and drug sales. *Journal of Drug Issues* 36 (2): 393–416.

Fowler, Floyd J., Jr., Patricia M. Gallagher, Vickie L. Stringfellow, Alan M. Zaslavsky, Joseph W. Thompson, and Paul D. Cleary. 2002. Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. *Medical Care* 40 (3): 190–200.

Galea, Sandro, Jennifer Ahern, and David Vlahov. 2003. Contextual determinants of drug use risk behavior: A theoretic framework. *Review of Economics and Statistics* 80 (S3): 257–68.

Gatewood, George. 2001. *A monograph on confidentiality and privacy in the U.S. Census*. Washington, DC: U.S. Census Bureau.

Gelman, Andrew, and John B. Carlin. 2002. Poststratification and weighting adjustments. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 289–303. New York, NY: John Wiley.

Geronimus, Arline T., John Bound, and Lisa J. Neidert. 1996. On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. *Journal of the American Statistical Association* 91 (434): 529–37.

Gfroerer, Joseph, Judith Lessler, and Teresa Parsley. 1997. Studies of nonresponse and measurement error in the National Household Survey of Drug Abuse. In *The validity of self-reported drug use: Improving the accuracy of survey estimates*, eds. Lana Harrison and Arthur Hughes, 273–95. NIH Publication No. 97-4147, NIDA Research Monograph 167. Rockville, MD: National Institute on Drug Abuse.

Gibson, James L. 1995. The political freedom of African Americans: A contextual analysis of racial attitudes, political tolerance, and individual liberty. *Political Geography* 14 (6/7): 571–99.

Gilbert, Christopher P. 1991. Religion, neighborhood environments, and partisan behavior: A contextual analysis. *Political Geography Quarterly* 10 (2): 110–31.

Glaeser, Edward L., and Spender Glendon. 1998. Who owns guns? Criminals, victims, and the culture of violence. *American Economic Review* 88 (2): 458–62.

Goyder, John, Jean Lock, and Trish McNair. 1992. Urbanization effects on survey nonresponse: A test within and across cities. *Quality & Quantity* 26 (1): 39–48.

Groves, Robert M. 2005. Research synthesis: Nonresponse rates and nonresponse error in household surveys. Paper presented at the 16th International Workshop on Survey Nonresponse, August 24–26, Stockholm, Sweden. Available from http://www.nonresponse.org/uploadi/editor/1129103567Research%20Synthesis%20nonresponse%20rates.pdf (accessed 4 June 2012).

Groves, Robert M. 2006. Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly* 70 (5): 646–75.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in household interview surveys*. New York, NY: John Wiley.

Groves, Robert M., and Mick P. Couper. 2001. Designing surveys acknowledging nonresponse. In *Studies of welfare populations: Data collection and research issues*, eds. Michele Ver Ploeg, Robert A. Moffitt, and Constance F. Citro, 11–54. Washington, DC: National Academies Press.

Groves, Robert M., and Lars Lyberg. 2010. Total survey error: Past, present, and future. *Public Opinion Quarterly* 74 (5): 849–79.

Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. Leverage-saliency theory of survey participation: Description and an illustration. *Public Opinion Quarterly* 64 (3): 299–308.

Grubesic, Tony H., and Timothy C. Matisziw. 2006. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *International Journal of Health Geographics* 5 (58): 1–15.

Harter, Rachel, Stephanie Eckman, Ned English, and Colm O'Muircheartaigh. 2010. Applied sampling for large-scale, multi-stage area probability designs. In *Handbook of survey research*, 2nd ed., eds. Peter V. Marsden and James D. Wright. Bingley, UK: Emerald Group.

Harvey, Bart J., Annette L. Wilkins, Gillian A. Hawker, Elizabeth M. Badley, Peter C. Coyte, Richard H. Glazier, J. Ivan Williams, and James G. Wright. 2003. Using publicly available directories to trace survey nonresponders and calculate adjusted response rates. *American Journal of Epidemiology* 158:1007–11.

Hill, Mark E., and Ira Rosenwaike. 2001–2002. The Social Security Administration's Death Master File: The completeness of death reporting at older ages. *Social Security Bulletin* 64 (1): 45–51.

Holmes, Thomas J. 1999. Localization of industry and vertical disintegration. *Review of Economics and Statistics* 81 (2): 314–25.

House, James S., and Sharon Wolf. 1978. Effects of urban residence on interpersonal trust and helping behavior. *Journal of Personality and Social Psychology* 36 (9): 1029–43.

Jencks, Christopher. 1999. What is the optimal level of inequality? Unpublished paper, Kennedy School of Government, Harvard University, Cambridge, MA.

Jencks, Christopher, and Susan E. Mayer. 1990. The social consequences of growing up in a poor neighborhood. In *Inner-city poverty in the United States*, ed. Laurence E. Lynn, 111–86. Washington, DC: National Academies Press.

Jenkins, Stephen P., Peter Lynn, Annette Jäckle, and Emanuela Sala. 2005. Linking household survey and administrative record data: What should the matching variables be? Discussion Paper 489, DIW Berlin, Institute for Economic Research, Berlin. Available at http://www.diw.de/documents/publikationen/73/diw_01.c.43285.de/dp489.pdf (accessed 8 June 2012).

Johnson, Timothy, and Young Ik Cho. 2004. Understanding nonresponse mechanisms in telephone surveys. In *JSM proceedings*, Survey Research Methods Section, 4952–59. Alexandria, VA: American Statistical Association.

Johnson, Timothy P., Young Ik Cho, Richard T. Campbell, and Allyson L. Holbrook. 2006. Using community-level correlates to evaluate nonresponse effects in a telephone survey. *Public Opinion Quarterly* 70 (5): 704–19.

Johnston, S. Claiborne, Charles B. Wilson, Van V. Halbach, Randall T. Higashida, Christopher F. Dowd, Michael W. McDermott, Carol B. Applebury, Thomas L. Farley, and Daryl R. Gress. 2000. Endovascular and surgical treatment of unruptured cerebral aneurysms: Comparison of risks. *Annals of Neurology* 48 (1): 11–19.

Jones, Dale E. 2002. *Religious congregations and membership in the United States, 2000*. Nashville, TN: Glenmary Research Center.

Kalsbeek, William D., Juan Yang, and Robert P. Agans. 2002. Predictors of nonresponse in a longitudinal survey of adolescents. In *JSM proceedings*, Survey Research Methods Section, 1740–45. Alexandria, VA: American Statistical Association.

Kalton, Graham, and Daniel Kasprzyk. 1986. The treatment of missing survey data. *Survey Methodology* 12 (1): 1–16.

Kawachi, Ichiro, Bruce P. Kennedy, and Kimberly Lochner. 1997. Long live community: Social capital and public health. *The American Prospect* 35 (spring): 56–59.

Kawachi, Ichiro, Bruce P. Kennedy, Kimberly Lochner, and Deborah Prothrow-Stith. 1997. Social capital, income inequality, and mortality. *American Journal of Public Health* 87 (9): 1491–98.

Kennedy, Courtney, Scott Keeter, and Michael Dimock. 2008. A "brute force" estimation of the residency rate for undetermined telephone numbers in an RDD survey. *Public Opinion Quarterly* 72 (1): 28–39.

Kennickell, Arthur B. 1997. Analysis of nonresponse effects in the 1995 Survey of Consumer Finances. In *JSM proceedings*, Survey Research Methods Section, 377–83. Alexandria, VA: American Statistical Association.

Kennickell, Arthur B. 2005. Darkness made visible: Field management and nonresponse in the 2004 SCF. In *JSM proceedings*, Survey Research Methods Section, 3187–94. Alexandria, VA: American Statistical Association.

Key, Wanda L., and Stephanie Miracle. 2009. What are "bad" addresses and what do you do with them—exactly? Paper presented to the National Postal Forum, May, Washington, DC.

Kim, Jibum, Tom W. Smith, and John Sokolowski. 2006. Ecological correlates of cooperation rate in the 2002 and 2004 General Social Survey. Paper presented to the American Association for Public Opinion Research, May, Montreal, Canada.

Kojetin, Brian A. 1994. Characteristics of nonrespondents to the Current Population Survey (CPS) and Consumer Expenditure Interview Survey (CEIS). In *JSM proceedings*, Survey Research Methods Section, 966–71. Alexandria, VA: American Statistical Association.

Krieger, Nancy, Jarvis T. Chen, Pamela D. Waterman, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson. 2002a. Geocoding and monitoring of U.S. socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? *American Journal of Epidemiology* 156 (5): 471–82.

Krieger, Nancy, Jarvis T. Chen, Pamela D. Waterman, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson. 2002b. Zip code caveat: Bias due to spatiotemporal mismatches between zip codes and U.S. Census–defined geographic areas. *American Journal of Public Health* 92 (7): 1100–1102.

Latkin, Carl A., and Aaron D. Curry. 2003. Stressful neighborhoods and depression: A prospective study of the impact of neighborhood disorder. *Journal of Health and Social Behavior* 44 (1): 34–44.

Lee, Barrett A., R. Sal Oropesa, and James W. Kanan. 1994. Neighborhood context and residential mobility. *Demography* 31 (2): 249–70.

Lepkowski, James M., and Mick. P. Couper. 2002. Nonresponse in the second wave of longitudinal household surveys. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 259–74. New York, NY: John Wiley.

Lessler, Judith T., and William D. Kalsbeek. 1992. *Nonsampling error in surveys*. New York, NY: John Wiley.

Loosveldt, Geert, and Ann Carton. 2002. Utilitarian individualism and panel nonresponse. *International Journal of Public Opinion Research* 14 (4): 428–38.

Luttmer, Erzo F. P. 1998. Understanding income redistribution: The role of interpersonal preference, information, and mechanism design. PhD diss., Department of Economics, Harvard University, Cambridge, MA.

Lynn, Peter. 2002. PEDAKSI: Methodology for collecting data about survey nonrespondents. *Quality & Quantity* 37 (3): 239–61.

Lynn, Peter, Paul Clark, Jean Martin, and Patrick Sturgess. 2002. The effect of extended interviewer efforts on nonresponse bias. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 135–49. New York, NY: John Wiley.

Marcus, Pamela M., Erik J. Bergstralh, Mark H. Zweig, Ann Harris, Kenneth P. Offord, and Robert S. Fontana. 2006. Extended lung cancer incidence follow-up in the Mayo Lung Project and overdiagnosis. *Journal of the National Cancer Institute* 98 (11): 748–56.

Marker, David A., David R. Judkins, and Marianne Winglee. 2002. Large-scale imputation for complex surveys. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 329–42. New York, NY: John Wiley.

Massey, Douglas S., and Mitchell L. Eggers. 1990. The ecology of inequality: Minorities and the concentration of poverty. *American Journal of Sociology* 95 (5): 1153–88.

Massey, Douglas S., Andrew B. Gross, and Kumiko Shibuya. 1994. Migration, segregation, and the geographic concentration of poverty. *American Sociological Review* 59 (3): 425–45.

McLeod, Jane, and Kevan Edwards. 1995. Contextual determinants of children's responses to poverty. *Social Forces* 73 (4): 1487–516.

Mellor, Jennifer M., and Jeffrey Milyo. 2005. State social capital and individual health status. *Journal of Health Politics, Policy, and Law* 30 (6): 1101–30.

Meng, Xiao-Li. 2002. A congenial overview and investigation of multiple imputation inferences under uncongeniality. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 343–56. New York, NY: John Wiley.

Merkle, Daniel, Gary Langer, Jon Cohen, Linda Piekarski, Robert Benford, and David Lambert. 2009. Unintended consequences: The cost of purging business numbers in RDD surveys. *Public Opinion Quarterly* 73 (3): 484–96.

Minato, Hiroaki, and Lidan Luo. 2004. Towards a better estimate of the working residential number (WRN) rate among the undetermined: An application of survival analysis. In *JSM proceedings*, Survey Research Methods Section, 3933–39. Alexandria, VA: American Statistical Association.

Montaquila, Jill M., and Michael J. Brick. 1997. Unit and item response rates, weighting, and imputation procedures in the 1996 National Household Education Survey. National Center for Education Statistics Working Paper 97-40, Washington, DC.

Moore, Danna L., and John Tarnai. 2002. Evaluating nonresponse error in mail surveys. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 197–212. New York, NY: John Wiley.

Moore, Laura M. 1999. A multi-level analysis of attitudes regarding women in politics: Why is the United States South different? PhD diss., Department of Sociology, University of Maryland, College Park.

Moore, Laura M., and Reeve Vanneman. 2003. Context matters: Effects of the proportion of fundamentalists on gender attitudes. *Social Forces* 82 (1): 115–39.

Murray, Mary Cay, Jessica Cardoni, Erin Foster, Chris Becker, Paul Buckley, and Marcie Cynamon. 2003. Impact of changes in the telephone environment on RDD telephone surveys. Paper presented to the American Association for Public Opinion Research, May, Nashville, TN.

Nolin, Mary Jo, Jill Montaquila, Patricia Nicchitta, Kwang Kim, Brian Kleiner, Jean Lennon, Christopher Chapman, Sean Creighton, and Stacey Bielick. 2000. *National Household Education Survey of 1999: Methodology report*. Washington, DC: National Center for Education Statistics.

Obenski, Sally. 2006. Models and applications of administrative records research. Paper presented at the Workshop on Data Linkages to Improve Health Outcomes, September 19. Washington, DC: National Committee on Vital and Health Statistics, Subcommittee on Population, U.S. Department of Health and Human Services.

O'Hare, Barbara C., Sonja Ziniel, and Robert M. Groves. 2005. Modeling components of response propensity in centralized telephone surveys. Paper presented to the American Association for Public Opinion Research, May, Miami Beach, FL.

O'Muircheartaigh, Colm. 2003. *There and back again: Demographic survey sampling in the 21st century*. Washington, DC: Federal Committee on Statistical Methodology. Available from www.fcsm.gov/events/papers2003.html (accessed 12 June 2012).

Oreopoulos, Philip. 2003. The long-run consequences of living in a poor neighborhood. *Quarterly Journal of Economics* 118 (4): 1533–75.

Pappas, Gregory, Wilbur C. Hadden, Lola Jean Kozak, and Gail F. Fisher. 1997. Potentially avoidable hospitalizations: Inequalities in rates between U.S. socioeconomic groups. *American Journal of Public Health* 87 (5): 811–16.

Peel, Jennifer L., Kristi Busico Metzger, Mitchel Klein, W. Dana Flanders, James A. Mulholland, and Paige E. Tolber. 2005. Ambient air pollution and respiratory emergency department visits. *Epidemiology* 165 (6): 625–33.

Peeples, Faith, and Rolf Loeber. 1994. Do individual factors and neighborhood context explain ethnic differences in juvenile delinquency? *Journal of Quantitative Criminology* 10 (2): 141–57.

Powell, Lisa M., Sandy Slater, Donka Mirtcheva, Yanjun Bao, and Frank J. Chaloupka. 2006. Food store availability and neighborhood characteristics in the United States. *Preventive Medicine* 44 (3): 189–95.

Raghunathan, Trivellore, and John Van Hoewyk. 2005. Disclosure risk assessment for survey microdata. Unpublished paper, Institute for Social Research, University of Michigan, Ann Arbor. Available from http://www-personal.umich.edu/~teraghu/Raghunathan-Van%20Hoewyk-2.pdf (accessed 12 June 2012).

Rahn, Wendy M., and Thomas J. Rudolph. 2005. A tale of political trust in American cities. *Public Opinion Quarterly* 69 (4): 530–60.

Raudenbush, Stephen, and Anthony S. Bryk. 2002. *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Regnerus, Mark D. 2003. Moral communities and adolescent delinquency: Religious contexts and community social control. *Sociological Quarterly* 64 (4): 523–54.

Rosenfeld, Richard, Timothy M. Bray, and Arlen Egley. 1999. Facilitating violence: A comparison of gang-motivated, gang-affiliated, and nongang youth homicides. *Journal of Quantitative Criminology* 15 (4): 495–516.

Rosenfeld, Richard, and Steven F. Messner. 1998. Beyond the criminal justice system: Anomie, institutional vitality, and crime in the United States. Paper presented to the American Sociological Association, August, San Francisco, CA.

Rosenfeld, Richard, Steven F. Messner, and Eric P. Baumer. 2001. Social capital and homicide. *Social Forces* 80 (1): 283–309.

Safir, Adam, Rebecca Steinbach, Timothy Triplett, and Kevin Wang. 2002. Effects on survey estimates from reducing nonresponse. In *JSM proceedings*, Survey Research Methods Section, 3024–29. Alexandria, VA: American Statistical Association.

Salvo, Joseph, and Peter A. Lobo. 2003. The effect of neighborhood characteristics on nonresponse in the Bronx test site of the American Community Survey. In *JSM proceedings*, Survey Research Methods Section, 3624–35. Alexandria, VA: American Statistical Association.

Sampson, Robert J., Jeffrey D. Morenoff, and Felton Earls. 1999. Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review* 64 (5): 633–60.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* 277 (5328): 918–24.

Scheuren, Fritz. 2000. Macro and micro paradata for survey assessment. Unpublished paper, United Nations, Economic Commission for Europe, New York, NY. Available from http://www.unece.org/fileadmin/DAM/stats/documents/2000/11/metis/crp.10.e.pdf (accessed 22 June 2012).

Shaw, Clifford R., and Henry D. McKay. 1969. *Juvenile delinquency and urban areas*. Chicago, IL: University of Chicago Press.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. Who overreports voting? *American Political Science Review* 80 (2): 613–24.

Smith, Tom W. 1983. The hidden 25 percent: An analysis of nonresponse on the General Social Survey. *Public Opinion Quarterly* 47 (3): 386–404.

Smith, Tom W. 1984. Estimating nonresponse bias with temporary refusals. *Sociological Perspectives* 27 (4): 473–89.

Smith, Tom W. 1994. *Anti-Semitism in contemporary America*. New York, NY: American Jewish Committee.

Smith, Tom W. 1999. *A methodological analysis of HIV risk behavior from the 1988–1998 General Social Survey*. GSS Methodological Report No. 92. Chicago, IL: NORC.

Smith, Tom W. 2002. Developing nonresponse standards. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 27–40. New York, NY: John Wiley.

Smith, Tom W. 2005. Total survey error. In *Encyclopedia of social measurement*, ed. Kimberly Kempf-Leonard, 857–62. New York, NY: Academic Press.

Smith, Tom W. 2006. Estimating the status of cases with unknown eligibility in telephone surveys. Paper presented to the Second International Conference on Telephone Survey Methodology, January, Miami, FL.

Smith, Tom W. 2007. Survey non-response in cross-national perspective: The 2005 ISSP nonresponse survey. *Survey Research Methods* 1 (1): 45–54.

Smith, Tom W., and John Sokolowski. 2011. Using audio-visuals in surveys. In *The handbook of emergent technologies in social research*, ed. Sharlene Hesse-Biber, 437–52. Oxford: Oxford University Press.

Snedker, Karen A., Jerald R. Herting, and Emily C. Walton. 2009. Contextual effects and adolescent substance use: Exploring the role of neighborhoods. *Social Science Quarterly* 90 (5): 1272–97.

South, Scott, and Eric P. Baumer. 2001. Community effects on the resolution of adolescent premarital pregnancy. *Journal of Family Issues* 22 (8): 1025–42.

South, Scott, Eric P. Baumer, and Amy Lutz. 2003. Interpreting community effects on youth educational attainment. *Youth and Society* 35 (1): 3–36.

Steeh, Charlotte G., Nicole Kirgis, Brian Cannon, and Jeff DeWitt. 2001. Are they really as bad as they seem? Nonresponse rates at the end of the twentieth century. *Journal of Official Statistics* 17 (2): 227–47.

Stoop, Ineke A. L. 2004. Survey nonrespondents. *Field Methods* 16 (1): 23–54.

Stoop, Ineke A. L. 2005. *The hunt for the last respondent: Nonresponse in sample surveys*. The Hague, the Netherlands: Social and Cultural Planning Office.

Taylor, Marylee C. 1998. How white attitudes vary with the racial composition of local populations: Numbers count. *American Sociological Review* 63 (4): 512–35.

Taylor, Marylee C. 2002. Fraternal deprivation, collective threat, and racial resentment: Perspectives on white racism. In *Relative deprivation: Specification, development, and integration*, eds. Ian Walker and Heather J. Smith, 13–43. New York, NY: Cambridge University Press.

Thomas, Avis J., Lynn E. Eberly, George Davey Smith, and James D. Neaton. 2006. ZIP-code-based versus tract-based income measures as long-term risk-adjusted mortality predictors. *American Journal of Epidemiology* 164 (6): 586–90.

Tolbert, Charles M., Thomas A. Lyson, and Michael D. Irwin. 1998. Local capitalism, civic engagement, and socioeconomic well-being. *Social Forces* 77 (2): 401–27.

Traub, Jane, Kathy Pilhuj, and Daniel T. Mallett. 2005. 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, May, Miami Beach, FL.

Turrell, Gavin, Carla Patterson, Brian Oldenburg, Trish Gould, and Marie-Andree Roy. 2003. 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 (2): 181–89.

U.S. Department of Housing and Urban Development. 1998. *A picture of subsidized households—1998*. Washington, DC: U.S. Department of Housing and Urban Development. Available from http://www.huduser.org/datasets/assthsg/statedata98/ (accessed 4 June 2012).

van Goor, Henk, Folbert Jansma, and Rene Veenstra. 2005. Differences in undercoverage and nonresponse between city neighborhoods in a telephone survey. *Psychological Reports* 96 (3): 867–78.

Voogt, Robert J. J., and Hetty van Kempen. 2002. Nonresponse bias and stimulus effects in the Dutch National Election Study. *Quality & Quantity* 36 (4): 325–45.

Weakliem, David L. 1997. Race versus class? Racial composition and class voting 1936–1992. *Social Forces* 75 (3): 939–56.

Weakliem, David L., and Robert Biggert. 1999. Region and political opinion in the contemporary United States. *Social Forces* 77 (3): 863–86.

Williams, Jason, Alisa McWilliams, Tina Mainieri, Peter J. Pecora, and Karin La Belle. 2006. Enhancing the validity of foster care follow-up studies through multiple alumni location strategies. *Child Welfare* 85 (3): 499–521.

Wirth, Louis. 1938. Urbanism as a way of life. *American Journal of Sociology* 44 (1): 3–24.

Zanutto, Elaine, and Alan Zaslavsky. 2002. Using administrative records to impute nonresponse. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 403–17. New York, NY: John Wiley.