# Table of Contents

- Executive Summary .................................................. 3
- Introduction and Background ........................................ 7
- Summary of Studies and Key Findings ............................. 13
  - Sample Frames .................................................... 13
  - Sampling Strategy and Size ....................................... 22
  - Data Collection Mode and Procedures ......................... 23
  - Finding Hard-to-Reach Populations ............................. 33
  - Issues for Surveys Involving Linguistically and Culturally Diverse Populations .................................................. 37
- Summary and Recommendations ..................................... 39
- References .................................................................... 45
- Appendix A: Summary of Surveys .................................. 53
Executive Summary

Introduction and Background

The First 5 Los Angeles (F5LA) Strategic Plan for FY 2009-2015 represents a new commitment by the F5LA Commission to direct funding to specific communities in Los Angeles County, called "Best Start Communities." To monitor progress toward the goals of the Strategic Plan, F5LA will develop a Community Change Dashboard. The dashboard will function as a tool that each Best Start community can use to identify and monitor risks, as well as mobilize resources from within and without. In the absence of existing databases that provide timely sources of information for the dashboard, F5LA initiated the development of the F5LA Family Survey. The Family Survey is intended to provide a snapshot of each Best Start community at one point in time and will be the main source of information for the Community Change Dashboard.

F5LA hired Harder+Company Community Research to assist in developing a plan for the design and implementation of the Family Survey in the 14 Best Start communities. The plan should be based on cost-effective but methodologically rigorous strategies drawn from best practices and lessons learned in the survey literature. The following background paper reviews the literature on sample designs and administration methods currently being used in surveys of representative samples of children and families. This paper will examine the issues and discuss the implications for the design of the F5LA Family Survey.

Literature Review Methods

Harder+Company used the following criteria to select the surveys to review: Large-scale surveys that are probability-based; include families with children, preferably from birth through five but at least families with children under 18 years of age; conducted in urban and rural settings where estimates are available for small geographic units, such as census tracts; and incorporate families from low income communities. Twelve studies met the criteria and were included in this literature review:

- American Community Survey
- 2003 California Health Interview Survey (CHIS)
- 2005 California Health Interview Survey (CHIS)
- 2007 California Health Interview Survey (CHIS)
- 2005 Los Angeles County Health Survey
- 2007 Los Angeles County Health Survey
- 2009 District of Columbia Health Insurance Survey
- 2008 Massachusetts Health Insurance Survey
- Los Angeles County WIC Survey
- Los Angeles Mommy and Baby (LAMB) Survey
- Los Angeles Family and Neighborhood Survey (LA-FANS)
- Making Connections Survey
Summary of Studies and Key Findings

The findings are grouped according to three key steps involved in designing a household survey: a) identifying the sample frame, b) determining sample strategy and size, and c) selecting data collection mode and procedures. Within each of these steps, a variety of issues central to the design of the F5LA Family Survey will be discussed, including cultural and linguistic diversity, use of indigenous interviewers, migration patterns, computer-assisted interviewing, and locating hard-to-reach populations.

Sample Frame

- Several sample frames were used most frequently in the studies reviewed: Random Digit Dialing (RDD), Address-Based Sampling (ABS), dual-frame RDD plus ABS, and other forms of constructed lists in a multi-stage area sample design.
- The highest response rates tended to come from studies that used either RDD or ABS, but not both. Given the ways in which response rates were calculated, caution is needed when interpreting these differences. But the two studies that used a combined RDD and ABS frame had much lower response rates than the studies that used one or the other, with considerably higher costs.
- Many of the most recent, well-done RDD surveys, such as CHIS 2007, include a cell phone sample frame that appears to mitigate or eliminate an important source of coverage bias. Although controversial, it may be possible to adjust for any remaining bias through a weighting procedure that better aligns the sample with the population.
- If an ABS approach is employed, the need to continuously update and verify the USPS Delivery Sequence File of addresses must be considered. Still, the ABS approach appears to be less costly than a multi-stage area sample, such as that employed by LA-FANS. This approach involved considerable time and expense constructing the sample frame of households within selected census tracts.
- For an ABS approach, continuously updating address lists, as the American Community Survey (ACS) does, will be necessary to draw samples for the twice per year surveys, as per the original plan for the Family Survey. The need for timeliness in data collection may conflict with the time and expense required to maintain updated lists.
- If using an ABS approach, another consideration is whether to include group quarters and, if so, how they are added to the frame and updated so as to reduce coverage bias.
- To address the issue of “hidden” or hard-to-reach populations in the sample frame, the respondent-driven sampling (RDS) approach appears promising. But further study is needed to ensure that the populations defined as hard-to-reach meet the criteria for using the RDS approach.
- Finally, regardless of the survey design selected, it will be important to revisit the notion that the two yearly samples drawn from the sample frame should be separate and independent. While it may be intuitively obvious that each sample would have greater precision if drawn independently, this may not allow for the detection of changes in the communities over time. Consideration should be given to an overlapping sample design in which each sample selected from the sample frame is not independent.

Sampling Strategy and Size

- Most of the ABS studies reviewed used various forms of a multi-stage area sampling approach, such as sampling census tracts, then households within census tracts and finally individuals within households. The largest macro unit of the sample (e.g. census
tract) typically employed the probability proportional to size (PPS) strategy in order to take into account the effects of different population densities across census tracts on the probabilities of selection.

- The Family Survey will need to obtain a sample large enough to provide for good precision of the population estimates, particularly at the largest geographic level, such as the census tract or the Best Start community or neighborhood, however that is defined.
- There are considerably more marginal costs involved in adding macro units (e.g. census tracts) to the sample than there are in adding smaller analytic units (such as households).
- The study that spent more time and effort constructing its own list-based frame needed to sample fewer households than other surveys. The two dual-frame RDD/ABS studies also required somewhat fewer individuals or households.
- Many of the studies oversampled for specific sub-populations, such as levels of poverty within the census tracts or ethnic and/or language groups, including “linguistically isolated” households.
- Migration of residents into and out of Best Start communities is an important factor to consider, but currently insufficient information exists to guide sample design.

Data Collection Mode and Procedures

- The literature suggests that using indigenous interviewers (who come from the same communities as those in the survey, with similar backgrounds and characteristics to respondents), offers no strong benefits—and there may be some disadvantages. It is more important to employ interviewers who are well-trained and experienced.
- Computer-assisted interview technologies offer the best methods for maximizing response rates and data quality. For RDD designs, a seasoned telephone research center using up-to-date computer-assisted telephone interview (CATI) technology should be employed. For ABS designs, in-person interviews using computer-assisted personal interview (CAPI), combined with CATI for telephone interviews, is recommended.
- Many of the studies reviewed used mixed methods of data collection. The primary data collection method typically involved telephone or mail, with more intensive methods reserved for sampled respondents who did not respond to the initial inquiry.
- To improve precision and data quality, the data collection mode must be considered within the larger context of the survey sample frame and sampling plan. Mixed methods may be more costly but have the potential for improving response rates, particularly for key population sub-groups such as cultural and language minorities.
- However, in the studies reviewed, the use of different modes did not always improve the response rates. Caution must be used when interpreting response rates across surveys when their uses and definitions may not necessarily be consistent.
- Many of the studies reviewed did not use any incentives. The pilot study for one national RDD survey, National Household Education Survey, found that higher incentives led to higher response rates.

Issues for Surveys Involving Linguistically and Culturally Diverse Populations

- Most of the studies reviewed included languages beyond English and Spanish, including Mandarin, Cantonese, Vietnamese, Korean, Russian, and in one study, Armenian. Contrary to this trend, LA-FANS only used English and Spanish.
Excluding linguistic minorities in surveys (by not providing translations for those languages) may lead to non-response bias. It is important to consider how interview language may contribute to measurement error differentially by language.

Important cultural differences may exist during the recruitment stages for members of minority groups. Researchers should pay attention to the separate stages of respondent recruitment and improve strategies to locate members of racial and ethnic minority groups during sample recruitment.

**Recommendations**

Overall, two main designs are recommended, each of which attempt to balance the inherent tradeoffs between cost and methodological rigor. In addition, a third option is presented as a more expensive, but potentially more rigorous, variation on the higher-cost option. The two main designs (including the variation) are as follows:

- **Lower cost approach:** Telephone-based RDD with cell phone coverage.

- **Higher cost approach:** ABS approach using the USPS Delivery Sequence File purchased through a vendor, with sampling proportional to size within specific census tracts or communities. An even higher-cost option is a multistage area sample using a constructed list of households within selected areas (e.g. census tracts) based on existing addresses combined with local field operations, which should improve coverage and response rates.

Both of the two main recommended designs have methodological advantages and disadvantages:

- **RDD:** High response rates in comparison to other methods, relatively low costs, good coverage especially with inclusion of a cell phone sample, but low sample efficiency.

- **ABS:** High responses rates when lists are updated frequently, larger costs due to the efforts needed for updating and keeping valid address lists, excellent coverage, and high sample efficiency with multiple mode data collection strategies.
Introduction and Background

Best Start and Best Start Family Survey

In June 2009, the First 5 Los Angeles (F5LA) Board of Commissioners approved its FY 2009-2015 Strategic Plan.¹ This strategic plan represents a new commitment by the Commission to direct funding to specific communities in Los Angeles County, called “Best Start Communities.” F5LA has identified 14 Best Start communities throughout Los Angeles County. Through the Best Start framework, F5LA hopes to create environments in Los Angeles’ diverse communities where young children are born healthy and raised in supportive environments that allow them to grow up eager to learn with the ability to reach their full potential.

All of F5LA’s investments, particularly those in the Best Start communities, will be focused on changing outcomes in four priority areas representing F5LA’s long-term goals of ensuring that:

- Babies are born healthy
- Children maintain a healthy weight
- Children are safe from abuse and neglect
- Children are ready for kindergarten

The strategic plan outlined a theory of change linking its long-term goals to a number of intermediate and longer-term outcomes that will be tracked via the Family Survey.

The Best Start Communities investment represented a shift in F5LA’s grant making from primarily funding programs based on specific initiatives, to a community-based approach known as “place-based.” The place-based approach enables F5LA to focus its human and financial resources in entire communities to improve the lives of children and families. This shift to a place-based approach is driven by the growing literature validating the ecological model of children’s development. A place-based strategy operationalizes the ecological model by proposing that:

- Whole communities need to be healthy in order for children and families living in those communities to be healthy;
- The larger systems that support the health, safety, and early learning of children need to be integrated; and
- Existing infrastructures and relationships can be leveraged by F5LA in certain communities.

Thus, F5LA aims to strengthen the family as well as build the capacity of the communities. F5LA will also invest in county-wide efforts that will support children across the county and complement the work being done within each Best Start community. Through both the place-based strategy and the county-wide approach, F5LA seeks to collectively improve the larger community as well as directly improve the lives of the children and families living in these communities.

As part of the implementation of the strategic plan, F5LA developed an Accountability and Learning Framework to monitor its own progress. The purpose of this framework is to provide

---

timely information to F5LA and the Best Start communities through a regularly updated set of indicators, the Community Change Dashboard.

The Community Change Dashboard will allow F5LA to do the following:

- Tell a coherent story of its work and impact;
- Promote the creation of a common set of indicators and outcomes;
- Conduct a systematic self-assessment of its performance;
- Promote dialog between F5LA and its stakeholders through timely and regular feedback about progress in reaching desired outcomes;
- Assist in selecting realistic and meaningful performance targets; and
- Incorporate learning from the field of early child development.

Community partners and other stakeholders in each of these communities will be involved in assessing local needs and developing solutions and programs to meet these needs. The Community Change Dashboard will function as a tool that each Best Start community can use to identify and monitor risks and mobilize resources from within and without to help communities become stronger and healthier. This, in turn, will help families to raise their young children by providing a safe and nurturing environment. Ultimately, the Community Change Dashboard has the potential to help communities become stronger by allowing leaders to monitor their efforts to address key community challenges.

Ideally, data for the dashboard would come from existing sources including Healthy City, Los Angeles County Healthy Survey (LACHS), Los Angeles County Women Infants Children (WIC) Survey, California Health Interview Survey (CHIS) and data from organizations with which F5LA collaborates, such as the Department of Child and Family Services. However, these data sources have several limitations, as follows:

1. The surveys are either not conducted regularly or are conducted every few years, such as LACHS and CHIS, which are done every two years. The frequency of these studies is insufficient to provide timely feedback to F5LA and to the Best Start communities to inform their policy planning and decision making around programs and emergent issues.
2. These surveys often do not include the hard-to-reach populations that F5LA is specifically interested in identifying within Best Start communities, such as homeless families, multiple families living in one housing unit, families who do not speak English or Spanish or are “linguistically isolated,” and families that are highly transient.
3. Existing survey data from these studies do not overlap perfectly with the geographic boundaries that define the 14 selected Best Start communities. Many of the surveys report data at the Service Planning Area (SPA) and/or Health District (HD) levels, rather than the smaller geographical units such as blocks and sub-blocks within these larger-order units. F5LA would like to focus specifically on the Best Start communities and monitor changes in them, but the existing data cannot be compiled in such a way as to precisely overlap with these relatively smaller areas.

To overcome the limitations of existing survey data, F5LA has initiated the design of the F5LA Family Survey, which will examine semi-annual changes in the Best Start communities related to the four priority measures mentioned earlier. The Family Survey seeks to address two core research questions:

- How do the families within each Best Start community change over time?
Are there any trends in risk and protective factors within and/or across the communities?

It is important to note that the Family Survey is not intended to evaluate the impact of the interventions in Best Start communities. For this, F5LA, in consultation with the F5LA Research Advisory Committee (RAC), will implement other evaluation activities, including a longitudinal panel study of families from the Best Start communities to compare with a panel of families from matched non-Best Start communities. Additionally, F5LA will employ other place-based and strategic research, including evaluations of some F5LA key strategies, as well as data collection within each of the F5LA Best Start communities to monitor progress towards change.

F5LA has already done considerable work on developing an instrument for the Family Survey. The selection of core items for the Family Survey was guided by a model from the child and family literature that links risk and protective factors to F5LA’s four priority outcomes. To identify and select the core items for the Family Survey, F5LA conducted a literature review and vetted the potential items with five experts in the field of child and family research. Following this work, F5LA reviewed surveys of families conducted in Los Angeles County (e.g. Los Angeles Health Survey) to identify indicators that have been pilot tested in the local context. As much as possible, the specific items came from existing surveys, so that Family Survey data can be benchmarked against the larger geographic area, such as the county. Through this process, the following eleven core items were selected for the Family Survey:

- Parent’s rating of child’s physical activity level
- Exclusive breastfeeding 0-3 months
- Fast food and soda consumption (parent and child)
- Parent-child daily reading
- Parent-child interactions (storytelling, playing, etc.)
- Participation in any preschool and childcare outside of the house
- Availability of formal social supports (e.g., medical home, childcare)
- Availability of informal social supports
- Parent stress (stress levels and coping)
- Parental depression
- Health insurance utilization

The Family Survey is intended to provide a snapshot of each Best Start community at one point in time and will be the main source of information for the Community Change Dashboard. Each semi-annual data collection will draw a separate and unrelated random sample representative of each Best Start community at that time.

With the survey instrument already well-developed, F5LA turned its attention to the survey design. In March 2011, F5LA hired Harder+Company Community Research to assist in developing a plan for the design of the Family Survey in the 14 Best Start communities.

**Family Survey Design**

The survey design will be based on cost-effective but methodologically rigorous strategies drawn from best practices and lessons learned in the survey literature. There are several key features of the proposed Family Survey design. The design should:
Include a probability-based sampling plan with estimated sample sizes in each of the 14 Best Start communities to ensure that it is representative of children 0-5 and their families from diverse cultural backgrounds living in these communities;

Identify the sample frame, data sources for development of a frame, and the statistical sample methods;

Show expected statistical power with assumptions about non-response rates, particularly among hard-to-reach populations living in these communities, and potential methods for finding these populations; and

Be reproducible at least every six months and take into account changes in community composition over time.

One of the first tasks in developing the Family Survey design is to conduct a thorough literature review and prepare a background paper that will identify the best methods for sample design and survey administration. Recommendations for the design of the F5LA Family Survey will be based on the findings from the literature review. These recommendations will lead to a fully-specified draft study design and analysis plan that will be presented to the F5LA RAC for its review, along with the background paper. The feedback received from the RAC and F5LA Research and Evaluation staff will be incorporated into a final Family Survey design.

**Background Paper**

This background paper reviews the literature on survey sample designs and administration methods currently being used in surveys of representative samples of children and families that provide for probability sampling of households in small geographical units (such as census tracks, blocks or sub-blocks). The literature review examines prior surveys of disadvantaged families and families from diverse cultural backgrounds living in urban or rural areas similar to the Best Start communities. Another criterion for these surveys was that the sample should include families with children from birth through five, although, as a secondary criterion, surveys of families with children under the age of 18 years were also reviewed. In addition to the review of literature, the Harder+Company team included a survey design consultant who provided additional information throughout the document (see references to personal communications).

The main research questions that this paper addresses are as follows:

1. What is an effective sampling design to reach low income, hard-to-reach, and culturally diverse families living in both rural and urban areas that maximizes community-level representativeness?

2. What data collection procedures maximize cooperation and response rates?

3. What are some of the tradeoffs between methodological rigor and cost? Can the additional precision gained from more costly designs be justified?

The current literature on survey sampling and methodology is vast. Given the research questions posed by this paper, a well-defined set of criteria was selected and used to review the existing literature. Surveys that meet the following criteria were included in this literature review:

- Surveys that are probability-based, i.e., with a known sample frame and methods of selection based on the frame;

- Surveys that include families with children. The primary focus was on surveys that include families with children birth through five, but since many surveys did not meet this criterion, a secondary criterion consisted of families with children under 18 years of age;
- Large scale surveys conducted in urban and rural settings where estimates are available for geographical units at the county level or smaller; and
- Surveys that includes families from low income communities.

Many large and important surveys that were reviewed did not meet the above criteria and were not considered in this review. Table 1 below describes the excluded surveys and indicates which criteria they did not meet.

<table>
<thead>
<tr>
<th>Survey</th>
<th>Probability Based</th>
<th>FC(^a)</th>
<th>RC(^b)</th>
<th>LI(^c)</th>
<th>Small Area Estimates are Available(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral Risk Factor Surveillance System</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>California Healthy Kids Survey</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Centers for Disease Control and Prevention (CDC) Pregnancy Risk Assessment Monitoring System (PRAMS)</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Early Childhood Longitudinal Study</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Fragile Family and Child Wellbeing Study</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Joint Canada and US Survey of Health</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>National Longitudinal Survey (NLSY79)</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>National Survey of American Families</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>National Survey of Children with Special Health Care Needs</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>National Survey of Children’s Health</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>National Youth Risk Behavior Survey</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

\(^a\) FC = Includes families with children under 18 years of age
\(^b\) RC = Includes rural communities
\(^c\) LI = Includes low income communities
\(^d\) “Small area estimates” defined as probability-based estimates for populations at the county level or smaller

A number of surveys conducted in the Los Angeles area, statewide, and nationally met all of the necessary criteria:

- American Community Survey
- 2003 California Health Interview Survey (CHIS)
- 2005 California Health Interview Survey (CHIS)
- 2007 California Health Interview Survey (CHIS)
- 2005 Los Angeles County Health Survey
- 2007 Los Angeles County Health Survey
- 2009 District of Columbia Health Insurance Survey
- 2008 Massachusetts Health Insurance Survey
- LA County WIC Survey
- LA Mommy and Baby (LAMB) Survey
- Los Angeles Family and Neighborhood Survey (LA-FANS)
Making Connections Survey

Brief summaries of each included survey can be found in Appendix A. The summaries focus on the following aspects of survey design and data collection methods:

- Sample size
- Sampling plan
- Data collection procedures and methods
- Language
- Cooperation rate
- Response rate
- Estimating sampling error and weights

Appendix A also includes a table summarizing key features of their design and methods in order to make cross-survey comparisons. The table includes the following features:

- Sample frame
- Sampling strategy
- Inclusion of families with children under 18 years of age
- Inclusion of rural communities
- Inclusion of low income communities
- Oversampled population
- Areas where estimates are available
- Mode of data collection
- Incentives
- Language
- Sample size
- Cooperation rate
- Response rate

In the following section, a comprehensive review is presented of the sample design and data collection methods used by the surveys that met the criteria. Key considerations for the F5LA Family Survey are discussed, followed by a set of recommendations.
Summary of Studies and Key Findings

Since surveys of populations within LA County are of the greatest interest, a majority of the studies included in this review were conducted in Los Angeles or included specific probability estimates for LA County and smaller geographic units within the county. These studies employed a variety of sample frames, some of which represented newer approaches to dealing with the issues of coverage bias and non-response, particularly when looking at hard-to-reach populations or different language or ethnic groups. All of the studies included families with children under the 18 years of age (and, in some cases, children from birth through 5 years), rural households, and low-income communities. The surveys in this review also used a variety of data collection modes, including mail- and web-based surveys, telephone interviews, and, in a few cases, in-person interviews. Finally, the success of the studies, often measured in terms of response and/or cooperation rates, shows a high degree of variation. (The table in Appendix A summarizes features of each of the surveys included in this review.)

In order to sort out the advantages and disadvantages of each of these approaches, the characteristics of each of these studies are summarized according to three points typically important to the design of a household survey: a) identifying the sample frame, b) determining sample strategy and size, and c) selecting data collection mode and procedures.

Sample Frames

The choice of sample frame determines the underlying population upon which inferences from the sample will be made. In many respects it guides the sampling strategy, stratification, and sample selection procedures. Ideally, for the Family Survey design, the sample frame should identify all households with young children, ages birth through five years of age, who live within the Best Start communities. The challenge is to balance the tradeoffs between potential coverage biases with the costs/resources required to select a sample frame that will yield precise estimates, particularly of sub-populations.

Random Digit Dialing (RDD) Sample Frames

Many of the studies, including the three waves of the California Health Interview Survey (CHIS) and the two waves of the LA County Healthy Survey (LACHS) examined for this review, used Random Digit Dialing (RDD) methods that mainly relied on calls to landline telephones. More recently, CHIS 2007 began incorporating calls to wireless exchanges in California. According to the standard definition, an RDD survey is one in which households are randomly sampled within some geopolitical area using one of several random-digit dialing techniques that generate telephone numbers (AAPOR, 2011). RDD methods, sometimes enhanced with cell phone coverage, have both advantages and disadvantages. Extensive methodological literature now looks at the issue related to coverage bias, non-response, and incorporating cell phones into an RDD survey.

Single Frame Using Landlines. RDD sampling frame using landlines has been widely used in household surveys. But with increased use of wireless telephones (also known as cellular telephones, cell phones, or mobile phones), this method faces several challenges related to coverage. Over time there has been a steady increase in cell phone-only households in the United States and an increased use of cell phones. More than one of every four American homes (26.6 percent) had only cell phones during the first half of 2010—an increase of 2.1 percent since the second half of 2009 (Blumberg and Luke, 2010). The proportion of children living in wireless-only households increased by 3.1 percentage points from the last 6 months of 2009 to the first half of 2010, but this was not as large as the 4.6 percent increase from the first...
6 months of 2009 to the last 6 months of 2009. In addition, nearly one of every six American homes (15.9 percent) received all or almost all calls on wireless telephones despite having a landline (Blumberg and Luke, 2010). Boyle, Lewis, and Tefft (2009) suggest that the majority of adults living in “wireless-mostly” households are reachable using their landline telephone number. Among households with both landline and wireless telephones, 27.3 percent received all or almost all calls on the wireless telephones, based on data for the period January–June 2010. These wireless-mostly households make up 15.9 percent of all households.

Coverage bias with landline-based telephone surveys may differentially affect sub-groups within a population. Adults in households with only wireless telephones, compared to those living in households with landlines (with or without wireless telephones), tend to be younger, living with unrelated adult roommates, renting their homes, male, living in or near poverty, living in the Midwest, South and West (23.5 percent, compared with the 15.8 percent living in the Northeast), and Hispanic (Blumberg and Luke, 2010). Although these estimates did not change substantially during a 4-year period from 2007 to the first 6 months of 2010, there were differences related to gender, age, employment status, and household structure. There were also increases in the proportion of women, adults aged 30 years and over, adults who are unemployed or retired, as well as adults living with children, and decreases in the proportion of employed adults among all wireless-only adults (Blumberg and Luke, 2010). In the U.S., cell phone-only status is heavily influenced by age (more appealing for younger people) and marital status (more frequent among single people without children) (Keeter, 2006; Keeter et al., 2007). Thus, it is likely that excluding cell phones from an RDD survey may result in the under-representation of some groups.

**Dual Frame Using Landlines and Cell Phone Lines.** To address the coverage bias with the use of a landline-only RDD sample frame, there has been an increased use of a dual frame approach that includes both landlines and cell phone lines. A dual frame survey conducted in Portugal, in which both landline and cell phones were used to collect data, reported that cell phone-only respondents differed considerably from landline telephone respondents. But when blended with the landline telephone sample, cell-phone only respondents did not produce a significant change in the general population estimates on substantive variables (Vicente and Reis, 2009). This is explained by the fact that cell phone-only respondents are the smaller part of the population; moreover, the differences between these two respondent groups are relatively small. Further, Vicente and Reis (2009) point out that there are no existing sampling frames of cell phones, either for dual users or for cell phone-only users. This means that samples have to be randomly generated and significant amounts of screening are necessary to identify active numbers and cell phone-only users. Given the methodological and cost implications associated with cell phone-based designs, the authors conclude that the benefit of dual frame designs appears to be marginal, at least at present.

Findings from other studies dispute these conclusions and, instead, support the inclusion of cell phone-only households to reduce coverage bias in RDD studies. Using 2006 data from the National Center for Health Statistics' National Health Interview Survey, Blumberg and Luke (2007) conclude that the population of cell phone-only households in RDD studies is not sufficiently represented and can result in “non-negligible bias” for traditional RDD landline-only telephone surveys. Comparing wireless-only adults and adults with landline telephones on demographic characteristics and 13 key indicators (i.e., health status, health behaviors, health care service use, and health care access), Blumberg and Luke (2007) report that telephone surveys of landlines will underestimate the prevalence of health behaviors, such as binge drinking, smoking, and HIV testing—while obesity will be overestimated and physical activity will be underestimated for low-income young adults. These findings held even after using statistical adjustments to account for demographic differences between adults living in households with
and without landlines. However, CHIS 2007 found that complex weighting procedures are effective in adjusting for biases so that, following such a procedure, fully adjusted landline sample estimates are not significantly different from combined landline and cell phone sample estimates, even along health behaviors (CHIS 2008a). The use of post-stratification weights to adjust for non-coverage bias is a relatively controversial method and must be done with care.

Finally, a 2010 report from the Cell Phone Task Force of the American Association for Public Opinion Research (AAPOR) concludes that RDD frames that include cell phones have been demonstrated consistently to provide better coverage of a number of important demographic groups in the U.S. than landline-only RDD frames.

Brick, Edwards and Lee (2007) conclude that conducting a cell phone interview of the same scope and length as a landline RDD interview is feasible, but there will be lower response rates to the screening interview in cell-only households. This decreases efficiency and could increase costs. However, once cell phone-only households were identified in the study, the adults participated in the interview at the same rate as in CHIS 2005, despite less rigorous efforts in this pilot study to achieve high response rates (since this was not a primary goal of the pilot study). The relatively high level of participation by the adults in the cell phone sample was due, at least in part, to the fact that screened cell phone respondents enter the sample at a higher rate in cell phone samples compared with landline-only RDD surveys (where the person who answers the telephone may not be the adult selected to participate). This had the effect of increasing the adult response rate in the cell sample. The authors also found little or no evidence of cell phone sharing, and suggested that the efforts needed for screening and randomly selecting one adult from the household yielded little benefit in comparison to interviewing the adult who was screened (typically the one who answered the telephone). However, they caution that “it is too early to conclude that sampling adults in cell-only households is unwarranted” based on the results of this pilot study (Brick et al., 2007).

Several issues must be addressed when including cell phone lines in a RDD sample frame. To date, cell phone RDD samples in the U.S. have not been as efficient as landline RDD samples for many reasons. Thus including cell phone samples will lead to higher costs and the need for a larger initial sample size (AAPOR, 2010). In addition, telephone use in general (landline or cell) is not static but changes within households. Substantial segments of the population go without any telephone service for varying periods of time. As part of the 2007 LACHS, four large-scale surveys measured the characteristics of the non-telephone household population in California for the state’s Public Utilities Commission, each of which included large sub-samples in Los Angeles County. These surveys consistently reported that the portion of the population without residential telephones in Los Angeles County is not fixed but highly dynamic, with three in four of all non-telephone households reportedly having had some residential telephone service for a time during the previous three years (Field Research, 2008).

The key decision for researchers is whether or not to use a cell phone sample to supplement a landline RDD sample (a dual-frame approach). Assuming a dual frame, the next decision is whether the dual frame design will be overlapping (i.e., with no screening for telephone service and usage) or non-overlapping (i.e., screening the cell phone sample for cell phone-only persons/households). The AAPOR Cell Phone Task Force concludes that one approach to dual frames is not better than the other and the decision depends on the survey objectives and research questions (AAPOR, 2011).

In addition, the CHIS surveys have done much in the way of piloting new methods and testing alternative strategies for traditional RDD surveys. In fact, largely as a result of its pilot work, CHIS introduced a dual-frame RDD sample (cell phone and landline) beginning with the 2007
surveys. To increase precision for concentrations of ethnic groups within some areas, the RDD approach was supplemented by telephone directory lists. As a result of this pioneering work, the CHIS surveys have raised a number of considerations for incorporating a cell phone frame within an RDD survey (i.e., the dual-frame approach). One issue is the feasibility of randomly sampling adults within a household for the interview using a procedure similar to that used in many RDD telephone surveys in the United States. Many telephone surveys, including the 2005 and 2007 LACHS surveys, interviewed any adult who answered the phone, irrespective of household composition (Brick et al., 2007). The sampling of adults is necessary in RDD surveys to avoid the potential for bias, because the set of persons answering the telephone may not constitute a random sample of all adults; for example, these samples tend to be disproportionately female (Brick et al., 2007).

Another important issue is whether, in cell phone-only households, the cell phone is shared or used by only one individual. Interviewing the adult who answers the telephone in cell phone surveys would yield a probability sample of adults provided there is a one-to-one link between cell phones and adults in the household. But according to estimates from the cell phone supplement to the 2004 Current Population Survey (CPS), 66 percent of cell phone-only households where there was more than one adult shared the cell phone (Brick et al., 2007). This highlights the importance of conducting screening in households with cell phones as well as randomly selecting adults within households, particularly when a cell phone is shared. However, the 2004 CPS occurred at a time when cell phone-only households represented a much smaller proportion of the population than they do now. Although it is possible that shared cell phones in cell phone-only households may be related to lower economic status, the authors failed to confirm this hypothesis (using family income, home ownership, race, or education). They did report that households with shared cell phones were more likely to have married persons, people of Hispanic ethnicity, or five or more persons sharing cell phones.

Several other issues around cell phone respondents remain. A major operational concern is the cell phone respondent's ability to stay on the telephone long enough to complete the survey (compared with landline telephone users). Additionally, proper safety precautions must be incorporated when contacting cell phone users to ensure the respondent is not driving or otherwise occupied. In addition, researchers face issues with the transfer or porting of cell phone numbers. Individuals who originate their cell phone plan out of state and then move into the state will usually maintain the same phone numbers. Other cell phone users might transfer, or “port,” their landline telephone numbers to their cell phones. While the individual answers the call on their cell phone, the call may have originated from their landline, which is in service but clearly not their telephone of choice. As a result, the target state used to select the cell phone sample may exclude cell phone users with out-of-state exchanges. Moreover, cell phone numbers within the target state may not always be assigned consistently within the dedicated cell phone exchanges (e.g., the 1000 series blocks in California dedicated to wireless service). Fleeman (2007) reported that about five percent of respondents in a national sample of cell phone numbers identified an “outside” state in which their cell phone originated, and thus was not the same as the state used to select the sample. CHIS researchers reduced the influence of ported or misclassified cell phone numbers by obtaining a list of cell phone numbers from the regular “landline” RDD sample for the prior CHIS (2005) and adding these to the cell phone-only sample for this pilot study.

**Geographic Specificity.** The geographic specificity of RDD based surveys is of particular interest to the Family Survey since the Best Start communities are smaller than the county level and have a set of unique boundaries. RDD surveys that use either single frames (landlines only) or dual frames (landlines plus cell phones) are generally able to provide estimates of areas at the county level. These surveys may, within limits, provide estimates for smaller geographic
areas by using lists of exchanges or area codes limited to specific geographic areas within the county. A caveat for cell phone samples is that these exchanges may not provide optimal coverage due to the porting and transfer issues discussed above. Additionally, cell phone numbers are not listed by geographic area, an important limitation of using cell phone samples for geographic areas smaller than the county. The smallest geographic area to stratify would be the area code and/or the telephone exchange (the next three digits following the area code). Examples from CHIS and LACHS reveal how RDD surveys can provide some estimates for areas smaller than the county, with some important limitations.

In CHIS 2007, the landline sample was comprised of 44 sampling strata, of which 41 were at the county level and three others combined several smaller counties into the same geographic areas. For Los Angeles County, the survey used an enhanced stratification based on the LA County Service Planning Areas (SPAs). SPAs are areas established by the LA County Department of Public Health to plan and manage health service delivery. Los Angeles County was stratified by SPA to provide smaller geographic estimates (CHIS, 2007). Specifically, zip code information allowed telephone exchanges to be classified into eight subsampling strata, each representing an SPA. Telephone exchanges that crossed SPAs were assigned to the SPA with the most listed households. There were no targets for individual SPAs, so the sample for Los Angeles was allocated proportionally by these substrata.² To increase the precision of estimates for Koreans and Vietnamese, areas within counties with relatively high concentrations of these groups were sampled at higher rates. These geographically targeted oversamples were supplemented by telephone numbers associated with group-specific surnames drawn from listed telephone directories to further increase the sample size for Koreans and Vietnamese (CHIS, 2007). The RDD cell phone sample in CHIS 2007 was stratified by area code because numbers assigned to cellular service were not available to support the same level of geographic stratification as the landline sample. This approach can provide estimates for areas smaller than the county level, although the boundaries may not match up with the defined Best Start community boundaries.

LACHS provided estimates at the health district level (SPA), as well as areas aggregated from census tracts, which are smaller than the SPA. Using RDD telephone listings purchased from a supplier of telephone samples to the survey research industry, LACHS systematically selected from all eligible 10-digit telephone numbers assigned to Los Angeles County. The sample frame consisted of all possible numbers in active blocks of 100 telephone numbers, using the first eight digits, including the telephone area code and exchange plus one digit. Only the blocks that were found to have at least 3 directory-listed numbers were eligible for inclusion in the LACHS sample frame. Thus, it is possible to identify telephone exchanges serving cities, towns, and zip codes within the county.

It appears that the use of telephone lists to target geographic areas smaller than the county may be somewhat inexact. While it is possible to use telephone lists for both landlines and cell phones in this way, it is not clear how well these areas map onto the boundaries of the 14 Best Start communities. Further, to obtain precise probability-based estimates for each community, the starting sample must be considerably larger than that required to obtain estimates for larger areas, due to the lack of fit between telephone-based lists and the target geographic areas. This reduces the efficiency of the sample.

² The eight SPAs in LA County were as follows: San Fernando, San Gabriel, Metro, West, South, East, South Bay, and Antelope Valley. The sample sizes for the individual SPAs were selected proportional to size, except for the sample for Antelope Valley, which was increased to yield 250 adult interviews more than what would be expected from proportional allocation.
Address-Based Sample Frames

A new sampling approach has emerged in response to the limitations of RDD studies, including under-coverage and low (and declining) response rates: Address Based Sampling (ABS). Historically, it has been difficult to undertake large-scale household surveys because lists of addresses have not been generally available in the United States. As a result, national household surveys relied on either RDD sampling of telephone numbers or multistage area sampling in which the list of addresses in constructed. A multi-stage area sample is a form of area probability sampling in which geographic areas are sampled with known probability. Typically these designs select areas as part of a clustered or multi-stage design whereby households and families or individuals within households are sampled within the geographical areas selected for the sample (Hall, 2008). At each stage, the units under consideration, such as households, are selected using either simple random sampling or proportional to size methods.

The most promising and cost-effective method of implementing ABS uses frames that are based on the United States Postal Service (USPS) computerized Delivery Sequence File (DSF), as exemplified by the American Community Survey. A number of the state health surveys in the table in Appendix A used an ABS approach in conjunction with RDD strategies (e.g. the 2008 Massachusetts Health Insurance Survey and the 2009 District of Columbia Health Insurance Survey).

One of the strengths of an ABS sample frame is that it allows for coverage of small geographic areas, such as neighborhoods and communities. This is demonstrated by the Making Connections household survey, in which addresses from the frame were selected if they were in any of the targeted census tracts. The disadvantage to this approach is that, often, defined neighborhoods or communities do not easily fit within census tracts. To address this challenge, the LA-FANS survey defined its target communities according to census tracts, which then allowed the investigators to sample census tracts.

To make estimates of small geographic areas, such as communities, it is important to match the defined community boundaries with an available sample frame, and this process has implications for the Family Survey. As part of its place-based initiative, F5LA conducted a study of boundaries that defined the 14 Best Start communities. The initial boundary for each community was based on census sub-blocks, high school feeder boundaries, municipal boundaries, and a requirement that communities contained approximately 5,000 children under the age of 5. With the exception of two communities (Metro LA and Long Beach), each Best Start community will have the opportunity to determine whether the initial boundaries are reflective of their community and propose revisions to the boundaries. The proposed revised boundaries will be vetted by F5LA to ensure that they adhere to certain parameters such as maintaining birth rates within five percent of those within the original community boundaries; built and natural environment features that are commonly identified as boundaries to community members or provide for natural community boundaries; demographic and community attributes that identify the community as high need; and, the race/ethnicity of the diverse communities. Based on this boundary revision process, F5LA concluded that clusters of census tracts between 1600-1900 residents would comprise the smallest area for selection. These clusters would be amenable to either a multistage area sample design or an ABS approach.

When using ABS, it is important to understand what is included or excluded from the list of addresses. In raw form, the USPS list of addresses has important limitations in terms of coverage, notably for households in areas without residential mail delivery that receive their mail either at a general mail delivery facility or a Post Office Box. For these households, the initial USPS list would not include the physical address of the household. However, this is partially
corrected in the Delivery Sequence File, which lists all delivery point addresses serviced by the USPS, with the exception of general delivery. Similarly, the rural route, in which the address consists of a route number and not a recognizable street number, is a source of undercoverage. Even if the route was identifiable, the box for the route might not be near the residence. One study comparing counts of residential mailing addresses to Census projections of households estimated that 1.3 percent of households nationwide receive mail at P.O. Box but not their physical address, and 3.9 percent have unlocatable rural route addresses (Staab and Iannacchione, 2003). In a further study of household addresses in a single metropolitan area, Iannacchione, Staab, and Redden (2003) found that 1.9 percent of randomly selected households had missing addresses and that the majority of persons who maintained a residential Post Office Box also had mail delivered to their street address—thereby demonstrating the utility and completeness of using mailing addresses to develop a sampling frame for a metropolitan household survey.

An important consideration when using an ABS frame is the identification of households consisting of adults living in non-institutional, civilian settings such as dorms, assisted living facilities, halfway homes, and shelters. These “group quarters” are not identified on the raw USPS lists and often represent, along with institutional settings such as prisons and army barracks, key sources of potential under-coverage when using ABS. In all of these situations, the address list needs constant updating, which requires the attendant costs and resources needed to ensure that all current household addresses are included.

The American Community Survey (ACS), by the U.S. Census Bureau, is instructive for understanding the advantages and disadvantages of an ABS approach. The sample frame for the ACS is the Master Address File (MAF), which is the Census Bureau’s inventory of known residential addresses (housing units and group quarters) and selected nonresidential units in the United States and Puerto Rico. The addresses are selected from the MAF to represent both the Housing Unit (HU) and Group Quarters (GQ) populations, although different methods are used to select and interview these two groups. When all interviews are combined, the survey data provide estimates of the characteristics of the total residential population.

The MAF contains mailing and location address information and other attribute information about each address. It also contains geographic codes, such as county and place codes, obtained by linking to the Census Bureau’s Topologically Integrated Geographic Encoding and Referencing (TIGER®) database. The geographic codes in the MAF, some of which come from the TIGER® database, identify a variety of areas, including states, counties, county subdivisions, Native American areas, Alaska Native areas, Hawaiian Homelands, census tracts, block groups, and blocks. At the conclusion of Census 2000, the MAF contained a complete inventory of known housing units (HUs) nationwide. This served as the starting point for the ACS sample frame, but it continues to be updated on a six-month basis as the ACS is rolled out.

Sampling for Group Quarters (GQs) was based on a separate, group quarter-specific MAF, which was constructed from an updated GQ inventory file following the last decennial census. This was merged with files from federal prisons and detention centers from the U.S. Bureau of Prisons, as well as a file of military bases and vessels from the Department of Defense. In addition, the Census Bureau identified new state prisons and state prisons that had closed based on internet research.

For city-style addresses, typically found in urban and suburban areas, updating the MAF is based almost entirely on the US Postal Service’s DSF. The Census Bureau receives updated versions every 6 months. There are many known problems with the DSF as the source of updated households for the MAF. First, it misses many addresses in new construction areas,
where it takes time to establish separate mailboxes and mailing addresses. Second, portions of the DSF are not updated at the same rate all around the country, and third, it does not clearly identify addresses in small multi-unit structures where mail may be delivered to a central hall or desk and not to the individual apartments (U.S. General Accounting Office, 1998).

To mitigate coverage issues for the 2000 decennial Census, the Census Bureau used a complete canvass of all 8.2 million blocks in 1999. To improve coverage for the 2010 decennial Census, it repeated this costly block canvass operation in 2009, supplemented by a Local Update of Census Addresses (LUCA) program in 2008, in which local governments are given the opportunity to review and update the residential address listings for their jurisdiction. These efforts improved the completeness of the 2010 Decennial Master Address File, giving better point in time decennial estimates. But the ACS, which uses time period estimates, still suffered from less precision by under-representing growing areas of the country, and not accurately representing residences in small multi-unit structures (those with 2–9 apartments). Further, evidence from the 2000 Census indicates that the problem of missed or erroneously identified addresses in these types of structures persisted in the 2000 MAF even after the block canvass and LUCA programs (Citro and Kalton, 2007).

For rural addresses that involve route numbers and P. O. Boxes, a different approach to improving the sample frame was taken. Starting with the MAF from the last census, the frame added new housing information from the Community Address Updating System (CAUS), which covers largely rural blocks where use of the DSF does not provide adequate coverage. A set of administrative procedures consider the address characteristics of existing MAF records for the county, changes in housing unit estimates for the county since the last decennial census, and changes in the DSF tallies for the county. The procedures also identify blocks that would be expected to yield the most new units. Dean and Petersen (2005) found that CAUS was successful in adding addresses to the MAF that would not have been added by other means, but the study was limited and did not address the constraints of using CAUS to improve the identification of rural addresses.

It is clear that maintaining an updated Master Address File that accurately covers the entire population of a given geographic area (or nationally) is a continuous and time-consuming process. Although it is relatively easy to purchase available address lists that originate from the USPS Address Management System (AMS), these cannot be purchased directly from the USPS. Rather an organization already having a list of residential addresses can get the delivery information for those addresses, and thus confirm that those addresses are correct. These organizations are usually vendors dealing with direct mail or other marketing agencies. The better vendors will have one of two licenses with the USPS: a license to the Delivery Sequence File (now in its second generation and known as the DSF2) or a license to the CDS file, a 5-digit ZIP code-based file that provides the same delivery information as the DSF2. For a vendor with a CDS license, the USPS will attach delivery information for the units that the vendor and USPS have in common. It will also update the vendor’s list in the process. The costs of the lists vary by vendor with prices ranging from as high as $25 per 1,000 addresses to as little as $8 per 1,000 addresses and, in some cases, the cost structure was based on the total number of addresses purchased (P. Fletcher, personal communication, April 18, 2011). A single, well-known vendor was involved in generating address-based lists for several ABS studies that were included in this review, including the 2008 Massachusetts Health Insurance Survey and the 2009 District of Columbia Health Insurance Survey.

Finally, ABS surveys must take into account individuals who move residences or are transient. To count individuals in both housing units and group quarters, the ACS uses a “current residence” rule: “people who live for more than 2 months at a sample address are assumed to
be residents of that unit” (Citro and Kalton, 2007). It is also applied prospectively to include those who just moved to a residence but intend to live there for at least 2 months. This differs significantly from the decennial census which employed the “usual place of residence” rule. The change in the residency rule was a result of the ACS being a continuous monthly survey rather than a point in time survey such as the decennial census. There were a number of exceptions to the two-month rule to cover several typical situations such as children living in boarding schools or summer camps, children in joint custody arrangements, and people who commute to a residence close to work but who also maintain a home residence. The Panel on the Functionality and Usability of Data from the American Community Survey (Citro and Kalton, 2007) criticized this definition because no explanation was provided for choosing this time frame over another. The panel also noted that many exceptions were missed, including “people with weekday and weekend residences, people who live and travel throughout the year in recreational vehicles, and people who move among the residences of several relatives or friends” (Citro and Kalton, 2007). It is not clear the extent to which this rule might miss or overestimate people living in specific residences but who change their residences over time.

Several studies included in this review employed multistage area sampling to produce the sample frame of households and then sampled households from the frame. These studies either constructed their own lists of addresses within census tracts (LA-FANS) or used an address-based approach that fit addresses to selected census tracts (Making Connections). Several studies used other forms of list-assisted sampling designs, including enrollment lists of WIC participants (LA County WIC Survey), making the sample representative of WIC recipients but not necessarily of the general population. One survey used the birth certificate registry (LA Mommy and Baby Survey), which has problems in terms of coverage and accessibility to the registry.

The multistage area sample approach, particularly the strategy employed in LA-FANS, selected census tracts proportional to size from all census tracts in LA County, and then enumerated households within these tracts. This is very expensive and requires a considerable amount of time to arrive at the final sample frame of households compared to RDD surveys. The advantage over using an existing list frame is better overall coverage of the population and somewhat better cooperation and response rates, since verification of households can also serve as an initial screening for eligibility.

Finally, some studies, notably the 2009 District of Columbia Health Insurance Survey and the 2008 Massachusetts Health Insurance Survey employed a dual-frame RDD plus ABS design. While this might seem an appealing choice for the Family Survey design to minimize coverage bias and non-response, there are several important disadvantages, including costs and response rates, which are discussed later in this paper.

Issue of Migration

Regardless of the sample frame used, migration patterns and the rate that the population moves into and out of the target communities need to be taken into account when developing an appropriate sampling design and plan. The existing literature on migration patterns of Los Angeles residents appears outdated and focuses primarily on the economic reasons for residents moving out (Gabriel and Mattey, 1996). For the design of a household survey, it would be important to know the degree of resident migration and the degree to which such migration is to adjacent areas as opposed to outside the LA County area. This information can help to ascertain the stability of the community populations and resident demographic characteristics so that variations in survey results from one semi-annual period to the next can be placed within context. Furthermore, if migration patterns are known, then adjustments to the sample frame for
each successive survey will ensure adequate coverage of sub-groups (e.g. ethnic, language or economic groups) moving into or out of the Best Start communities.

F5LA recently commissioned a Demographic Migration Study to provide information about the migration patterns of the 14 Best Start communities in order to assist in the Family Survey design. It is anticipated that the Demographic Migration Study will include information on home ownership and rental rates, as well as forecast population flows and demographic descriptions of the 14 communities for the next five years. This study is still in progress, but when the results are available they will be used to inform the Family Survey sample design plan.

**Sampling Strategy and Size**

Variations of multistage or cluster sampling strategies were used in all of the ABS or dual-frame ABS/RDD studies reviewed. This approach involved selecting households within census tracts and then individuals within households, with either simple random sampling (SRS) or sampling with probability proportional to size (PPS), and with or without some stratification for subpopulations of interest. Since communities may vary considerably in population, SRS does not adequately adjust for differential selection probabilities for communities of different sizes. As a result, sampling with PPS may provide greater sample precision. In PPS, each household in the sample will typically have an equal chance of being selected and thus the sample is said to be self-weighting, which can simplify the analysis.

All of the studies reviewed used various forms of stratification to ensure that key subgroups in the population were adequately represented in the sample. Studies stratified on one of several factors: housing density size to adjust for rural populations (the ACS), poor vs. non-poor census tracts (LA-FANS), and subareas within cities (Making Connections). In most of the RDD surveys, geographic regions were used as the strata. The CHIS studies stratified by large counties in California including Los Angeles County, in order to provide a large enough sample to allow for separate estimates in these counties. In CHIS 2007, stratification also included several Service Planning Areas around LA County.

When using multi-stage samples with stratification, it is often important to estimate the effects at different levels of the model. For example, LA-FANS employed multi-level modeling to estimate the effects of neighborhood-level characteristics, as well as between-household characteristics within neighborhoods. Multi-level modeling is also important when analyzing hierarchical data because it adjusts for the inherent clustering of smaller units within larger units, such as individuals within households within neighborhoods. The goal is to generally choose the sample size at each level that balances precision with cost—and often costs are determined by the largest unit chosen and the number of observations per unit. This also assumes that the sample size is relatively constant across the higher level of unit, such as community or neighborhood (for our purposes).

While it is important to keep in mind the minimum sample size requirement for analyses of key hypotheses, this must be balanced with the need for sufficient sample to make precise estimates of the population. Although increasing the number of macro units will tend to reduce the standard errors for estimating a given set of parameters, there will be a tradeoff in terms of the number of observations that can be collected at the smaller unit of analysis. Costs tend to be driven more by adding the macro units (i.e. census tracts) than by adding observations at the smaller units (i.e., households and individuals within households) because of the marginal data collection costs associated with adding each macro unit. When estimating the random components in multilevel models, designs with larger numbers of observations per group tend to
be more efficient. Hox (1998) suggested that to estimate models with random slopes there should be approximately 50 groups (macro level) and at least 20 individuals per group.

An example of balancing costs with precision is seen in the LA-FANS survey, in which the main purposes for arriving at its sample size were to support multilevel statistical analyses and generate reliable estimates for LA County. They concluded that a target sample of 3250 households was optimal, based on a generic test of proportions between two comparison groups, which is similar to a logistic regression with a single explanatory variable. The resulting effect size of 0.16, considered small by Cohen’s classification, required a sample size of 325 households per neighborhood. The neighborhoods were defined by census tracts which averaged 5600 households per census tract. However, this assumes drawing a simple random sample from the population of LA County. In multilevel, clustered designs, the sample size generally has to be increased to take into account the design effect.

Clustering produces a design effect greater than 1.0. The design effect for statistical estimates is a function of the intraclass correlation (ICC) and the cluster size. The LA-FANS survey had to take into account both intra-community correlations as well as the larger intra-household correlations. Assuming a cluster size of 50 households per community (65 communities overall) their design effects ranged from 1.45 to 3.45, which are substantial. The researchers concluded that they needed 1125 households per stratum and the study called for three strata (proportions of households in poverty in census tracts from very poor, near-poor and non-poor). The researchers then conducted a simulation study using the 1990 Census STF3A tables for LA County in which census tracts were sampled, as well as block groups within census tracts, and finally individuals within the block groups. The simulation study was designed to identify the trade-offs associated with changing cluster size (the number of census tracts) while holding sample size constant. They found that precision improved as the number of sampled clusters increased from 51 to 66, but that little was gained by increasing the cluster sizes to 75 or 81 (and of course the costs would be higher). They concluded that a total of 65 tracts sampled from approximately 1600 eligible tracts best balanced cost with precision. This study illustrates one method for choosing sample size and highlights the cost and design considerations that pointed towards selecting the smallest number of macro units or clusters that met the study needs.

Finally, regardless of the survey design selected, it would be important to revisit the notion that the two yearly samples drawn from the sample frame should be separate and independent. For periodic surveys of time series, designed to detect trends in risk and change over time, constant-size samples with partial overlaps are a favored design (Fletcher, personal communication, April 18, 2011). The gains in precision due to correlation over time are directly proportional to the overlapping portion of the sample. In an analytic model of the difference between means for two characteristics or means for the same characteristic at two points in time, the variance of the difference should include a covariance term. When the two samples do not overlap at all, the covariance term is effectively removed, which has the effect of increasing the overall variance of the difference. While it may be intuitively obvious that each sample would have greater precision if drawn independently, this may not allow for the detection of changes in the communities over time due to the lack of a covariance term in the analytic model. Thus, consideration must be given to the greater detection of change over time in the communities, which may require an overlapping sample design.

**Data Collection Mode and Procedures**

Studies reviewed for this paper used a variety of data collection modes. Many of the more complex studies involving dual-frame ABS and RDD employed mixed modes of data collection.
While many RDD studies used telephone interviews, the dual frame ABS and RDD surveys combined telephone interviews with either mail surveys and in some cases web-based surveys.

An important issue for the mode of data collection is whether to conduct interviews, either in person or by telephone, or use more “hands-off” approaches, such as mailings or web-based surveys. Using interviewers who can employ structured but conversational interviewing techniques often ensures that respondents understand the questions, thus allowing for the uniform interpretation of the intent of each question, which should lead to more accurate responses (Conrad and Schober, 2000). However, when used improperly or without sufficient protocols in place, this technique can lead to biased results (Beatty, 1995). It is important that each interviewer handle and interpret each question in the exact same manner even though they may use their own personal style to how the questions are asked.

The interviewer bias or interviewer effect has been defined as the “tendency for answers provided by the respondent and recorded in a questionnaire to vary depending on which interviewer is assigned to the respondent” (Johnson, Fendrich, & Shaligram, 2000, p. 77), and has been detected even in telephone surveys (Johnson et al., 2000; Lueptow et al., 1990). Evidence shows that interviewer gender and race may have an impact in telephone surveys (Johnson et al., 2000; Lueptow et al., 1990), but the effects may be relatively benign except in questions related to race/ethnicity attitudes. With race/ethnicity attitude questions, respondents may tend to give answers that avoid offending the interviewers (Davis, 2010; Cotter, Cohen, and Coulter, 1982; Reese, Danielson, Shoemaker, Chang, and Hsu, 1986). While some studies found support for higher cooperation rates when interviewers and respondents share similar social statuses or educational backgrounds (Durrant, Groves, Staetsky, and Steele, 2010), others reported inconclusive findings due to differences in the definitions or measures of social status across studies (Davis, 2010).

Use of Indigenous Interviewers. One method long discussed in the literature to improve cooperation, response rates, and data quality is the use of individuals who come from the target survey communities to serve as interviewers (i.e., “indigenous interviewers”). In this approach, front-line research workers, often those doing screening, recruitment, or data collection, share important characteristics with research participants, which give them a social and experiential context different from that experienced by the traditional research interviewers, who work with similar target research populations but are not indigenous to the target study population. Due to their familiarity with the social and physical environments of the target research population, indigenous interviewers may possess contextual knowledge that provides “improved access to an otherwise highly restricted group, enhanced rapport-building capabilities, and greater understanding of the language unique to the study population” (Berg et al., 2004, cited in Alexander and Richman, 2008). Any set of characteristics may make individuals indigenous if they allow these individuals to become the “ultimate insiders,” such as living in the same neighborhood, coming from the same racial or ethnic group, speaking the same language(s), having the same social class or educational background (typically community college or less), and/or having personal experience with the focal problem addressed by a research project (Alexander and Richman, 2008). The use of these research workers may be especially beneficial for surveys in poor communities with subgroups of linguistic minorities and hard-to-reach individuals where respondents may be distrustful of researchers’ intentions. For example, community residents may not want to participate in the survey if they feel the community is not deriving any benefit from the research (Josephson, 1970).

Using indigenous interviewers can strengthen the researchers’ credibility in the community being studied. It also maximizes the “process of exchange,” that is, the larger participation and involvement of community members with aspects of the research endeavor such as survey
design, questionnaire development and survey data collection (Holbrook, Farrar and Popkin, 2006). By using indigenous interviewers, the community may have a greater investment in the survey, while the researchers “give back” by providing employment (Holbrook et al., 2006). Moreover, when sensitive questions are administered to minorities, such as interviewing Latino migrants about risky sexual behavior, a greater degree of trust would improve data quality and response rates (Parrado, McQuiston, and Flippen, 2005). Indigenous interviewers are often used in communities where “in-house” interviewers hired by the researchers may not feel safe, such as in surveys of residents of Chicago public housing developments (Gwiasda, Taluc, and Popkin 1997), urban youth (Torres 1998), and studies designed to count the homeless (e.g., Devine and Wright 1992; Stark 1992; Applied Survey Research 2002, 2005).

However, there are also several important disadvantages to using indigenous interviewers. Weinberg (1971, cited in Holbrook et al., 2006) emphasized that indigenous interviewers require much more training and that the researcher usually needs to initially hire from one-quarter to one-third more indigenous interviewers to account for attrition (Weinberg, 1971 cited in Holbrook et al., 2006). Indigenous interviewers typically are less well-educated and, as a result, may make more errors in data collection. In addition, they may not follow instructions consistently or engage in the probing needed, thereby making the interview less standardized (Gwiasda et al., 1997; Weinberg, 1971 cited in Holbrook et al., 2006). These studies also suggest that indigenous interviewers may be more prone to falsifying data and that these problems have been reported even when indigenous interviewers have been given additional training and supervision (Gwiasda et al., 1997).

Several ethical issues may also be raised by the use of indigenous interviewers, for which there currently is minimal formal guidance or evaluation (Alexander and Richman, 2008). Indigenous interviewers are often not trained formally on the regulations and ethical guidelines for research as covered by the Belmont Report and the Common Rule governing research with human subjects (45 C.F.R. 46.111). In some cases where training had taken place, indigenous interviewers were vague about the content of the training or what specific guidance they had taken from it (Alexander and Richman, 2008). During focus groups, Alexander and Richman (2008) report that indigenous interviewers tended to be suspicious of the need for following a research protocol; they were concerned about their own personal safety; they voiced a strong desire to maintain existing relationships in their communities; and they did not want to coerce respondents. The authors suggested that more targeted and nuanced training approaches for indigenous interviewers are needed to protect research participants, enhance research integrity, and promote collection of valid data in vulnerable populations (Alexander and Richman, 2008). However, most of the research reviewed on this subject involved only indigenous interviewers and thus it was impossible to empirically test the advantages or disadvantages of these interviewers when compared with in-house interviewers hired by the research team.

Holbrook, Farrar and Popkin (2006) were the first to empirically compare indigenous interviewers with in-house interviewers on the quality of survey data. In a survey of residents in a Chicago neighborhood housing project that asked several sensitive questions regarding lease violations or their status as residents in the project, the authors found no differences between indigenous and in-house interviewers in their ability to gain trust from respondents. Indigenous interviewers did not obtain better cooperation, nor did they appear to elicit more honest responses to sensitive questions, and, when there was a difference between in-house and indigenous interviewers, in-house interviewers obtained more honest responses to sensitive questions. Holbrook, Farrar and Popkin (2006) also reported administrative and training difficulties associated with using indigenous interviewers. An important limitation of this study is that both the in-house and indigenous interviewers were African-American, and that the only difference was that the indigenous interviewers came from the same Chicago housing complex.
where the sample was recruited. By being similar in racial/ethnic group to the indigenous interviewers, the likelihood of finding significant differences was greatly reduced. Further, the distinction between the two interviewer groups was not based on language or social class but rather on residency thereby limiting the generalizability of the findings. Nevertheless, the authors point to the greater training and experience of in-house interviewers, as well as the greater social distance between in-house interviewers and respondents, which allowed respondents to feel more comfortable and to give more honest answers. The effect of social distance is supported in a study of outreach strategies in a face-to-face survey of individuals seeking HIV prevention services in New Jersey. Crowley, Roff, and Lynch (2007) report that caseworkers affiliated with community-based organizations (CBOs) had lower response rates compared with “outsider” researchers.

The importance of interviewer training and experience is consistent with other research showing that interviewers’ experience, expectations, attitudes, and behavior during their interactions with respondents, as well as their confidence, improve respondent cooperation (Grove and Couper, 1998; Durrant et al., 2010), as does the technique used and the interviewer’s behavior (Beatty, 1995; Keeter et al., 2000). Thus, the ability of the interviewer to develop rapport and sensitivity with each respondent may be more important than the similarity in backgrounds between interviewer and respondent and whether the interviewer is indigenous.

The importance of interviewer training and experience was found by Project Director Resnick in his experiences training interviewers and conducting quality control in the field for the national Head Start FACES study. Head Start centers that were initially reluctant to participate in the study during the first data collection later welcomed back the field research team, asking for the same interviewers, in subsequent rounds of data collection. Many of the parents who were interviewed the first time remembered the interviewer and gladly participated in later interviews, a result largely due to the rapport the interviewer established during the initial round. These factors appeared to explain the relatively high response rates for the parent interviews across multiple rounds of FACES, which were at least 80 percent, often higher (Resnick, 2010; Zill and Resnick, 2006).

**Computer-Assisted Interviewing.** Computer-assisted interviewing (CAI) includes computer-assisted personal interviewing (CAPI), computer-assisted telephone interviewing (CATI), and computer-assisted self interviewing (CASI). Most large scale federal surveys as well as many of the studies that were reviewed for this paper (e.g., CHIS and LACHS) utilized CATI and CAPI strategies. These methods constitute the gold standard for surveys involving telephone as well as in-person interviews, particularly for large samples (Tourangeau, 2004). There are several significant steps in designing and implementing a CATI/CAPI survey beyond what may be anticipated in a paper-and-pencil survey, particularly the development and programming of the survey instruments using either off-the-shelf or customized software.

Commercial off-the-shelf computer-assisted interviewing (CAI) software is available and has often been used in national surveys including the Early Childhood Longitudinal Studies, the National Health Interview Survey, and CHIS. Available software for each type of interview method generally supports structured questionnaire design, complex routing and checking, interactive data entering, top-down editing with interactive tables, and forms-based data entry. The best software should allow for an unlimited number of questions and records and should provide tools for survey management, data manipulation and tabulation, interactive coding, and the export of data to statistical packages. Typically, large, national survey research firms customize the off-the-shelf software to maximize the advantages of a CAI system when conducting telephone or in-person interviews.
An example of customization can be found in the telephone research centers developed by a national survey firm for its implementation of such studies as CHIS. Each center uses the same proprietary telephone system and a computerized survey management system. Employing specially-designed computer software programs, their CATI system allows for reducing costs by using a computer program to route telephone calls to the least expensive long-distance telephone supplier. The system features computerized scheduling of interviewing assignments and callbacks, computerized reports of production, call status, and response rates, call detail reporting for control and billing purposes, an autodialing capability that interacts with the computerized interview scheduling system, and specific strategies for the release of the sample in the CATI automated scheduler. Additionally, there are a variety of quality assurance procedures in place, from monitoring a sample of telephone calls in real time to ongoing data quality checks and procedures for callbacks, if needed.

Computer-assisted strategies also require extensive consideration in the hiring and training of interviewers, particularly to train interviewers to administer the survey in different languages. For example, CHIS 2007 interviewed sampled respondents in up to five languages. An important consideration for in-person household interviews is the number of laptops that must be purchased and the maintenance required to keep them operating, including the regular upload of completed survey data to the home computer for production quality checks and data cleaning.

The key features of CATI/CAPI systems may carry more initial costs than a paper-and-pencil survey. An important question is whether the additional costs are justified by the benefits. Studies comparing CATI with paper-based methods found that the computer assisted surveys provided a relatively small advantage in cost, timeliness, and data quality, although these studies were conducted when CATI/CAPI methods were in their relative infancy (Bergman et al. 1994, Catlin & Ingram 1988, Groves & Mathiowetz 1984). At the same time, the use of computer assisted interviewing (CATI or CAPI) typically produces lower rates of missing data and the ability to administer more complicated questions and skip patterns, especially for telephone interviews. There are economies of scale when considering very large sample sizes, so that these benefits including lower error rates and improved response rates become much more important as sample sizes increase. When comparing the two methods, telephone and in-person computer assisted interviewing methods, some studies found that CAPI had less of an effect on survey costs, response rates, the composition of the samples, or the respondents’ answers (Baker et al. 1995, Martin et al. 1993, Tourangeau et al. 1997).

**Data Collection Mode.** Many of the studies reviewed used mixed data collection modes—mail, web, telephone or in-person—in order to improve data quality and to boost response rates. But when mixing data collection modes, an important consideration is the potential effect of measurement error and differential respondent coverage inherent in each mode of data collection. For example, Bech and Kristensen (2009) compared response rates from postal and web-based survey data collection modes in a study of preferences for the design of future nursing homes. Keeping the questions the same, they randomly assigned their sample of 10,000 individuals aged 50-75 years to receive either a postal questionnaire or a letter with a Web link to an online version of the same questionnaire (to control for initial introduction to the survey). The results showed that response rates were significantly lower for the Web-based survey. While the Web-based version improved the sample representativeness with respect to gender, it underrepresented older respondents. Respondent characteristics in the Web-based survey differed significantly from those of respondents in the postal survey with respect to income, education, civil status and health status. The Web-based version improved data quality by significantly lowering the amount of item non-response and ‘don’t know’ answers but the cost per response was significantly higher for the Web-based survey due to the lower response
rates. However, the authors admit that the lower response rate in the Web survey may have been a result of requiring respondents to switch modes, that is, taking the letter to a computer and typing in the Web link before the questionnaire could be completed. Nevertheless, the coverage bias in Web-based surveys found in this study is consistent with other research and the differences in response rates between the two modes (42 percent for the postal survey compared with 17 percent for the Web-based survey) cannot be explained entirely by “mode switching.” The authors also point out that, in their country, access to computers and the Internet is over 86 percent for individuals aged 40-59 years and 57 percent for those aged 60-74 years thus removing any effects of differential computer/Internet literacy (Bech and Kristensen, 2009).

To improve precision and data quality, it is important to consider data collection mode within the larger context of the survey sampling plan and sampling frame. The DC Health Insurance Survey used a dual-frame ABS and RDD approach and relied on two interview modes: telephone and web. Respondents were also given the option of requesting a paper copy of the survey. Information for the RDD-sample was obtained using traditional telephone interviewing methods, with a web-based option offered to the sample members for whom addresses could be obtained (Ormond, Triplett, Long, Dutwin and Rapoport, 2010). Advance letters were sent to the ABS sample as well as the RDD sample for which addresses were available, followed by telephone interviews. Reminder post cards were sent to all households with a listed address. A final round of mailings was sent to all non-responding households two weeks after the post card reminders were mailed (Ormond et al, 2010).

The American Community Survey shows how different data collection modes may be used together in the same survey and it arguably may be considered the “gold standard” for mixed methods of data collection in household surveys. One generally accepted measure of the success of a data collection effort is the final response rate, although there are limitations of using this metric that are explored in some depth later in this paper. Assuming general acceptance of this standard, the ACS stands apart from most other studies in achieving, through multiple modes of data collection, a high response rate of 97 percent for the first five-year survey segment. This standard is often aspired to but rarely achieved by most surveys, including the rest of the surveys reviewed here. The high response rate in the ACS seems due, at least in large part, to the efforts of the interviewing staff in the telephone centers and regional offices and the use of a mixed mode of data collection that focused on the non-responders. But the high response rate is also related to the mandatory nature of the survey by the Census Bureau and thus cannot be expected to be achieved by most other surveys.

Before discussing the issue of response rates as a measure of survey success, it would be useful to examine the ACS data collection modes in greater depth. The data collection operation for housing units (HUs) consists of three modes: mail, telephone, and personal visit. For most HUs, the first phase includes a questionnaire mailed to the sample address, with a request to the household to complete the questionnaire and return it by mail. If no response is received, the Census Bureau follows up with CATI when a telephone number is available. If the Census Bureau is unable to reach an occupant using CATI, or if the household refuses to participate, the address may be selected for CAPI (U.S. Census Bureau, 2009). The ACS includes 12 monthly independent samples.

Data collection for each sample lasts for 3 months, with mail returns accepted during this entire period. This three phase process operates in continuously overlapping cycles so that, during any given month, three samples are in the mail phase, one is in the CATI phase, and one is in the CAPI phase. Thus, for data collection of housing units in the ACS, there were three data
collection strategies with increasing levels of intensity (mail, telephone interviews, and in-person interviews) to obtain survey data for the panel over a three-month time period.

There are three to four mailings to each sample address, depending on when a return is received and respondents can request Spanish or English mailing packages via a Telephone Questionnaire Assistance (TQA) telephone number. The mailings include a pre-notice letter, an initial mail package, and a reminder postcard. A replacement mail package is sent to sample addresses when there is no response 3 weeks after mailing the initial mail package (U.S. Census Bureau, 2009).

Data collection for the GQ sample in 2006 involved personal visits by field representatives, where they administered an automated Group Quarters Facility questionnaire (for the first visit) and a bilingual paper ACS questionnaire for each sampled resident. Although face-to-face interviewing was preferred, a variety of other data collection methods were allowed. The field representative may fill in the questionnaire by telephoning the sample person; conduct an in-person interview with a proxy, such as a relative or guardian; leave the questionnaire with the sample person to complete after ascertaining that the person is physically and mentally able to do so; or leave questionnaires with the contact person for the GQ to distribute them to sample persons and collect them after they are filled in (Citro and Kalton, 2007).

Response Rates

Clearly, achieving a high response requires a substantial investment of time and effort. But many factors influence the final response rates, and often, depending on the nature of the population in the sampling frame and the types of sample selection procedures employed, the theoretical upper limit to the response rate may fall much shorter than 100 percent. As a result, it can be difficult to compare response rates across studies.

Care should be taken in ascribing too much importance to the response rates obtained in various studies. Response rates are often interpreted as an indicator of survey accuracy, hence, an indicator of potential bias in surveys (Backstrom and Hursh, 1963; Babbie, 1990; Aday, 1996; Rea and Parker, 1997). But lower response rates do not necessarily suggest higher non-response bias in surveys because non-response error is a function of both non-response rate and the differences between responders and non-responders on the statistics of interest (Keeter et al, 2000; Groves 2006; Groves and Peytcheva 2008). Lower response rates will mainly affect survey estimates if there are differences between respondents and non-respondents on the dimensions or variables of interest (Holbrook, Krosnick, and Pfent, 2008). If there are no differences between respondents and non-respondents, then it can be concluded that non-respondents are essentially a random subset of the full survey sample, or at least random with respect to the variables being measured. Naturally, it is not always possible to fully account for all factors correlated with the dimensions or variables of interest. High non-response rates may still signal some potential problems interpreting the survey results, even if the researchers cannot account for these differences in the variables they tested.

This issue has been addressed in a number of studies, including, for example, Keeter and colleagues (2000), who compared the results of a 5-day survey fielding period (response rate of 36 percent) to the results from fielding the same survey for 8 weeks (response rate 61 percent), and found no significant differences between the two surveys in the outcomes of interest. Thus, differences in the response rate may not necessarily be an accurate indicator of survey quality (Brick, Ferraro, Strickler, and Rauch 2003). Further, there is a notable lack of consistency in the way response rates have been defined and reported, making it difficult to compare response rates across surveys (Brick, Ferraro, Strickler, and Rauch 2003). As Groves and Lyberg (1988)
have noted, “there are so many ways of calculating response rates that comparisons across surveys are fraught with misinterpretations” (AAPOR, 2010).

The American Association of Public Opinion Research (AAPOR) has established standardized codes researchers can use to catalogue the dispositions of sampled cases and provided a common language and definitions for the research industry. They recommend using these codes in all technical reports so that “researchers [can] find common ground on which to compare the outcome rates for different surveys” (AAPOR, 2010). They also emphasized the importance of disclosing all elements that went into disposition codes, such as completion rate and response rate and not just selected ones, in order to fully evaluate a survey. They caution that there is no single number or measure that reflects the total quality of a sample survey (AAPOR, 2010).

Briefly, AAPOR defines response rate as the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample. AAPOR (2010) provides six definitions of response rates, ranging from the definition that yields the lowest rate to the definition that yields the highest rate, depending on how partial interviews are considered and how cases of unknown eligibility are handled. The most commonly used estimate is called Response Rate 3 (RR3), which bases its estimates on the proportion of cases of unknown eligibility that are actually eligible. The basis for the estimate must be explicitly stated and detailed. It may consist of separate estimates (e.g., Estimate 1, Estimate 2) for the sub-components of unknowns and/or a range of estimators based of differing procedures. In each case, the basis of all estimates must be indicated and the resulting response rate depends heavily on how eligible vs. ineligible cases are defined.

Similarly, the cooperation rate is defined as the proportion of all cases interviewed to all eligible units ever contacted. AAPOR provides four definitions of cooperation rates, ranging from a minimum or lowest rate, to a maximum or highest rate. It gives separate calculations for household-level and respondent-level cooperation rates, although household-level rates are most typically given, based on contact with households, including respondents, rather than contacts with respondents only. Respondent-level cooperation rates could also be calculated using only contacts with and refusals from known respondents. The most frequently used cooperation rate is Cooperation Rate 3 (COOP3), which excludes those unable to do an interview, since they are incapable of cooperating.

All of the studies reviewed, with the exception of the LA Mommy and Baby Survey, provided detailed descriptions of how they calculated response rates and the most frequently used definition was the AAPOR Response Rate definition 3 (RR3): “The response rate is the number of people who completed the survey divided by the number of people in the sample who were eligible for the survey.” However, relatively fewer of these surveys reported cooperation rates, including the American Community Survey, the 2008 Massachusetts Health Insurance Survey, the 2009 District of Columbia Health Survey, the LA Mommy and Baby Survey, and the Making Connections household survey. Yet cooperation rate can be an important and useful indicator of survey success over and above the response rate.

The CHIS is notable for its in-depth explanation of how response rates were calculated, even if they did not apply standard AAPOR definitions. CHIS 2003 through 2007 employed the screener completion rate, defined as the proportion of screeners administered to a household and an adult randomly selecting for interviewing compared to all screeners attempted, and the extended interview completion rate, which was defined as the proportion of sampled individuals who completed the full interview. For CHIS 2003, the extended interview completion rate was 60.0 percent for the adult interview, 81.4 percent for child interview, and 57.3 percent for
adolescent interview. Multiplying the screener and extended rates yielded an overall response rate of 33.5 percent. CHIS 2005 arrived at an overall response rate of 29.5 percent while CHIS 2007, using the same formula, reported an overall landline sample response rate of 21.1 percent and cell phone sample response rate of 52.0 percent.

The 2005 and 2007 LACHS studies used the standard, accepted AAPOR definitions for both the cooperation rates (Cooperation Rate 3, COOP3) and the response rates (Response Rate 3, RR3). They reported response rates of 22.8 percent and 18 percent, respectively, for the adult interviews, which are lower than most of the RDD surveys and some of the ABS surveys. The rates are substantially lower than the LA-FANS constructed list approach. In fact, the LA-FANS response rates compare favorably with many ABS approaches using the standard USPS DSF as the sample frame. The LA WIC study used the AAPOR Response Rate 2 (RR2) for calculating the overall response rate and reported a rate of 66.5 percent. Finally, the Making Connections household survey did not report how their response rates were calculated, and they did not provide information on cooperation rates.

A major challenge in RDD- and AB-samples is in determining the number of people in the sample who were eligible for the survey (Brick et al, 2002; Groves, 2006). While eligibility for the survey can be determined for many sample members in both RDD- and AB-samples, there are parts of both samples for which eligibility cannot be determined directly. For example, in an RDD-sample, some telephone numbers are never answered, while in an AB-sample with address-correction requested, some addresses that do not respond to the survey may be vacant or second homes (Brick, Ferraro, Stickler and Rauch, 2002).

One method to allocate the unknown eligibility group in RDD-samples is to use the “survival method” developed by Brick, Montaquila and Scheuren (2002). This method assigns sample members who fall within the AAPOR Response Rate category three (RR3; eligibility is unknown) to eligible and ineligible status. This method has been used in a number of RDD surveys, including the AHRQ Consumer Assessment of Healthcare Providers and Systems (CAHPS), the California Health Interview Survey, and the National Survey of America’s Families. Brick, Montaquila and Scheuren (2002) reported that this method reduced the effective response rate for the RDD component of the 2008 National Health Interview Survey (NHIS) from 49 percent to 42 percent, suggesting that surveys not using this adjustment may be overestimating their response rates.

Factors Associated with Higher Response Rates. Due to considerable pressure at the National Center for Education Statistics (NCES) to find an alternative to RDD interviewing, because of falling response rates, a pilot study was conducted for the National Household Education Survey (NHES). It is instructive to look at this pilot study that tested a dual-frame ABS/RDD survey with different screeners, mail and telephone surveys, and incentives, and its effect on response rates.

The 2009 pilot study explored the use of address-based sampling as an alternative to land-line RDD in the NHES (Montaquila, Brick, Hagedorn, and Williams, 2010; Brick, Williams, and Montaquila, forthcoming). There was particular interest in addressing the particular needs of what they call the “linguistically isolated” population, i.e., those households where no household members over the age of 14 speak English well, according to their self-report (U.S. Census Bureau, 2003). The 2000 decennial census reported 21 million (8.1 percent) individuals nationally who were considered to be “linguistically isolated” (U.S. Census Bureau 2003).

Screeners were mailed out in the first stage to find and match the school-age population, achieving an overall nationally representative response rate of 59 percent, which is slightly
higher than RDD screening methods. The response rate to a subsequent topical questionnaire was 75 percent, again higher than RDD interview response rates. The mail survey also identified greater numbers of children.

Three different screeners were tested: the screen-out, core, and engaged versions. While the screen-out and engaged versions were in both English and Spanish, the core was only in English. Although this may have contributed to differences in response rates, the effect appears marginal and only occurred for the screen-out version. The screen-out consisted of a single page asking for the names, age, and gender of only the children in the household. Pictures of children on the cover sheet were intended to encourage the respondents. The four-page core screener added an additional 9 questions about the household and its members. Finally, the six-page engaging screener included the core questions with 16 additional questions, designed to encourage rapport by asking for respondent perceptions of education.

Overall response rates for the screen-out in English and Spanish were 63.6 and 80.7 percent and similar rates were obtained for the engaging screener with 60.1 and 71.9 percent for English and Spanish versions, respectively. These rates were higher than those for the core screener, which was only presented in English thus making comparisons difficult, at 58.3 for all speakers and 70.3 percent for bilingual Spanish respondents. With a sample of 10,200, the screen-out and engaging screeners out-performed the core screener, but there were few differences in the characteristics of respondents who completed all screeners compared with the larger population. For the entire sample, the percentages of families with young babies and respondents 20 years old or younger were somewhat lower than those found in the American Community Survey (ACS).

Each screener included a $2 incentive and was followed by a second follow-up either by mail or phone. Phone numbers were retroactively matched to street addresses for this purpose, achieving a nationally representative match rate of 57 percent. The pilot study found that mailings out-performed land-line phones, but this was mainly because 19 percent of the address-matched phones were no longer in service when a contact with the household was attempted. Only 17 percent of the topical questionnaire non-response cases sent to telephone operators for follow-up was completed. However, there was also some evidence that follow-up phone calls encouraged screener responses.

Once the screener had been returned, households with school-age children were mailed the topical questionnaire. The study tested a variety of levels of randomized incentives, including $0, $5 and $15. They found that response rates increased with the size of incentives: 72 percent response rate for $0 incentives, 77 percent for $5 response rates, and 83 percent for $15 response rates. However, they found no differences in household characteristics by incentive level.

In looking at the census tracts containing large proportions of linguistically isolated households, language was still a key barrier to participation, in addition to literacy. Although address-matched mailings performed better than RDD interviews overall, there were significant challenges faced by mail surveys in the face of language and literacy difficulties. The main advantage of mailings over RDD was that mailings brought into the sample the 25 percent of households with cell phones only.

When comparing the costs between RDD and ABS, the research team found that address-matched mailings, sent out all at once, were approximately 25 percent cheaper than RDD. However, mailings still have problems, including the inapplicability of selected sections of the
questionnaire in a given household (in this study as much as 60 percent of the questionnaire) and lack of control for sampling children within households.

The study authors concluded that, overall, screening for households with children by mail in a two-phase approach was successful. Mail follow-ups tended to bring in respondents with lower socioeconomic households, while landline phone follow-ups did not. However, they also conclude that language and literacy issues are particularly salient when using mail (Montaquila, Brick, Hagedorn, and Williams, 2010; Brick, Williams, and Montaquila, forthcoming).

The study findings also highlight the relationship between incentives and response rates, with relatively small increases in the incentive producing ordered, linear increases in response rates. While it is generally accepted that cash incentives are more effective than gifts of equal value (Singer 2002), the 2008 Massachusetts Health Insurance Survey provided an option of winning Red Sox tickets through a lottery, and this seemed to help the response rate, given the extreme popularity of the team in the state (Mass Health Insurance Survey, 2008).

Finding Hard-to-Reach Populations

An important challenge for the F5LA Family Survey is how to adequately represent “hard-to-reach” populations. Hard-to-reach may be defined as those in a population that are “hidden” by virtue of not being available for sampling using either RDD or ABS frames. In general, no sample frames exist for the hard-to-reach, making it difficult to estimate the size or boundaries of this group. Members in these groups are often stigmatized or engaged in illegal behavior (Heckathorn, 1997).

“Hard-to-reach,” as defined within the context of the Family Survey goals, may not include some of the populations that are typically included in this category, such as drug users, HIV positive individuals, and sex workers. Instead, the hard-to-reach for the purpose of the Family Survey could include homeless and indigent individuals, and/or migrant and seasonal workers who move frequently. This list might also include those identified by the NHES researchers as “linguistically isolated,” as defined earlier. In California, 39.5 percent of people spoke a language other than English at home, with 20.0 percent classified as linguistically isolated—the highest among all 50 states in the United States (Lee, Nguyen, Jawad, and Kurata, 2008). In LA County, households in some communities or small areas may include linguistically isolated populations who speak a variety of languages including Chinese, Vietnamese, Korean, Khmer, and Tagalog, according to the 2000 decennial census (U.S. Census Bureau, 2003).

Methods for reaching “hard-to-reach” populations are classified as either extensive (quantitative) or intensive (qualitative). The choice of which method to use appears to be tied to the specific topic of interest and the degree to which the topic of interest is important in the larger survey. In a general population survey, the precision of the sample may be improved by locating and including the hard-to-reach populations defined above (homeless, mobile, migrant), ideally using either a sample frame or other form of bias adjustment.

In surveys or studies of specific populations where the goal is to understand these groups, such as intravenous drug users, the phenomenon under study affects one’s ability to locate them. In these cases, more intensive methods would be appropriate, with the most frequently used method being the snowball sampling approach, although key informant and targeting sampling strategies are also available (Heckathorn, 1997). Snowball sampling is a non-probability method that relies on referrals from initial survey participants to generate additional participants. Since this method is a nonrandom approach, it is important to understand that the results cannot be generalized to the universe of these populations. These methods often increase bias, since
those who are reached are those most cooperative and those identified tend to be those with larger social networks (Heckathorn, 1997). Some associates of those sampled are intentionally masked and kept from being identified.

Targeted sampling approaches offer some possibilities for including hard-to-reach populations in a survey sample because they utilize sites or places where these potential subjects are most likely to be situated. In this method, locations where the hard-to-reach population under study congregates are initially mapped using ethnographic methods. Then, researchers recruit a pre-specified number of participants at each location for interviews, ensuring that subjects from different areas and subgroups will appear in the final sample. However, the accuracy of targeted sampling depends on the accuracy and comprehensiveness of the ethnographic mapping and the identification of all potential locations for the population under question. This is also very time-consuming and labor-intensive work and may not be suitable for timely twice-a-year surveys.

The main issue with these intensive approaches is how to draw a random sample. All of the approaches described tend to produce biased samples. Moreover, it is difficult to draw a sample that is representative of the larger population of hard-to-reach subjects under study, and statistical error can not always be properly calculated.

To meet the goals of the Family Survey, it may be possible to use an approach that includes group quarters in addition to housing units, as defined and used in the American Community Survey. This may collect many of the most indigent and mobile individuals, particularly in rural areas. At the same time, it may be possible to use sub-area sampling frames where such individuals might be identified, such as hospital databases for emergency hospital cases, WIC participants, welfare cases, or neonatal intensive care units (NICU’s) in hospital for high-risk births. Using such frames will need to be coordinated with the larger overall sampling frame, in order to identify and remove duplicates. Such an effort may be very time-consuming and potentially difficult because the assumption is that these local databases are accurate and reliable. Nevertheless this may be an option for locating hard-to-reach populations for the Family Survey (e.g. crossing a household sample frame consisting of urban and rural addresses for housing units and group quarters with local WIC enrollment lists).

Respondent-Driven Sampling

A promising method for intensively sampling hard-to-reach populations called Respondent Driven Sampling (RDS) was initially described in a seminal article published in 1997 by Dr. Douglas Heckathorn. Over the past decade, respondent driven sampling (RDS) has become recognized as a viable option for rigorous sampling of hard-to-reach populations. It is the sampling method of choice for studying populations at risk for HIV around the world (Johnston and Sabin, 2010). The advantages to RDS are that it offers “a generally effective and efficient means” to gather a probability sample from underserved and understudied populations, costs are reasonable, survey staff work from fixed sites where their physical security can be protected, and results may be generalizable to the larger population.

RDS is based on Markov chains and a form of social network analysis called “biased networks theory” to develop a sample from a hidden population. It resolves many of the sources of bias listed earlier for snowball sampling and other chain-referral approaches and is not dependent on the initial subjects identified at the start of the sampling method. RDS is based on the notion of an incentive system, both material (e.g. money) and symbolic (e.g. helping others), in which a primary reward to the individual—the reward for being interviewed—is reinforced by a secondary incentive—a reward for recruiting others into the study. Thus, as found in other chain-driven sampling methods, a few initial subjects each produced chain-referral systems that
will yield a large number of recruits into the survey after several successive waves. Heckathorn (1997) showed how a single subject generated more than one hundred recruits of diverse race, ethnicity, gender and place of residence. Further, because subjects already recruited into the sample are asked to invite others to participate, it reduces masking bias associated with having subjects “name names” of others. They are not asked to give names to the investigator but instead the respondents doing the recruiting have the opportunity to allow the potential recruits to decide for themselves whether they want to participate. Finally, by using recruitment quotas for each recruiter, it reduces the bias that might occur from individuals with larger personal networks being more likely to be in the study.

Sample recruitment occurs over successive waves as more subjects enter the sample. The initial subjects, called the “seeds,” represent the first wave and recruit a new group of recruits for the second wave, who then go on to recruit another new group of recruits in the third wave, and so on. Typically 10 to 20 waves of subjects are required until equilibrium is reached, which is defined as the point at which the sample distribution on key variables remains stable within 2 percent of the equilibrium distribution, even though more individuals enter into the sample (Heckathorn, 1997; 2002). When this happens, the characteristics of the final sample are independent of those characteristics of the initial seeds, along such factors as ethnicity or gender or other characteristics of interest among the hard-to-reach population. When equilibrium is reached then the sample is representative of the larger hard-to-reach population (Salganik, 2006).

The characteristics of individuals who enter the RDS sample should correspond to the theoretical distribution in the larger population. However, if sampling does not continue until equilibrium is reached, then it is likely that the final sample will more closely reflect the characteristics of those subjects initially recruited into the sample; hence, it will be biased. An important feature of RDS is that equilibrium is reached at a very rapid, geometric rate that can be calculated. Fewer waves of sampling are required if the initial seeds varied with respect to key characteristics, e.g. ethnicity, gender, location, etc. Recruiting a higher number of waves allows researchers to penetrate the networks of target hard-to-reach populations more deeply. The greater the RDS sample reflects the networks of the hard-to-reach population, the greater the sample meets several theoretical assumptions indicative of representativeness (Heckathorn, 1997; 2002; Salganik & Heckathorn, 2004).

Done properly, an RDS strategy allows for a close approximation of covering the underlying population by modeling known sources of bias, such as bias from a tendency towards in-group affiliation (recruiting members of one’s own ethnic group) or out-group affiliation (recruiting those outside of one’s own ethnic group). If these biases do not occur, then it may be concluded that every subject is selected randomly from the hard-to-reach population, independent of group membership. That is, the probability of selecting a member of any sub-group of the hard-to-reach is equal to the proportion of that sub-group’s members in the larger system (Heckathorn, 1997). If this occurs, then it may be concluded that the RDS sample is unbiased. Even in the worst-case scenario where in-group affiliation (in-breeding) is greatest, according to simulations conducted by Heckathorn (1997), the association between the sample and the population remains positive and strong, with a mean discrepancy of 12.1 percent plus or minus 8 percent. Still, if the protocols for using RDS are not rigidly followed and carefully analyzed, then the sample becomes more of a convenience, snowball sample with all the attendant biases produced from this method. In fact, Johnston et al (2008) suggested that many studies claiming to use RDS ignore strict adherence to these requirements and become snowballs samples.

Given the hidden nature of hard-to-reach populations, the bias of an RDS sample cannot be empirically assessed by comparing it to the population from which it was drawn. It may be
possible to compare the demographic characteristics of the RDS sample to the characteristics of the larger community from which it was drawn, but this assumes that the hidden population mirrors the larger community. Even better would be a comparison between the RDS sample and a subset of the larger community that would likely be most comparable, such as the percent living below the poverty line.

This type of comparison is particularly applicable in the Family Survey sample because there are specific target neighborhoods from which the RDS recruitment can occur and where the numbers of certain characteristics, such as the percent living below the poverty line, are known. By using separate RDS recruitment methods within each Best Start site, the RDS sample characteristics can be matched to the specific characteristics of the community to determine representativeness. Thus, the heterogeneity of the target population is at least mirrored by the sample (Spreen and Zwaagstra, 1994). Put another way, if the degree of inbreeding bias can be independently assessed (not drawn from the RDS sample) then it may be possible to weight the sample to eliminate these biases. Further, if the consistency of RDS samples can be assessed by comparing the variability between the sample and the larger hard-to-reach population along a set of known factors it may be possible to compute standard errors and confidence intervals for these population estimators. For the hard-to-reach populations of most interest in the Family Survey, these techniques are feasible by using extant administrative data related to homelessness and other characteristics of the hard-to-reach population of most concern.

There are also some limitations and cautions to the RDS approach. RDS will only work if the characteristic that identifies individuals as members of the population under study is also one that requires them to create connections with other members of the population, for example, through exchanges with other drug users or shared use of homeless shelters. Additionally, the geographic boundaries of the subjects extend only as far as their migration or travel and may depend on the availability of transportation used by respondents. This has implications for finding members of the population in non-urban areas where public transportation is non-existent or difficult.

A caveat of using this approach is that methods of verifying identification of new recruits is necessary to prevent impersonation and to reduce the possibility that the same already recruited individuals obtain more of the incentives. As Johnston and Sabin (2010) note, RDS does not work in every setting with every population and it is important to conduct some formative explorations of the population of interest to ensure that RDS is the correct method given the target population and the geographic, cultural and social settings. To help determine whether RDS is appropriate to the Family Survey, there are several available tools that should be consulted (Johnston et al, 2010; University of California, San Francisco, 2009; Johnston, 2008; Johnston, Sabin & Prybylski, 2010). A set of key questions, derived from the caveats listed by Heckathorn (1997), will determine the potential feasibility of an RDS approach for the Family Survey. They are as follows:

1. Is the population of interest socially networked? If the answer is yes, then continue, but if no then RDS is not appropriate.
2. Does the population have social network sizes sufficient to maintain recruitment? That is, does the average population member have at least 3-5 peers, who would meet study eligibility criteria, who they could refer to the survey? If the answer is yes, then continue, but if no then RDS is not appropriate.
3. Does the social network consist of one social network component (that is, the members come from a single common group)? If yes, conduct further formative
research to determine if RDS is acceptable to the population. If no, then go to the next question (Q4).

4. If the population does not consist of one social network component, are there some sub-network members (members of a sub-group of the larger network) that form ties with other sub-networks members? If yes, conduct further formative research to determine if RDS is acceptable to the population and ensure that those who have connections across sub-networks are selected as seeds.

Johnston and Sabin (2010) emphasize “RDS is not a magic method to easily recruit previously understudied and inaccessible populations.” While RDS offers the possibility of producing a representative sample, and its use attests to the ease with which it can be implemented, it is highly sensitive to deviations from the rigorous requirements for its use. When assumptions and analysis requirements are not met, the generalizability of a sample can be jeopardized. According to Johnston and Sabin (2010), in numerous RDS surveys reported in the literature, social network size was inadvertently not collected or recruiter-recruit connections were not tracked, which would produce highly inaccurate estimates.

**Issues for Surveys Involving Linguistically and Culturally Diverse Populations**

Most of the studies reviewed included languages beyond English and Spanish. Surveys in California included the largest diversity of language groups including six languages for the CHIS studies (English, Spanish, Mandarin, Cantonese, Vietnamese and Korean) and the ACS (English, Spanish, Chinese, Russian, Korean and Vietnamese). Most of the surveys based in LA County included translations of survey materials into those languages most common to the county population, including English, Spanish, Cantonese, Mandarin, Korean, and Vietnamese, with the 2005 LACHS surveys adding Armenian. LAMB had translators available in languages other than the English, Spanish and Chinese languages (both Cantonese and Mandarin). Contrary to this trend, LA-FANS only used English and Spanish.

The languages in which survey interviews are translated may determine who is eligible to participate and influence the survey estimates. Using data from the 2003 California Health Interview Survey (CHIS), Lee, Nguyen, Jawad and Kurata (2008) studied potential biases arising from the hypothetical exclusion of linguistic minorities. They looked at differences in survey estimates between English and non-English respondents as a result of excluding linguistic minorities. There were disparities between the English and non-English respondents on a set of health indicators and the magnitudes of the differences were surprisingly large. More importantly, although the proportion of persons unable to carry out survey interviews in English was not large, differences on the health indicators were found to be considerable. Excluding linguistic minorities potentially overstated better health status and understated problems with health care access. When survey interviews were conducted in English and Spanish, the overall differences in the health estimates disappeared, but for interviews conducted among Asian minorities, the differences in the estimates for Asians remained unchanged. They conclude that surveys conducted only in English may not adequately represent the general population and differences in estimates for Asians may exist if only English and Spanish are used (Lee et al, 2008).

Another important potential effect of language translation was also suggested by this study; that interview language may contribute to measurement error differentially. For example, interviewing style and respondents’ behaviors may differ systematically by interview language and cultural background. As Lee et al. (2008) point out, “it is not guaranteed that translated questions measure the same concept with the same scale.”
When including a variety of cultural groups in the population of interest, it is sometimes important to consider that response rates to different modes of data collection might differ by cultural group. For example, various group-specific cultural norms regarding the sharing of information considered personal might require different methods for collecting the data. In fact, there may be important cultural differences during the recruitment stages for members of different minorities.

A study using data from birth certificates and survey administrative data from a large-scale statewide experiment examined differences in survey response by race and Hispanic origin (Nam, Mason, Kim, Clancy, and Sherraden, 2011). The sample consisted of mothers of infants selected from Oklahoma birth certificates using a stratified random sampling method (N=7,111). Using probit analysis, the study looked at the probability of survey respondents being located by the survey team and, separately, the probability that they would complete the survey questionnaire. The results indicated that African Americans, Native Americans, and Hispanics were significantly less likely to be located during the study recruitment than Whites, controlling for other demographic, socioeconomic, and environmental factors. However, once located, there were no differences among the four groups in the probability of completing the survey questionnaire. Nam et al. (2011) suggest that researchers should pay attention to separate stages of respondent recruitment and improve strategies to locate members of racial and ethnic minority groups during recruitment.

There is evidence from the Census 2000 Supplementary Survey that response rates may differ by the three modes of data collection: mail, CATI, and CAPI (Citro and Kalton, 2007). Those responding by mail were more likely to be white, be headed by an older person, and own their own homes. The response rates after mail contacts were highest in predominantly white census tracts and lowest in census tracts that were predominantly Native Americans and Alaskan.

Response rates to the initial mail surveys were also lower in census tracts that were predominantly Hispanic or black. At the same time, response rates were highest in these groups when CAPI was conducted to follow up those who did not respond to the initial mail contact. For census tracts that were predominantly Hispanic, the efforts of CAPI resulted in a response rate that did not substantially differ from the overall rate (95.8 vs. 95.4 percent) suggesting the importance of conducting CAPI follow-up in census tracts with predominantly Hispanic samples. Even with the CAPI follow-up, the overall response rates for census tracts that were predominantly Native American/Alaskan or African-American were lower than for census tracts with predominantly white respondents. From an overall response rate of 95.4 percent for all addresses, response rates for census tracts that were predominantly American Indian/Alaskan or African-American were 89.1 and 92.4 percent, respectively.

One of the important lessons for the F5LA survey from these response rate differences is that any household survey based on address lists should also employ the more labor intensive in-person interviewing for non-responders to the initial mail or telephone strategies, particularly for blocks that have large Hispanic, African-American, or Native American/Alaskan populations. Further, it would be important to understand differences in how survey items translated into other languages are understood by members of those language groups. To this end, F5LA recently initiated a project to develop and test a measure of social support using up to eight languages, to equate the test across languages. It is expected that this measure would be included in the Family Survey.
Summary and Recommendations

In this section, key issues to emerge from the literature review are summarized and two designs are recommended. First, the themes that emerged from the literature review and their implications for the Family Survey will be discussed according to each of the key steps in designing a household survey: sample frame, sample strategy and size, and data collection mode and procedures. This will be followed by the recommendations, summarized in Table 2.

Sample Frame

Several sample frames were used most frequently in the studies reviewed: Random Digit Dialing (RDD), Address-Based Sampling (ABS), dual-frame RDD plus ABS, and other forms of constructed lists. The highest response rates in the studies reviewed tended to come from those that used either RDD or ABS, but not both. Given the ways in which response rates were calculated, caution is needed when interpreting these differences. Despite higher costs, the two studies that used a combined RDD and ABS frame had much lower response rates than the studies that used one or the other.

Recently there has been much more attention paid to cell phone users and, in particular, households where cell phones are the only telephones used. Many of the most recent and well-done RDD surveys, such as CHIS, now include a cell phone sample frame that appears to mitigate or eliminate an important source of coverage bias. Although controversial, it may be possible to adjust for any remaining bias through a weighting procedure that better aligns the sample with the population. If an RDD sample frame is used with coverage for cell phone users, a choice should be made as to whether the frame is overlapping, which includes both landline and cell phone users and excludes duplicates without pre-screening, or non-overlapping, in which cell phone-only households are first screened and those with only cell phones are added to the landline users in the frame.

If an ABS approach is employed, the need to continuously update and verify the USPS Delivery Sequence File of addresses must be considered. There are a number of advantages and disadvantages to working with this existing address-based list as opposed to constructing a list specific to the survey, as LA-FANS did. Continuous updating of the address lists, as the ACS does, will be necessary to draw samples twice per year, as per the original plan for the Family Survey. The need for timeliness in data collection may conflict with the time and expense required to maintain updated lists. It may not be possible to complete the list construction and verification before the sample must be drawn in order to field the survey within a given six-month window. One option is to consider the ACS approach, in which sampling is done on a continuing and rolling basis, which may allow time for the lists to be as up to date as possible before they are used for a given sample draw. Another consideration is whether to include group quarters and, if so, how they are added to the frame and updated so as to reduce, rather than enhance, coverage bias.

To address the issue of “hidden” or hard-to-reach populations in the sample frame, the respondent-driven sampling (RDS) approach appears promising. But further study is needed to ensure that the populations defined as hard-to-reach meet the criteria for using the RDS approach. Specifically, the social networks of members in a given hard-to-reach population must be studied to ensure they meet the criteria that members are socially networked in some way (e.g. they all show up at soup kitchens or food distribution sites).
Sample Strategy and Size

The Family Survey will need to obtain a sample large enough to provide good precision of the population estimates, particularly at the largest geographic level, such as the census tract or the Best Start community or neighborhood, however that is defined. At the same time the importance of the design effect, that is, the degree to which the sample varies from a simple random sample, should be considered because this can make the estimates for sub-populations with smaller sample sizes less precise (if the design effect is large relative to the sample size). Finally, there are considerably more marginal costs involved in adding macro units (e.g. areas) to the survey than there are in adding more cases to the smaller analytic units (such as households).

The studies reviewed all grappled with these issues. They consistently selected as many macro units or clusters as possible in order to ensure that precise estimates were available for smaller geographic areas (e.g. census tracts). In general, the adult samples for state or county areas were on the order of tens of thousands of cases. The study that spent more time and effort constructing its own list-based frame needed to sample fewer households than other surveys. The two dual-frame RDD/ABS studies also required somewhat fewer individuals or households.

The issue of migration into and out of the Best Start communities should be taken into account in the semi-annual sample design, since F5LA will want to identify potential changes in community composition. However, at this point, very little is known about the nature and size of this issue, as more conclusive data are still being collected.

Most of the samples in the studies reviewed used a multi-stage approach, such as sampling census tracts, then households within census tracts and finally individuals within households. The largest macro unit of the sample (e.g. census tract) typically employed the probability proportional to size (PPS) strategy in order to take into account the effects of different population densities on the probabilities of selection. Many of the studies oversampled for specific sub-populations, such as levels of poverty within the census tracts or ethnic and/or language groups that have a lower “natural” probability of occurrence within a given population. This is also one way of taking into account areas with a concentration of linguistically isolated households, defined as households where individuals report speaking English less than very well (U. S. Census Bureau, 2003). Another approach, used in CHIS 2007 was to oversample geographically targeted concentrations of certain ethnic groups, in this case Koreans and Vietnamese, by supplementing the RDD with listed telephone directories from these geographic areas using, group-specific surnames. This was designed to increase the precision of their estimates for these ethnic groups (California Health Interview Survey, 2009b). In the Family Survey it would be important to calculate the power of the sample in the neighborhood clusters or sub-group concentrations that are most salient, such as areas of poverty, linguistically isolated households or ethnic groups.

Data Collection Mode and Procedures

The literature suggests that the use of interviewers who come from the same communities as those in the survey, with similar backgrounds and characteristics to respondents, offers no strong benefits—and there may be some disadvantages. While employing indigenous interviewers may help improve the participatory research process, it may come at a cost of lower data quality and response rates. The literature indicates that interviewers who are well-trained and experienced tend to obtain the highest data quality and response rates.
Computer-assisted interview technologies are very well-developed and, in combination with a computerized survey tracking database, offer the best methods for ensuring standardized data collection and quality control. This helps maximize response rates and data quality. For RDD designs, a seasoned telephone research center using up-to-date computer-assisted telephone interview (CATI) technology should be employed. For ABS designs, in-person interviews using computer-assisted personal interview (CAPI) is the gold standard, but costs can include outfitting interviewers with laptop computers, as well as maintenance of this equipment.

Many of the studies reviewed used mixed methods of data collection. The primary data collection method typically involved telephone or mail, with more intensive methods reserved for sampled respondents who do not respond to the initial inquiry. To improve precision and data quality, the data collection mode must be considered within the larger context of the survey sample frame and sampling plan. Mixed methods may be more costly but have the potential for improving response rates, particularly for key population sub-groups such as cultural and language minorities.

However, in studies reviewed, the use of different modes did not always improve the response rates. While the high ACS response rate could be attributed to the mixed modes of data collection, it could also be due to the methods in place to find non-responders and the diligence of the field staff. The CHIS surveys show good response rates despite using only telephone interviews after an initial telephone screening, while response rates for the LACHS surveys using telephone interviews were much lower. Surveys that employed dual-frame RDD and ABS approaches, offering telephone, web and mail options did not necessarily show higher response rates; indeed, they appear lower than other surveys. Caution must be used when interpreting response rates across surveys when their uses and definitions may not necessarily be consistent.

Many of the studies reviewed did not use any incentives. The pilot study for one national survey, NHES, found that higher incentives led to higher response rates. But while incentives may lead to higher response rates, it is not clear from this review that this is always the case. The use of higher incentives for cell phone users is justified on the basis of compensating respondents for their air time, and probably will continue to occur for surveys where cell phone users make up an increasingly important segment of RDD surveys.

**Recommendations**

Table 2 summarizes the key findings from the review of survey design literature considering key factors such as population coverage, response rate, cost and sample efficiency. Each factor was given a qualitative rating (e.g., low, moderate, high, etc.) consistent with the main findings from the literature review. A color-coding scheme is provided to highlight the advantages (green) and disadvantages (red) of each main survey design approach.
Based on this information, two designs are recommended, each at different cost points balanced against methodological considerations. A third recommended approach is a variation of one of the two main designs. The lower cost approach recommended is the telephone-based Random Digit Dialing (RDD) with cell phone coverage. The higher cost approach is the Address-Based Sampling (ABS) approach, using the USPS Delivery Sequence File purchased through a vendor, with sampling proportional to size within specific census tracts or communities. An even higher-cost option that is recommended is the use of a multistage area sample using a constructed list of households within selected areas (e.g. census tracts) based on existing addresses combined with local field operations, which should improve coverage and response rates.

In addition to cost, both recommended designs have advantages and disadvantages from a methodological perspective. Although a potential weakness of the RDD design is in the coverage of households without landlines, the inclusion of cell phone coverage has been shown to reduce this problem, and careful screening and recruitment by well-trained telephone operators can mitigate this issue.

### Table 2. Summary of survey designs by key factors, with recommendations

<table>
<thead>
<tr>
<th>Design</th>
<th>Population Coverage</th>
<th>Response Rate</th>
<th>Cost</th>
<th>Sample Efficiency</th>
<th>Verdict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Digit Dialing (RDD) w/cell phone</td>
<td>Good</td>
<td>High (CHIS, Moderate)</td>
<td>Low to Moderate</td>
<td>Low</td>
<td>Recommended</td>
</tr>
<tr>
<td>Address-Based Sample (ABS) using USPS DSF</td>
<td>Excellent</td>
<td>High</td>
<td>High</td>
<td>Moderate</td>
<td>Recommended</td>
</tr>
<tr>
<td>Multistage Area Sampling using Constructed Lists</td>
<td>Excellent</td>
<td>High</td>
<td>High</td>
<td>Moderate to High (LA-FANS, High)</td>
<td>Recommended</td>
</tr>
<tr>
<td>RDD plus ABS</td>
<td>Excellent</td>
<td>Low</td>
<td>Moderate</td>
<td>Low</td>
<td>Not Recommended</td>
</tr>
<tr>
<td>Respondent-Driven Sampling (RDS) 7</td>
<td>Excellent</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Needs Further Review</td>
</tr>
</tbody>
</table>

Color- Coding: Green = advantage, Red = disadvantage, Clear = neither advantage nor disadvantage

---

3 Population Coverage is defined as how well the sample frame covers all members of the population. Ratings are relative to the range of rates among the studies reviewed: Fair, Good, and Excellent.

4 Response rate is based on definitions used by study researchers so they may not all be consistent. Household or Parent rates were used and qualitative ratings of Low, Moderate, or High are relative to the range of rates among the studies reviewed (shown in the Appendix Table). High = 60 percent or higher, Moderate = 30 to 60 percent, Low = below 30 percent.

5 Cost is defined as the cost per completed interview and includes total costs of sample planning, sample frame development, sample selection, and data collection operations, including multi-mode strategies. Costs do not include instrument development or data analyses, which are considered relatively constant. Ratings are based on rough estimates of overall costs of each method in the literature, with relative ratings of High, Moderate, and Low assigned.

6 Sample efficiency is defined as the amount of the sample frame that is required to be contacted to obtain the initial sample size. High efficiency occurs when fewer of the sample frame yields the required sample size. Factors affecting efficiency include sample frame construction, screening and cooperation rates.

7 High ratings assume that the RDS approach is appropriate to the Family Survey and that best practices in RDS are followed. If any assumptions underlying RDS are not met (see Johnson and Sabin, 2010) then the rating for any factor would be low.
interviews should also help. Weighting procedures to adjust for any remaining coverage bias may also be effective, although this strategy is controversial and may not always work. Further, the feasibility of RDS (Respondent Driven Sampling) as an adjunct to RDD should be considered. The increased costs might be marginal relative to the increased coverage of hidden or hard-to-reach populations. However, further investigation is required to ensure that all assumptions inherent in the RDS approach are met.

In RDD surveys, sample efficiency tends to be relatively low because of the need to call a large amount of telephone numbers for a given sample yield. The use of a vendor with existing, updated telephone lists should improve efficiency. Another advantage of a telephone-based survey is the ability to use a variety of languages without requiring extensive scheduling of in-person interviews. To improve coverage, particularly for concentrations of linguistic minorities and hard-to-reach populations, the RDD approach could be supplemented by local telephone directory lists for targeted geographic areas. With high response rates and relatively low costs, an RDD design with cell phone coverage is recommended. An RDD design may also fit more readily into a semi-annual data collection because the effort required is less time-consuming than is an ABS approach and does not require extensive updating or field effort. Screening and data collection can be done largely by telephone, although there may still be a need for in-person field work for non-responders or geographic concentrations of hard-to-reach subgroups.

The second recommended design is the ABS approach, using the US Postal Service DSF (Delivery Sequence File). ABS provides excellent coverage and can yield high response rates, but at higher cost. There are vendors through which updated address lists can be obtained. Given the semi-annual nature of the Family Survey, the amount of time required for updating and screening households within selected census tracts may require a rolling sampling schedule rather than keeping to a discrete semi-annual data collection timeline. Consideration should be given to the procedures used by the American Community Survey (ACS) in which the sample frame is split into specific months within a given year. Later months of the sample frame may undergo updating at the same time that samples are drawn from earlier months, and data collection for those earlier months, proceeds.

The ABS design should include a multi-mode data collection strategy, beginning with mail and/or web-based surveys followed by telephone interviews for non-responders (or a sample of non-responders) and finally in-person interviews particularly focused on the hard-to-reach and linguistic minority groups in a given census tract. Costs would primarily be driven by list development, updating and verification working with an approved vendor as well as multiple modes of data collection to improve response rates.

A related design with higher costs is the use of a multistage area sample in which a list of households is generated for each Best Start community based on extensive field operations to locate households in sampled census tracts. The first stage would sample census tracts proportional to size (PPS) within each community, and then all households in the sampled census tracts would be identified using field-based verification and screening. This approach is costly and requires a considerable amount of time to arrive at the final sample of households, but can produce a more accurate and updated sample with excellent coverage and good sample efficiency.

Finally, regardless of the survey design selected, it would be important to revisit the notion that the two yearly samples drawn from the sample frame should be separate and independent. While it may be intuitively obvious that each sample would have greater precision if drawn independently, this may not allow for the detection of changes in the communities over time.
Consideration should be given to an overlapping sample design in which each sample selected from the sample frame is not independent.

In conclusion, it is hoped that this examination of best practices in household survey design from the literature will assist F5LA in designing the Family Survey. This paper identified the issues involved in the design of a household survey, from developing the sample frame to sampling and collection of the data, based on best practices in the literature as well as lessons learned. The paper concluded with several recommended survey designs, and further discussion between F5LA staff, the RAC, and the Harder+Company team will be needed to choose the best approach that balances costs with methodological rigor.
References


Applying GIS to residents’ perceptions of their neighborhoods. Washington, DC: The Urban Institute.


Appendix A: Summary of Surveys

Appendix A includes a summary of surveys that were reviewed for this paper. First, a table that briefly outlines the main components of selected surveys is presented. The surveys are presented in the order of their methodological rigor and relevance to informing the design of First 5LA Family Survey.
<table>
<thead>
<tr>
<th>Survey</th>
<th>Sample Frame</th>
<th>Sampling Strategy</th>
<th>FC*</th>
<th>RC*</th>
<th>LI*</th>
<th>Oversampled Population</th>
<th>Areas Where Estimates are Available*</th>
<th>Mode of Data Collection</th>
<th>Incentives*</th>
<th>Language</th>
<th>Sample Size</th>
<th>Cooperation rate</th>
<th>Response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Community Survey</td>
<td>Address-Based: Census Bureau’s Master Address File (MAF)</td>
<td>Sampling proportional to size using housing density and five year blocks.</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Areas with small populations</td>
<td>Census sub-blocks</td>
<td>Mail, Telephone, and In-person</td>
<td>None</td>
<td>Mailed surveys - English and Spanish; Telephone interviews - English, Spanish, Chinese, Russian, Korean, and Vietnamese</td>
<td>Approximat etly 1.9 million households; 145,000 people in group quarters facilities</td>
<td>Unknown</td>
<td>97%</td>
</tr>
<tr>
<td>2003 California Health</td>
<td>RDD: All eligible households with landline telephones</td>
<td>Stratified sampling; Strata include geographic regions</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>LA and Alameda County; Koreans and Vietnamese</td>
<td>State; County; Service Planning Area (SPA) level</td>
<td>Telephone</td>
<td>None</td>
<td>English, Spanish, Cantonese, Mandarin, Korean, and Vietnamese.</td>
<td>42,044 adults; 8,526 children; and 4,010 adolescent s</td>
<td>61.4% overall state cooperation rate</td>
<td>Adult (60.0%); Child (81.4%); and Adolescent (97.3%) - AAPOR RR3</td>
</tr>
<tr>
<td>Interview Survey (CHIS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 California Health</td>
<td>RDD: All eligible households with landline telephones</td>
<td>Stratified sampling; Strata include geographic regions</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Marin, Humboldt, and Solano Counties; Koreans and Vietnamese</td>
<td>County</td>
<td>Telephone</td>
<td>$2 cash in the advance letter</td>
<td>English, Spanish, Mandarin, Cantonese, Vietnamese, and Korean</td>
<td>43,020 adult; 11,358 child; 4,029 adolescent s</td>
<td>57% overall state cooperation rate</td>
<td>Household (59.3%); Adult (54.0%); Child (75.2%); Adolescent (48.5%) - AAPOR RR4</td>
</tr>
<tr>
<td>Interview Survey (CHIS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007 California Health</td>
<td>RDD: For landlines, all eligible households with landline telephones.</td>
<td>Stratified sampling; Strata include geographic regions</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>LA County and San Diego County; Koreans and Vietnamese</td>
<td>State; County; Service Planning Area (SPA) level</td>
<td>Telephone - landlines and cell phone lines</td>
<td>For landline sample, $2 cash in the advance letter; For cell phone sample, $5 for screener and $25 for completed interview; For area sample, $25 for completed interview</td>
<td>English, Spanish, Mandarin, Cantonese, Vietnamese, and Korean</td>
<td>51,048 adults; 9913, children; and 3638 adolescent s</td>
<td>41.7% landline/ surname sample</td>
<td>Household (59.4%); Adult (52.9%); Child (73.7%) and Adolescent (44.1%) - AAPOR RR4</td>
</tr>
<tr>
<td>Interview Survey (CHIS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey</td>
<td>Sample Frame</td>
<td>Sampling Strategy</td>
<td>FC</td>
<td>RC</td>
<td>LC</td>
<td>Areas Where Estimates are Available</td>
<td>Mode of Data Collection</td>
<td>Incentives</td>
<td>Language</td>
<td>Sample Size</td>
<td>Cooperation Rate</td>
<td>Response Rate</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>------------------------------------</td>
<td>-------------------------</td>
<td>------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>2005 Los Angeles County Health Survey</td>
<td>RDD: All eligible LA County households with landline telephones</td>
<td>Simple random sampling; additional child interviews conducted</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Antelope Valley (rural communities)</td>
<td>Telephone</td>
<td>None</td>
<td>English, Spanish, Cantonese, Mandarin, Korean, Vietnamese and Armenian.</td>
<td>8,648 adult interviews and 6,032 child interviews</td>
<td>46.6% adult interviews; 46.7% child interviews - AAPOR COOP3</td>
<td>Adult (22.8%); Child (26.0%) - AAPOR RR3</td>
<td></td>
</tr>
<tr>
<td>2007 Los Angeles County Health Survey</td>
<td>RDD: All eligible LA County households with landline telephones</td>
<td>Simple random sampling; additional child interviews conducted</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Antelope Valley (rural communities)</td>
<td>Telephone</td>
<td>None</td>
<td>English, Spanish, Cantonese, Mandarin, Korean, Vietnamese and Armenian.</td>
<td>7,200 adult interviews and 5,728 child interviews</td>
<td>40% adult surveys</td>
<td>40% child surveys - AAPOR COOP3</td>
<td></td>
</tr>
<tr>
<td>2008 Massachusetts Health Insurance Survey</td>
<td>Dual-Frame: RDD and Address-Based: US Postal Service Delivery Sequence File (DSF) residential addresses; All residential telephone numbers currently in use in DC</td>
<td>List-assisted simple random sample, stratified by listed vs. non-listed numbers</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>City</td>
<td>Telephone, Web, and Mail</td>
<td>$20 gift card to a grocery store or the Metro or donation to one of four local charities and entered into a drawing to win $100</td>
<td>English and Spanish</td>
<td>2,099 for RDD sample; 2,618 for address-based sample</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 Massachusetts Health Insurance Survey</td>
<td>Dual-Frame: RDD and Address-Based: US Postal Service Delivery Sequence File (DSF) residential addresses; All residential telephone numbers currently in use in Massachusetts</td>
<td>List-assisted simple random sample, stratified by listed vs. non-listed numbers</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Region</td>
<td>Telephone, Web, and Mail</td>
<td>All respondent entered into a drawing to win $100 or a pair of Red Sox tickets; Additional $20 for address-based sample with unlisted numbers.</td>
<td>English, Spanish, and Portuguese</td>
<td>1,619 or RDD sample: 3,291 for address-based sample</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prepared by Harder+Company Community Research for First 5 LA
<table>
<thead>
<tr>
<th>Survey</th>
<th>Sample Frame</th>
<th>Sampling Strategy</th>
<th>FC (^a)</th>
<th>RC (^b)</th>
<th>LI (^c)</th>
<th>Oversampled Population</th>
<th>Areas Where Estimates are Available (^d)</th>
<th>Mode of Data Collection</th>
<th>Incentives (^e)</th>
<th>Language</th>
<th>Sample Size</th>
<th>Cooperation rate</th>
<th>Response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA County WIC Survey</td>
<td>List-Based: WIC Participants in LA County and selected Service Planning Area</td>
<td>Simple random sampling</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Antelope Valley (rural communities)</td>
<td>County; Service Planning Area (SPA) level</td>
<td>Telephone</td>
<td>$10 gift card</td>
<td>English and Spanish</td>
<td>approximat ely 5000</td>
<td>92.60% - AAPOR CCOP1</td>
<td>66.50% - AAPOR RR2</td>
</tr>
<tr>
<td>LA Mommy and Baby (LAMB) Survey</td>
<td>Birth certificate file</td>
<td>Stratified sampling; strata not specified</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Low weight birth; pre-term births; African American; Asian/Pacific Islander; Native American</td>
<td>County; Service Planning Area (SPA) level</td>
<td>Telephone; Mail</td>
<td>$20 Ralphs or Food4Less gift card</td>
<td>English, Spanish; Chinese; and translators available in other languages</td>
<td>5,500 (in 2005)</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Los Angeles Family and Neighborhood Survey (LA-FANS)</td>
<td>List-Based: Dwelling units within sampled census tracts in Los Angeles County; constructed listing file developed through on-the-ground operations</td>
<td>Adults within households within census tracts; Selected households were screened in-person; Census tracts stratified by poverty level, sampled proportional to size; within strata</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Poor neighborhoods</td>
<td>County</td>
<td>In-person</td>
<td>None</td>
<td>English and Spanish</td>
<td>3090 household surveys; 3165 child surveys</td>
<td>Parents (85%); Primary Caregiver (89%); Child (87%); Sibling (86%)</td>
<td>Parents (61%); Primary Caregiver (63%); Child (62%); Sibling (61%)</td>
</tr>
<tr>
<td>Making Connections Survey</td>
<td>Address-Based: list-assisted probability sampling of households using US Postal Service Delivery Sequence File (DSF), all zip codes overlapping with designated MC areas</td>
<td>Simple random sampling and stratified sampling; strata include subareas within selected cities</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>City and subareas of</td>
<td>None</td>
<td>In-person</td>
<td>None</td>
<td>English, Spanish, and additional languages that were prevalent in the site (10 sites total)</td>
<td>7,498 households was interviewed. The average sample was approximat ely 750 in each city</td>
<td>Unknown</td>
<td>Average response rate across 10 sites was 69%</td>
</tr>
</tbody>
</table>
a Includes families with children under 18 years of age
b Includes rural communities
c Includes low income communities
d Small areas estimates are at the county level or smaller
e Includes survey reports where incentives are not mentioned.
American Community Survey

Overview

The American Community Survey (ACS) represents an exciting advance from the Census Bureau beyond the decennial census because it allows for timely estimates of populations particularly in small geographical areas including census tracts and block groups. Further, it represents an impressive implementation of an address-based sampling design as part of a mixed-mode probability-based survey that also employs mail and computer-assisted telephone interviewing (CATI). Unlike the decennial census, which represents a fixed point in time sample of households on Census Day (April 1), the ACS is a continuing monthly survey designed to provide estimates for 1-year, 3-year, and 5-year periods (Citro and Kalton, 2007). The first full implementation of the ACS, involving the selection of county8 level samples, took place in 2004 and was fielded in 2005.

After a decade of testing, the first set of ACS data products were released in August–November 2006 and reported the social, demographic, economic, and housing characteristics in 2005 of cities, counties, and other areas with 65,000 or more people. Later products, after a full five years of data have been collected, will provide similar estimates for smaller geographical areas including census tracts, neighborhoods and sub-blocks. With the advent of the ACS, there will no longer be a long-form sample as part of future decennial census surveys.

The ACS has three major benefits compared with the long-form sample:

   a) Timeliness: ACS data products are released 8–10 months after data collection;
   b) Frequency: ACS data products are updated every year instead of every 10 years, which will make it possible in many areas to track trends in such important population characteristics as educational attainment, employment, poverty, diversity, and others; and
   c) Higher Data Quality: Responses to survey items are much more complete compared with the 2000 long-form sample because of the supplemental computer-assisted telephone and personal interviewing of households that do not respond by mail. Some of this quality is also due to the high degree of training of the ACS interviewers and their level of experience, in contrast to the lightly trained temporary enumerators that were used for nonresponse follow-up in the 2000 census. In addition, ACS telephone interviewers contact mail respondents to obtain answers to missing items, a step not done in 2000.

In general, to make interpretations about neighborhoods and small areas in the ACS, it is important to examine margins of error and to review the non-sampling errors for estimates of interest. The Panel on the Functionality and Usability of Data from the American Community Survey (Citro and Kalton, 2007) also recommend using the same time frame when comparing estimates among areas or population groups and when analyzing trends over time using the one-year estimates. As well, to compare time periods, geographic boundaries may have changed making some comparisons difficult, and that overlapping period estimates (e.g., 5-year period estimates for 2010–2014 and 2011–2015) should not be used.

8 “County” refers to counties, county equivalents, and municipalities.
Sample Plan

The sample frame for the ACS is the Master Address File (MAF), which is the Census Bureau’s inventory of known residential addresses (housing units and group quarters) and selected nonresidential units in the United States and Puerto Rico. The addresses are selected from the MAF to represent both the Housing Unit (HU) and Group Quarters (GQ) populations, although different methods are used to select and interview these two universes. When all interviews are combined, the survey data provide estimates of the characteristics of the total residential population. We mainly focus on the sampling of Housing Units (Hu’s), which is closest to our purposes for the Family Survey Design, but we will also summarize the methods whereby Group Quarters (GQ’s) are sampled.

The MAF contains mailing and location address information and other attribute information about each address. It also contains geographic codes, such as county and place\(^9\) codes, obtained by linking to the Census Bureau’s Topologically Integrated Geographic Encoding and Referencing (TIGER) database. The geographic codes in the MAF, some of which come from the TIGER® database, identify a variety of areas, including states, counties, county subdivisions, places, Native American areas, Alaska Native areas, Hawaiian Homelands, census tracts, block groups, and blocks. The MAF in its state of existence at the conclusion of Census 2000, containing a complete inventory of known housing units (HUs) nationwide, served as the initial sample frame, but it continues to be updated on a six-month basis as the ACS is rolled out.

Beginning in 2005, the MAF was divided equally into five yearly segments, each of which was assigned to one year within each five-year period, such as 2005-2009. Additional samples for 2010-2014 and 2015-2019 will also be selected in the future based on a continually updated MAF. By segmenting the sample frame into these five-year periods, no specific address will be included in the ACS sample more than once every 5 years. The first five years of data will be required to produce estimates for the smallest areas, which comprise neighborhoods and census sub-blocks. Once the Census Bureau has collected the first 5 years of data, new small-area data estimates will be produced annually. Period estimates for larger geographic areas, such as states, counties, county subdivisions and census tracts, require only one and three years of data within each five year sample.

The initial sample is selected using systematic sampling from the MAF each month and spread throughout the United States without any clustering. Sampling occurs in two phases; the first phase divides the MAF into yearly segments while the second phase selects addresses for the ACS sample for each data collection year from the applicable segment. In the first phase, every January and August, newly added addresses are assigned equally to one of the five segments for the period then in progress: for example, new addresses identified in January 2006 were assigned equally to years 2005–2009. At the end of the five-year period, all addresses that still exist as housing unit addresses on the MAF at that time will be reassigned to the same yearly segments for the next five year period. For example, in August 2009 all still-valid addresses assigned to segments for 2005–2009 were reassigned to the same yearly segments for 2010–2014, and the process of assigning newly added addresses to segments for the next five years

---

\(^9\) “Place” is defined by the Census Bureau as “A concentration of population either legally bounded as an incorporated place, or delineated for statistical purposes as a census designated place (in Puerto Rico, a comunidad or zona urbana). See census designated place, consolidated city, incorporated place, independent city, and independent place.” From <http://www.census.gov/geo/www/tiger/glossary.html#glossary>.
proceeds each January and August until August 2014, when the process will begin anew (Citro and Kalton, 2007).

The second-phase sampling is designed to select ACS sample addresses from a given year's first-phase segment to meet specified sampling rates that are chosen to improve the precision of estimates for small governmental units. In this phase, the sample is divided into 12 panels of approximately 240,000 addresses each. Each month a new panel of sample addresses is introduced, and data will be collected using three sequential data collection strategies of increasing intensities (mail, telephone interviews and finally in-person interviews) to obtain survey data for the panel over a three-month time period.

The smallest sampling unit was defined as a geographic block, which represented an area of, on average, approximately 15–20 housing units. Blocks were assigned to one of five strata representing differential densities of housing units as follows:

1. Stratum #1, blocks in the smallest governmental units, with fewer than 200 occupied housing units;
2. Stratum #2: blocks in small governmental units with between 200 and 800 occupied housing units;
3. Stratum #3: blocks in small governmental units with between 800 and 1,200 occupied housing units;
4. Stratum #4: blocks in large census tracts with more than 2,000 occupied housing units; and
5. Stratum #5, all other blocks.

To keep the total yearly sample to 3 million housing unit addresses, and considering budget limitations, the initial sampling rate for census tracts in strata 4 and 5 (the largest units) was reduced by 8 percent. This allowed for a higher level of computer-assisted personal interviewing (CAPI) nonresponse follow-up in blocks with lower-than-average response rates by the initial mail and CATI. At least 75 percent of addresses in census tracts in strata 4 and 5 were mailable and response rates from mail and computer-assisted telephone interviewing (CATI) were estimated to be at least 60 percent.

Sampling rates were calculated for five strata in order to produce approximately equal precision for estimates of a given characteristic for both small governmental units and large census tracts outside these units. The initial sampling rates were based on estimates of occupied rather than total housing units because blocks in governmental units or census tracts with large numbers of seasonally vacant housing units would be under-sampled if total housing units were the criterion. The first phase sampling rates for 2008 ranged from a low of 1.5 percent to a high of 10 percent of the addresses in each block, across the five strata. Approximately 3 million housing unit addresses for one year within each five-year period were selected in 2005, representing approximately 2.3 percent of 129.5 million housing units on the MAF.

Sampling for Group Quarters (GQ's) was based on a separate group quarter MAF, which was constructed from an updated GQ inventory file following the last decennial census, merged with files from of federal prisons and detention centers from the U.S. Bureau of Prisons and a file of military bases and vessels from the Department of Defense. In addition, the Census Bureau identified new state prisons and state prisons that had closed based on internet research.

In the August preceding the data collection year for ACS, the main sampling phase selects all GQ samples are selected from two strata: small GQs which were estimated to have 15 or fewer
people as well as GQs listed as closed on Census Day, 2000, and larger GQs estimated to have more than 15 people. Sampling within the small GQ stratum follows the same two-phase approach used for the housing unit sample. There are five yearly segments in the first-phase followed by selection of the second-phase sample designed to yield a 2.5 percent systematic sample of small GQs within each state, sorted by GQ type and geography, which means that for each year, 12.5 percent (1 in 8) of GQs in the yearly sample are selected. In each GQ thus selected, every person is eligible to be interviewed, except where there are more than 15 people residing in the GQ (in which case the sample is reduced to 10 people).

For the larger GQ stratum, the GQs are divided into five yearly segments but are all included, sorted by type and geography. In the first phase, based on a measure of size using the estimated number of residents divided by 10, a 2.5 percent (1 in 40) systematic sample of groups is selected each year. Depending on the measure of size, GQs may be allocated to different months for data collection, while some GQs may not be selected. GQs in the large stratum may be selected in any year regardless of whether or not they were previously selected. The second-stage sampling unit for larger GQs is the person. Field visit the selected GQs assigned for a given month and determine the total number of residents at the GQ. Using an automated listing instrument they select 10 residents to be interviewed for that month. If a large GQ was selected for another month, the field representative will return to select another group of 10 individuals to be interviewed.

Due to budgetary and operational constraints, there were exceptions to this sampling procedure for correctional facilities and military barracks, in which all groups of 10 people in a state or local correctional facility or barracks with more than one sampled group are assigned to the same month, instead of being spread across months as is the case for other GQ types. For federal prisons, all sampled groups of 10 people are assigned to September, with a period of up to 4.5 months allowed for data collection. The U.S. Bureau of Prisons generates the person samples for the federal prisons that are selected by the Census Bureau for the year. For the 2006 ACS, about 18,000 GQ facilities were in the sample, including 850 military facilities and 148 federal prisons.

Data collection for the GQ sample in 2006 involved personal visits by field representatives, where they administered an automated Group Quarters Facility questionnaire (for the first visit) and a bilingual paper ACS questionnaire for each sampled resident. Although face-to-face interviewing was preferred, a variety of other data collection methods were allowed; the field representative may fill in the questionnaire by telephoning the sample person; conduct an in-person interview with a proxy, such as a relative or guardian; leave the questionnaire with the sample person to complete after ascertaining that the person is physically and mentally able to do so; or leave questionnaires with the contact person for the GQ to distribute them to sample persons and collect them after they are filled in (Citro and Kalton, 2007).

Response Rates

For 2008 and 2009, the Census Bureau received responses from about 51 percent of the initial sample. About 67 percent of the addresses selected in the first phase sample resulted in a completed CATI interview, but after a second phase in which nonresponses were subsampled for CAPI, the overall survey response rate reached 97 percent. The survey response rate is the ratio of the estimate of units interviewed after data collection is complete to the estimate of

---

all units that should have been interviewed, which appears to be a version of the AAPOR RR3 response rate definition, although it is not clear how the eligible respondents were defined.

**Methodological Issues**

The size and complexity of the ACS has produced a variety of challenges, including a number of important methodological issues that are useful to consider if using a mixed-method address-based approach for the Family Survey Design. The key issues are: coverage bias, differential response rates by mode and ethnic/cultural group, and the definition of a residence.

**Coverage Bias.** Because the ACS is a continuous monthly survey nationwide, the MAF as the sample frame must be as complete and accurate as possible. This requires that it be updated on a continuous basis in all areas of the country.

For city-style addresses, typically found in urban and suburban areas, updating the MAF is based almost entirely on the US Postal Service’s Delivery Service File (DSF) and the Census Bureau receives updated versions every 6 months. There are many known problems with the DSF as the source of updated households for the MAF. First, it misses many addresses in new construction areas, where it takes time to establish separate mailboxes and mailing addresses. Second, portions of the DSF are not updated at the same rate all around the country, and third, it does not clearly identify addresses in small multi-unit structures where mail may be delivered to a central hall or desk and not to the individual apartments (U.S. General Accounting Office, 1998:17-18).

To mitigate coverage issues, for the 2000 decennial Census the Census Bureau used a complete canvass of all 8.2 million blocks in 1999 supplemented by a Local Update of Census Addresses (LUCA) program in 2008, in which local governments are given the opportunity to review and update the residential address listings for their jurisdiction. While these efforts improved the completeness of the 2000 Decennial Master Address File, giving better point in time estimates, the ACS, which uses time period estimates, still suffered from less precision by under-representing growing areas of the country, and not accurately representing residences in small multi-unit structures (those with 2–9 apartments). Further, evidence from the 2000 census indicates that the problem of missed or erroneously identified addresses in these types of structures persisted in the 2000 MAF even after the block canvass and LUCA programs (Cito and Kalton, 2007).

For rural addresses, a different approach to improving the sample frame was taken because these addresses often involve route number and PO boxes rather than specific housing addresses. Starting with the MAF from the last census, the frame adds new housing information from the Community Address Updating System (CAUS), which covers largely rural blocks where use of the DSF does not provide adequate coverage. A set of administrative procedures that consider the address characteristics of existing MAF records for the county, changes in housing unit estimates for the county since the last decennial census, changes in the DSF tallies for the county and identifies blocks that would be expected to yield the most new units. Dean and Petersen (2005) found that CAUS was successful in adding addresses to the MAF that would not have been added by other means, but the study was limited and did not address the constraints of using CAUS to improve the identification of rural addresses.

**Response Rate Differences by Mode and Ethnic Group.** Although we cannot find response rate data by mode or ethnic group in ACS, there is evidence from the Census 2000 Supplementary Survey that response rates may differ by the three modes of data collection –
mail, CATI, and CAPI. Households responding by mail were more likely to be White, headed by an older person and were more likely to own their own homes. The response rates after mail contacts were highest in predominantly White census tracts and lowest in census tracts that were predominantly Native American and Alaskan. Response rates were also lower in census tracts that were predominantly Hispanic or Black. At the same time, the response rates were highest in these latter census tracts when CAPI was conducted to follow-up with those who did not respond to the initial mail contact. For census tracts that were predominantly Hispanic, the efforts of CAPI resulted in a response rate that did not substantially differ from the overall response rate (95.8 vs. 95.4 percent) suggesting the importance of conducting CAPI follow-up in census tracts that were predominantly Hispanic. Even with the CAPI follow-up, which boosted response rates, the overall response rates for census tracks that were predominantly Native American/Alaskan, Black or Hispanic were lower than for census tracts with predominantly White respondents. From an overall response rate of 95.4 percent for all addresses, response rates for census tracts that were predominantly American/Indian/Alaskan, Black or Hispanic were 89.1, 92.4 and 95.8 percent, respectively. One of the important lessons for the F5LA survey from these response rate differences is that any household survey based on address lists should also employ the more labor intensive in-person interviewing for non-responders to mail or telephone approaches, particularly for blocks that have predominantly Hispanic, African-American or Native American/Alaskan populations.

**Definition of Residence.** To count individuals in both housing units and group quarters, the ACS uses a “current residence” rule, which is defined as “people who live for more than 2 months at a sample address are assumed to be residents of that unit” (Citro and Kalton, 2007). It is also applied prospectively to include those who just moved to a residence but intend to live there for at least 2 months. This differs significantly from the decennial census which employed the “usual place of residence” rule. The change in the residency rule was required as a result of the ACS being a continuous monthly survey and the usual place of residence is more common in person-based surveys. There were a number of exceptions to the two-month rule to cover typical situations such as children living in boarding schools or summer camps, children in joint custody arrangements, and people who commute to a residence close to work but who also maintain a home residence. The Panel on the Functionality and Usability of Data from the American Community Survey (Citro and Kalton, 2007) were critical of this definition because it was not clear why this value was chosen rather than another time period, and because there are many exceptions that are missed, including “people with weekday and weekend residences, people who live and travel throughout the year in recreational vehicles, and people who move among the residences of several relatives or friends.” (Citro and Kalton, 2007). It is not clear the extent to which this rule might miss some or overestimate others living in specific residences but who change their residences over time.
2003 California Health Interview Survey

Overview

The California Health Interview Survey (CHIS, 2005a; 2005b; 2005c; 2005d) is one of the largest health surveys in the nation and it collects information on health status, health conditions, health-related behaviors, health insurance coverage, access to health care services, and other health and health related issues. CHIS has been conducted every two years since 2001 by the UCLA Center for Health Policy Research in collaboration with the California Department of Health Services and the Public Health Institute. The latest available report covers the 2009 CHIS survey, but since the design and methods tend to be slightly modified from one time period to the next, we will summarize each two-year CHIS separately. This summary is for the 2003 CHIS survey methods.

In all CHIS surveys, the sample is designed to provide statewide estimates of California’s overall population, major racial/ethnic groups and subgroups. It also provides estimates for large- and medium-sized counties in the California, and for groups of the smallest counties. The CHIS sample is representative of California’s non-institutionalized population living in households at the time of the sample selection.

Sample Size

Interviews were completed with 42,044 adults, 8,526, children, and 4,010 adolescents.

Sampling Plan

CHIS 2003 used a multi-stage sample design. First, the state was divided into 41 geographic sampling strata (33 single-county strata and 8 groups that included the 25 other counties). Then, within each stratum, households were randomly selected and within each household, an adult was randomly selected. Additionally, one adolescent and one child were randomly selected in those households with adolescents and/or children. The Los Angeles sample was augmented to allow estimates at the Service Planning Area (SPA) level and the Alameda sample was augmented to allow estimates for the cities of Oakland and Hayward as well as the remainder of the county. In urban counties with a high portion of the state’s Latino, African American and Asian populations, a larger sample sizes were allocated. Areas with relatively high concentrations of Koreans and Vietnamese groups were oversampled by using phone numbers for group-specific surnames drawn from listed telephone directories.

Data Collection Procedures and Methods

Survey staff interviewed one randomly selected adult in each sampled household. In those households with children (under age 12) or adolescents (ages 12-17), one child and one adolescent was randomly sampled. The sampled adult was interviewed, and the parent or guardian most knowledgeable about the health and care of the sampled child was interviewed. The interviews were administered using a computer-assisted telephone interviewing (CATI) system. On average, the interviews took 33 minutes for adults, 14 minutes for children, and 21 minutes for adolescents to complete. In general, interviews conducted in non-English languages took longer.
Languages

Interviews were conducted in English, Spanish, Mandarin, Cantonese, Vietnamese, and Korean.

Cooperation Rate

The overall state level cooperation rate was 61.4%.

Response Rate

The response rate for CHIS 2003 was calculated using the screener completion rate (i.e., success in introducing the survey to a household and randomly selecting an adult to be interviewed), and the extended interview completion rate (i.e., success in getting the selected person to complete the full interview). The method used to calculate the response rate is AAPOR’s RR4, which is based on both the number of people who fully completed and those who partially completed the survey divided by the number of people in the sample who were eligible for the survey.

The extended interview completion rate was 60.0 percent for the adult interview, 81.4 percent for child interview, and 57.3 percent for adolescent interview. Multiplying the screener and extended rates gives an overall response rate of 33.5 percent.

Estimating Sampling Error and Weighting

The sample was weighted to represent the non-institutionalized population for each sampling stratum and statewide. Weights were applied to compensate for differential probabilities of selection for households and persons within households, to adjust for non-response, to adjust for undercoverage in sample frame and data collection, and finally to reduce the variability of the estimates using auxiliary information. First, household weights were created based on completed screener interviews adjusted for a variety of factors listed above including unknown residential status, refusal conversion attempts, screener nonresponse, and multiple telephone numbers. Household weights were then used to compute person-level weights which adjusted for the within household sampling of persons and nonresponse. Person-level weights were adjusted using a raking method to ensure that estimates were consistent with 14 dimensions of the population from the 2003 California Department of Finance estimates of the numbers of persons by age, race, and sex, and from the 2000 Census of Population counts from the U.S. Census Bureau.

Design effects (DEFT)\(^{11}\) describe the variability of the estimates for each sample and stratum, based on the degree to which the sample estimates are close to what would be obtained for a simple random sample. A design effect of 1.0 means that the sample estimates are similar to those that would occur with a simple random sample, without stratification or post-stratification weights. In CHIS 2003, for the adult items, the average DEFTs for estimates for the state is 1.28 and for counties it ranged between 1.08 to 1.25. For the child items, the average DEFT at the state level is 1.37 and at the county level it ranged from 1.03 and 1.30. The average DEFT for the state estimates is 1.40 for adolescent items and for the county estimates it ranged from 1.01 to 1.40.

\(^{11}\) The design effect was defined as the ratio of the variance of the sample estimate for the survey to the variance of a simple random sample of the same sample size.
2005 California Health Interview Survey

Overview

The California Health Interview Survey (CHIS, 2007a; 2007b; 2007c) is one of the largest health surveys in the nation and it collects information on health status, health conditions, health-related behaviors, health insurance coverage, access to health care services, and other health and health related issues. CHIS has been conducted every two years since 2001 by the UCLA Center for Health Policy Research in collaboration with the California Department of Health Services and the Public Health Institute. The latest available report covers the 2009 CHIS survey, but since the design and methods tend to be slightly modified from one time period to the next, we will summarize each two-year CHIS separately. This summary is for the 2005 CHIS survey methods.

In all CHIS surveys, the sample is designed to provide statewide estimates of California’s overall population, major racial/ethnic groups and subgroups. It also provides estimates for large- and medium-sized counties in the California, and for groups of the smallest counties. The CHIS sample is representative of California’s non-institutionalized population living in households at the time of the sample selection.

Sample Size

Interviews were completed with 43,020 adult, 11,358 child, and 4,029 adolescents.

Sampling Plan

CHIS 2005 used a multi-stage sample design. First, the state was divided into 44 geographic sampling strata (41 single-county strata and three groups that included the 17 other counties). Then, within each stratum, households were randomly selected and within each household, an adult was randomly selected. Additionally, one adolescent and one child were randomly selected in those households with adolescents and/or children. Samples in Marin, Humboldt, and Solano Counties were enhanced. Also, to increase the number of child interviews, additional samples were selected statewide and in San Diego County and screened to identify households with children ages 0 to 11. Areas with relatively high concentrations of Koreans and Vietnamese groups were oversampled.

Data Collection Procedures and Methods

Survey staff interviewed one randomly selected adult in each sampled household. In those households with children (under age 12) or adolescents (ages 12-17), one child and one adolescent were randomly sampled. In households with children where the sampled adult was not the screener respondent, children and adolescents could be sampled as part of the screening interview, and the extended child (and adolescent) interviews could be completed before the adult interview. This procedure was new for CHIS 2005 and it resulted in a higher number of completed child interviews. The interviews were administered using the computer-assisted telephone interviewing (CATI) system. On average, the interviews took 35 minutes for adults, 15 minutes for children, and 20 minutes for adolescents to complete. In general, interviews conducted in non-English languages took longer.
Languages

Interviews were conducted in English, Spanish, Mandarin, Cantonese, Vietnamese, and Korean.

Cooperation Rate

The overall state level cooperation rate was 57 percent.

Response Rate

The response rate for CHIS 2005 was calculated using the screener completion rate (i.e., success in introducing the survey to a household and randomly selecting an adult to be interviewed), and the extended interview completion rate (i.e., success in getting the selected person to complete the full interview). The method used to calculate the response rate is AAPOR’s RR4, which is based on both the number of people who fully completed and those who partially completed the survey divided by the number of people in the sample who were eligible for the survey.

The extended interview completion rate was 59.3 percent for the household, 54.0 percent for the adult interview, 75.2 percent for child interview, and 48.5 percent for adolescent interview. Multiplying the screener and extended rates gives an overall response rate of 29.5 percent.

Estimating Sampling Error and Weighting

The sample was weighted to represent the non-institutionalized population for each sampling stratum and statewide. Weights were applied to compensate for differential probabilities of selection for households and persons within households, to adjust for non-response, to adjust for undercoverage in sample frame and data collection, and finally to reduce the variability of the estimates using auxiliary information. First, household weights were created based on completed screener interviews adjusted for a variety of factors listed above including unknown residential status, refusal conversion attempts, screener nonresponse, and multiple telephone numbers. Household weights were then used to compute person-level weights which adjusted for the within household sampling of persons and nonresponse. Person-level weights were adjusted using a raking method to ensure that estimates were consistent with 11 dimensions of the population from the 2004 California Department of Finance estimates of the numbers of persons by age, race, and sex, and from the 2000 Census of Population counts from the U.S. Census Bureau, and from the 2004 American Community Survey estimates of education attainment and types of households.

Design effects (DEFT)\textsuperscript{12} describe the variability of the estimates for each sample and stratum, based on the degree to which the sample estimates are close to what would be obtained for a simple random sample. A design effect of 1.0 means that the sample estimates are similar to those that would occur with a simple random sample, without stratification or post-stratification weights. In CHIS 2005, for the adult items, the average DEFTs for estimates for the state is 1.37 and for counties it ranged between 1.08 to 1.33. For the child items, the average DEFT at the state level is 1.63 and at the county level it ranged from 1.08 and 1.51. The average DEFT for

\textsuperscript{12} The design effect was defined as the ratio of the variance of the sample estimate for the survey to the variance of a simple random sample of the same sample size.
the state estimates is 1.39 for adolescent items and for the county estimates it ranged from 1.02 to 1.35.
2007 California Health Interview Survey

Overview

The California Health Interview Survey (CHIS, 2009a; 2009b; 2009c; 2009d) is one of the largest health surveys in the nation and it collects information on health status, health conditions, health-related behaviors, health insurance coverage, access to health care services, and other health and health related issues. CHIS has been conducted every two years since 2001 by the UCLA Center for Health Policy Research in collaboration with the California Department of Health Services and the Public Health Institute. The latest available report covers the 2009 CHIS survey, but since the design and methods tend to be slightly modified from one time period to the next, we will summarize each two-year CHIS separately. This summary is for the 2007 CHIS survey methods.

In all CHIS surveys, the sample is designed to provide statewide estimates of California’s overall population, major racial/ethnic groups and subgroups. It also provides estimates for large- and medium-sized counties in the California, and for groups of the smallest counties. The CHIS sample is representative of California’s non-institutionalized population living in households at the time of the sample selection.

Sample Size

The interviews were completed with 51,048 adults, 9,913, children, and 3,638 adolescents.

Sampling Plan

CHIS 2007 used a multi-stage sample design. For the first time, both landline and cellular telephone numbers were included in the random-digit-dial (RDD) sample. For the landline RDD sample, the state was divided into 44 geographic sampling strata (41 single-county strata and three groups that included the 17 other counties). Then, within each stratum, households were randomly selected and within each household, an adult was randomly selected. Additionally, one adolescent and one child were randomly selected in those households with adolescents and/or children. Also, the samples in Los Angeles and San Diego Counties were augmented. Areas with relatively high concentrations of Koreans and Vietnamese groups were oversampled.

A separate RDD sample of cellular telephone numbers was drawn. The cell RDD sample was stratified by area code and screened to identify whether it was associated with cell-only households. Cell numbers from households with landlines were not included in the sample and only adults were interviewed.

Data Collection Procedures and Methods

For the landline RDD sample, survey staff interviewed one randomly selected adult in each sampled household. They also sampled one adolescent and one child if they were present in the household and the sampled adult was the parent or legal guardian. In households with children where the sampled adult was not the screener respondent, children and adolescents could be sampled as part of the screening interview, and the extended child (and adolescent) interviews could be completed before the adult interview. For the cell RDD and area samples, only one randomly selected adult in each household was interviewed. Interviews were
administered using the computer-assisted telephone interviewing (CATI) system. On average, the interviews took 35 minutes for adults, 17.5 minutes for children, and 20 minutes for adolescents to complete. In general, interviews conducted in non-English languages took longer.

Languages

Interviews were conducted in English, Spanish, Mandarin, Cantonese, Vietnamese, and Korean.

Cooperation Rate

The overall cooperation rate for landline/surname sample was 41.7 percent.

Response Rate

The response rate for CHIS 2007 was calculated using the screener completion rate (i.e., success in introducing the survey to a household and randomly selecting an adult to be interviewed), and the extended interview completion rate (i.e., success in getting the selected person to complete the full interview). The method used to calculate the response rate is AAPOR’s RR4, which is based on both the number of people who fully completed and those who partially completed the survey divided by the number of people in the sample who were eligible for the survey.

The extended interview completion rate was 59.4 percent for the household, 52.8 percent for the adult interview, 73.7 percent for child interview, and 44.1 percent for adolescent interview. Multiplying the screener and extended rates gives an overall landline sample response rate of 21.1 percent. The response rate for the cell phone sample was 52.0 percent.

Estimating Sampling Error and Weighting

The sample was weighted to represent the non-institutionalized population for each sampling stratum and statewide. Weights were applied to compensate for differential probabilities of selection for households and persons within households, to adjust for non-response, to adjust for undercoverage in sample frame and data collection, and finally to reduce the variability of the estimates using auxiliary information. First, household weights were created based on completed screener interviews adjusted for a variety of factors listed above including unknown residential status, refusal conversion attempts, screener nonresponse, and multiple telephone numbers. Household weights were then used to compute person-level weights which adjusted for the within household sampling of persons and nonresponse. Person-level weights were adjusted using a raking method to ensure that estimates were consistent with 11 dimensions of the population from the 2004 California Department of Finance estimates of the numbers of persons by age, race, and sex, and from the 2000 Census of Population counts from the U.S. Census Bureau, and from the 2004 American Community Survey estimates of education attainment and types of households.

Design effects (DEFT)\textsuperscript{13} describe the variability of the estimates for each sample and stratum, based on the degree to which the sample estimates are close to what would be obtained for a

\textsuperscript{13} The design effect was defined as the ratio of the variance of the sample estimate for the survey to the variance of a simple random sample of the same sample size.
simple random sample. A design effect of 1.0 means that the sample estimates are similar to those that would occur with a simple random sample, without stratification or post-stratification weights. The design effects for each of the samples in CHIS 2007 were the following:

**Landline/List Sample.** For the adult items, the average DEFTs for estimates for the state was 1.37 and for counties it was between 1.15 to 1.38. For the child items, the average DEFT at the state level was 1.45 and at the county level it ranged from 1.10 and 1.45. The average DEFT at the state level was 1.34 for adolescent items and at the county level was between 1.05 and 1.32.

**Combine Landline/List and Cell Phone Sample.** For the adult items, the average DEFTs for estimates for the state was 2.30 and for counties it was between 1.20 to 2.96.

**Area Samples.** For Los Angeles County area, the average DEFT for estimates for adult items was 1.16 and across SPAs, it was between 1.01 to 1.33.
2005 Los Angeles County Health Survey

Overview

The 2005 Los Angeles County Health Survey (LACHS) was commissioned by the Los Angeles County Department of Health Services and conducted by Field Research Corporation. LACHS includes the Adult and Child survey components (Los Angeles County Department of Health Services, 2007). The purpose of the survey is to collect data on key health status indicators, health-related behaviors, health insurance coverage and access to health care among adults and children living in Los Angeles County (across the county's 26 health districts and 8 service planning areas). The survey also provides information on the characteristics and health behaviors of major racial and ethnic groups and those households below the federal poverty level.

Sample Size

A sample of 8648 adults age 18 years or older were interviewed for the Adult survey and 6032 interviews was completed among parents/caregivers of children age 17 years or less for the Child Survey.

Sample Plan

Using an unrestricted random digit dial (RDD) sampling methodology random samples of Los Angeles County landline telephone households were drawn. The sample is systematically selected from all eligible 10-digit telephone numbers assigned to the county. Additional Adult and Child interviews were conducted in Antelope Valley.

Data Collection Method

For the Adult Survey, in households with one adult, the respondent was interviewed. When there were two adults in a household, one adult was randomly selected using the CATI system. In households with more than two adults, if the adult being screen is not selected by CATI, the adult with the most recent birthday was selected as the respondent.

For the Child Survey, if household chosen for the Adult Survey has one or more children, at the conclusion of the interview, the respondent was asked a series of questions to identify the mother or primary caregiver of one or more of the children residing in the household. If multiple families reside in the household, the mother/caregivers with the most recent birthday were selected. If the respondent has multiple children, child with the most recent birthday was selected.

Languages

Both interviews for the Adult Survey and Child Survey were conducted in English, Spanish, Chinese (Mandarin and Cantonese), Korean and Vietnamese.
Cooperation Rate

The method used to calculate the response rate is AAPOR’s COOP3, which is based on the number of people who fully completed divided by the number of people who completed the survey, partially completed the survey, and those who refused. The cooperation rate for the Adult Survey is 46.6 percent the cooperation rate for the Child Survey is 46.7 percent.

Response Rate

The method used to calculate the response rate is AAPOR’s RR3, which is based on the number of people who completed survey divided by the number of people in the sample who were eligible for the survey. The response rate for Adult Survey is 22.8 percent and for Child Survey is 26.0 percent.

Estimating Sampling Error and Weighting

The weighting process involved two stages. In the first stage of weighting the data files, a base weight is calculated that account household size, the number of children in the household, the number of dedicated landline telephone lines entering the home, landline-only/cell phone-only/non-telephone household status, and health district distribution of county households for both Adult and Child surveys. In stage two, the Adult and Child Surveys are adjusted to population-based estimates of the actual adult and child populations in Los Angeles County.
2007 Los Angeles County Health Survey

Overview

The Los Angeles County Health Survey (LACHS) was conducted for the fifth time in 2007 to gather health related information about Los Angeles County residents (Los Angeles County Department of Health Services, 2008). Topics covered in the survey include access to health care, health care utilization, health behaviors, health status, and knowledge and perceptions of health-related issues. The information collected has been used to update the key health indicators, track health issues over time, identify emerging public health issues among adults and children in Los Angeles County, and address the root causes of poor health. The survey also provides information on the characteristics and health behaviors of major racial and ethnic groups and those households below the federal poverty level. The LACHS includes an Adult survey and a Child survey that are administered to the parent or guardian (usually the mother) of a child 0-17 years old.

Sample Size

A total of 7,200 adults (ages 18 years or older) were interviewed for the Adult survey and 5,728 parents of children ages 17 years or under were interviewed for the Child survey.

Sample Plan

Using an unrestricted random digit dial (RDD) sampling methodology random samples of Los Angeles County landline telephone households were drawn. The sample is systematically selected from all eligible 10-digit telephone numbers assigned to the county. Additional Adult and Child interviews were conducted in Antelope Valley.

Data Collection Method

For the Adult Survey, in households with one adult, the respondent was interviewed. When there were two adults in a household, one adult was randomly selected using the CATI system. In households with more than two adults, if the adult being screened is not selected by CATI, the adult with the most recent birthday was selected as the respondent.

For the Child Survey, if household chosen for the Adult Survey has one or more children, at the conclusion of the interview, the respondent was asked a series of questions to identify the mother or primary caregiver of one or more of the children residing in the household. If multiple families reside in the household, the mother/caregivers with the most recent birthday were selected. If the respondent has multiple children, child with the most recent birthday was selected.

Languages

Interviews were conducted in English, Spanish, Chinese (Mandarin and Cantonese), Korean and Vietnamese. Approximately 26 percent of the Adult Survey was completed in non-English languages and 43 percent of the Child Survey was interviewed in a non-English language.
Cooperation Rate

The method used to calculate the response rate is AAPOR’s COOP3, which is based on the number of people who fully completed divided by the number of people who completed the survey, partially completed the survey, and those who refused. The overall cooperation rates for both the Adult and Child surveys were 40 percent each. The cooperation rate for households that received a letter informing them of the survey (47%) was higher than those that were not mailed the letter or whose letter was returned as undeliverable (38%).

Response Rate

The method used to calculate the response rate is AAPOR’s RR3, which is based on the number of people who completed survey divided by the number of people in the sample who were eligible for the survey. The response rate was 18 percent for the Adult Survey and 15 percent for the Child Survey (including homes ineligible because no children resided there.)

Estimating Sampling Error and Weighting

The weighting process involved two stages. In the first stage of weighting the data files, a base weight is calculated that account household size, the number of children in the household, the number of dedicated landline telephone lines entering the home, landline-only/cell phone-only/non-telephone household status, and health district distribution of county households for both Adult and Child surveys. In stage two, the Adult and Child Surveys are adjusted to population-based estimates of the actual adult and child populations in Los Angeles County.
2009 District of Columbia Health Insurance Survey

Overview

The 2009 District of Columbia Health Insurance Survey (DC-HIS) was conducted by the Urban Institute and its subcontractor, SSRS/Social Science Research Solutions, to document health insurance options and coverage and access to and use of health care for the non-institutionalized population in the District (Ormand, Triplett, Long, Dutwin, and Robyn, 2010).

Sample Size

A total of 4,717 interviews were completed; 2,099 for the random-digit-dialing sample (RDD) and 2,618 for the address-based sample (AB).

Sample Plan

DC-HIS used a dual sample frame that combined RDD landline sample with AB sample. For the RDD sample, all residential phone numbers currently in use in District of Columbia was generated and ran against a reverse directory services to identify phone numbers in the RDD sample with known address. Then, the sample was divided into two strata: those with listed phone numbers and those without listed phone numbers. The AB sample was developed from the United States Postal Service (USPS) Delivery Sequence File (DSF), a computerized file that contains residential addresses serviced by the USPS which is updated weekly. To identify addresses with a listed telephone number, DFS was run against all listed landline telephone numbers in the state databases from vendors. Then, the AB-sample was divided into addresses with listed landline numbers and those without listed landline numbers.

With the use of two sample frames, households in both the RDD and AB sample frames had a higher chance of being selected to the final sample. To address this issue, households without listed telephone numbers in the AB sample were oversampled and households without a listed landline number in the RDD sample were undersampled.

The initial sample was drawn based on the estimated distribution of telephone ownership and expected response rate of households. Approximately 57 percent of the sample was drawn from the AB sample frame and 43 percent was drawn from the RDD sample frame.

Data Collection Method

Most of the data was collected using the telephone or the web. A small number of respondents were mailed a paper copy of the survey upon request. Advance letters and reminder letter were sent to households with known addresses in both the RDD and AB samples. The advance letter gave the option of completing the survey by calling into a survey center with a 1-800 number or completing it on-line. Letters to RDD and AB samples with known telephone numbers also notified the household that would be receiving a call to complete the survey. Telephone interviews were conducted using the CATI system.
Languages

Interviews were conducted in English, and Spanish.

Cooperation Rate

Unknown

Response Rate

The response rate is based on AAPOR’s RR3 definition with an AAPOR-approved alternative method of addressing ineligible households. The overall response rate was 34.1 percent. The response rates for the RDD sample was 43.7 percent and for the AB sample it was 27.3 percent.

Estimating Sampling Error and Weighting

The weights were used to produce estimates of the total population of District of Columbia as well as the adult and child populations. Weights were first constructed separately for the RDD and AB samples and then for the combined sample. The weights adjusted for differential sampling probably, to reduce nonresponse bias, and coverage bias. First, household weights were constructed by adjusting the base weight so that all households have the same probability of selection and by correcting for nonresponse. Then, person-level and target person weights were adjusted to ensure that the estimates were consistent with Current Population Survey, March Supplement population estimates for DC for the age, race/ethnicity and gender of the population, rates of home ownership, and Claritas population estimates of respondent ward status. For the combined sample, the weights were adjusted to be consistent with estimates of percentage of District households and residents without a landline telephone based on the information collected in the survey for the AB-sample. These new combined-sample weights are then post-stratified to the Census control totals described above.

In DC-HIS, the average design effect for estimates for the target person in the household was 1.85.
2008 Massachusetts Health Insurance Survey

Overview

The 2008 Massachusetts Health Insurance Survey (HIS) was conducted by the Urban Institute, and its subcontractor, International Communications Research/ICR to collect information on health insurance coverage and access to and use of health care for the non-institutionalized population in Massachusetts (Long, Triplett, Dutwin, and Sherr, 2008).

Sample Size

A total of 4,910 interviews were completed; 1,619 for the random-digit-dialing sample (RDD) and 3,291 for the address-based sample (AB).

Sample Plan

HIS used a dual sample frame that combined the RDD landline sample with the AB sample. For the RDD sample, all residential phone numbers currently in use in Massachusetts was generated and run against a reverse directory services. This allows the identification of phone numbers in the RDD sample with known addresses. The RDD sample was then divided into two strata: those with listed phone numbers and those without listed phone numbers. The AB sample was developed from the United States Postal Service (USPS) Delivery Sequence File (DSF), a computerized file that contains all types of residential addresses serviced by the USPS that is updated weekly. To identify addresses with a listed telephone number, DFS was run against all listed landline telephone numbers in the state databases from vendors. As with the RDD sample, the AB-sample was divided into addresses with listed landline numbers and those without listed landline numbers.

With the use of two sample frames, households in both the RDD and AB sample frames had a higher chance of being selected. To address this issue, households without listed telephone numbers in the AB sample were oversampled and households without a listed landline number in the RDD sample were undersampled.

The initial sample was drawn based on estimated distribution of telephone ownership and expected response rate of households in Massachusetts. Approximately 60 percent of the sample was drawn from the AB sample frame and 40 percent was drawn from the RDD sample frame.

Data Collection Method

The computer-assisted telephone interviewing (CATI) system was used to randomly select one household member to be the “target” person for the interview. Although the data for the interview was collected using multiple modes including telephone, web, and mail, a large majority of the surveys were collected on the phone using a CATI interviewer or on-line. Advance letters and reminder letter were sent to household with known addresses in both the RDD and AB samples.
Languages

Interviews were conducted in English, Spanish, and Portuguese.

Cooperation Rate

Unknown

Response Rate

The response rate is based on AAPOR’s RR3 definition with an AAPOR-approved alternative method of addressing ineligible households. The overall response rate was 41 percent. The response rates were 50 percent for the RDD sample and 37 percent for the AB sample.

Estimating Sampling Error and Weighting

The weights were used to produce estimates of the total population of Massachusetts as well as the adult and child populations. Weights were first constructed separately for the RDD and AB samples and then for the combined sample. The weights adjusted for differential sampling probably, to reduce nonresponse bias, and coverage bias. First, household weights were constructed by adjusting the base weight so that all households have the same probability of selection and by correcting for nonresponse. Then, person-level and target person weights were adjusted to ensure that the estimates were consistent with 2008 Current Population Survey estimates for Massachusetts for age, race/ethnicity and gender of the population, and rates of home ownership; and 2008 Claritas estimates for residence by region of the state. For the combined sample, the weights were adjusted to be consistent with estimates of percentage of District households and residents without a landline telephone based on the information collected in the survey for the AB-sample. These new combined-sample weights are then post-stratified to the Census control totals described above.

In HIS, the average design effect for individual estimates based on the combined sample for all persons in the household was 1.8 and the average design effect for estimates for the target person in the household was 1.5.
Los Angeles County Women, Infants, Children (WIC) Survey

Overview

The Los Angeles County Women, Infants, and Children (WIC) Survey is a telephone survey that collects information about the health of women and children, parenting practices, early care and education, and home and community environments from a sample of WIC participants (Public Health Foundation WIC Program, 2010 WIC Data Mining Report, n.d.). WIC participants are pregnant and breastfeeding mothers and mothers of children age 0-5 who are enrolled in WIC in Los Angeles County. It was designed and commissioned by Public Health Foundation Enterprises (PHFE) WIC program and conducted by Field Research Corporation.

Sample Size

A total of 5,015 surveys were completed in 2005 and 4,998 in 2008.

Sampling Plan

A random sample was drawn from list of all WIC participants provided by the California WIC Program. WIC participants from the Antelope Valley Service Planning Area were oversampled.

Data Collection Procedures and Methods

All selected WIC participants were mailed a postcard with information about the survey and a phone number to call for more information. The interviews were conducted using the Computer Assisted Telephone Interviewing (CATI) system. Up to sixteen attempts were made to reach and interview the WIC participants and on the last attempt, an 1-800 number was provided to call back to participate in the survey.

Languages

The interviews were conducted in English or Spanish.

Cooperation Rate

Overall household cooperation rate was 92.6 percent.

Response Rate

The response rate is based on AAPOR’s RR definition. The response rate was 66.5 percent.

Estimating Sampling Error and Weighting

Sampling weights were develop that aligned with the estimates of the total distribution of WIC recipients across the county’s eight Service Planning Areas. These weights enabled the
interviews conducted in the Antelope Valley region to their proper proportion of all WIC recipients countywide.
Los Angeles Mommy and Baby (LAMB) Survey

Overview

The Los Angeles Mommy and Baby (LAMB) Survey is a population-based survey of mothers who recently delivered a baby (Los Angeles Department of Public Health, n.d.). The survey collects information about events that happened before, during, and after pregnancy. The survey's questions are adapted from several national surveys including the Maternal and Infant Health Assessment and CDC's Pregnancy Risk Assessment Monitoring System. The data from the survey is used to better understand the cause of poor birth outcomes, identify the needs of the county, and to improve the health and human services offered to mothers and babies in the county.

Sample Size

Approximately 5,500 surveys were completed in 2005.

Sampling Plan

The sampling frame was all mothers who delivered a baby during the preceding two to six months and who were residents of Los Angeles County. Adoptive mothers were not included in the survey because many of the survey questions were about attitudes and perceptions related to being the birth mother. Also, mothers who had babies over seven months old were not included because they may have difficulty with recalling events that happened several months prior to their pregnancy. A population-based stratified sampling was used with oversampling of low birth weight, preterm, African American, Asian/Pacific Islander, and Native American births.

Data Collection Procedures and Methods

A mixed method data collection was used for the LAMB survey. Advance letters were mailed to explain the survey and let the mothers know that they will receive a survey packet within 10 to 14 days. After 7-10 days, second survey packets were mailed to nonrespondents. Then the nonrespondents were called for a phone interview.

Languages

The surveys were conducted in English, Spanish, and Chinese. Translators were available in other languages.

Cooperation Rate

Unknown

Response Rate

Unknown
Estimating Sampling Error and Weighting

Unknown
Los Angeles Family and Neighborhood Survey

Overview

The Los Angeles Family & Neighborhood Survey (L.A.FANS) studies adults, teens, children and neighborhoods in Los Angeles County (Sastry, Ghosh-Dastidar, Adams, and Pebley, 2005). The survey was designed to understand the effects of neighborhoods and families in children's development and well-being, the effects of welfare reform at the neighborhood level, and to study the process of residential mobility and neighborhood change. L.A.FANS was conducted by the RAND Corporation in collaboration with the UCLA School of Public Health. Funding was provided primarily by the National Institutes of Health. The Los Angeles Family and Neighborhood Survey data collection for Wave 1 began in April 2000 and ended in mid-January 2002. Data collection for Wave 2 of L.A.FANS began in Fall 2006 and ended in November 2008.

Child outcomes of interest collected in L.A.FANS include: health status, health care utilization and health insurance, child's reading and problem-solving skills, social development and behavior, anti-social and self-destructive behavior, school attendance and attendance in after-school care, day care and extracurricular activities, and behavior in school, homework, educational expectations and employment and wages.

L.A.FANS also collected extensive information at the neighborhood level using interviews with adult respondents (asking about neighborhood social cohesion and collective action), systematic social observations of neighborhood social and physical characteristics in blocks where respondents resided, and GIS data for all census tracts in LA County from the Neighborhood Services and Characteristics (NSC) database collected by the US Census at the census tract level from the 1990, 1990 and 2000 censuses.

Sample Size

A total of 9,378 addresses was selected for the L.A.FANS sample. They were chosen at random from the listing files that were constructed for this project. The listing files contained a record of each dwelling unit that was located in the areas sampled for the survey through a systematic on-the-ground listing operation. The total number of addresses selected was based on estimates of completed interviews using estimated rates of non-response, household vacancies, and ineligibility from the 1990 Census, various rounds of the Current Population Survey, and other sources. The target sample size consisted of 65 census tracts out of 1,652 total census tracts in Los Angeles County. The listing file produced 4,110 households for sample selection and from these, there were 3,085 completed rosters after screening for eligibility (Sastry and Pebley, 2002). However, according to Sastry et al (2005) a total of 3090 households were interviewed. It should be noted that this was less than the target of 3250 households due to higher fieldwork costs. This sample size was sufficient to produce a small effect size of 0.16 between the baseline group and a comparison group.

The sample sizes for adults and children varied depending on whether the household was with or without children. Completed interviews were obtained for 2,309 households with children and with 777 households without children (Sastry and Pebley, 2002). In households with children, a child was randomly selected from the household roster (termed the Randomly Selected Child or
RSC) and the mother of the RSC was also interviewed if she lived in the household or, if not, the adult identified as the child’s primary caregiver, and this person was termed the Primary Caregiver or PCG. Additionally, if the RSC had one or more siblings under 18 years of age who shared the same biological or adoptive mother and the same PCG, one sibling was randomly selected for interview (this child was designated the SIB). Once selected, all child respondents (i.e., RSCs and SIBs) completed questionnaire modules based on only their age. In households without children the respondent was an adult chosen randomly from the household roster, or the Randomly Sampled Adult (RSA). In most cases the RSA and the PCG were the same person. Interviews were completed with 2,044 Randomly Sampled Adults, of which 1,081 were Primary Caregivers, as well as 2,001 Randomly Selected Children and 1,164 siblings.

**Sampling Plan**

The sample frame was based on a list of dwelling units that were located within the 65 sampled census tracts. This list of dwelling units was developed specifically for this survey using systematic on-the-ground listing operations. In this sample, neighborhood units were defined as census tracts. The sample was based on a stratified sampling design in which, at the first level, all census tracts were divided into three strata based on the percent of the tract’s population living in poverty, from very poor (top 10 percent of the poverty distribution), poor (60-89th percentiles) and non-poor (bottom 60 percent), based on tract-level estimates from county and administrative data. Census tracts and blocks were selected with probability proportion to size (PPS) and with equal numbers of households from each tract (n=50). This first stage of sampling produced a self-weighted sample with probabilities that are the same for all households in a given stratum. However, it was necessary to oversample tracts from the very poor and poor strata, since most of the tracts fell within the third, non-poor stratum. Based on a simulation analysis, an approximately equal number of census tracts were allocated to each of the three strata, with 20 tracts each from the very poor and poor strata and 25 from the non-poor strata. This meant that poor tracts were oversampled relative to their probabilities of occurrence, thereby ensuring that there were a sufficient number of welfare recipients and minorities, especially African-Americans, for separate analyses and estimates.

In the second sampling stage census blocks within each census tract were selected and then households were sampled from these blocks, rather than from the tract as a whole. This design allowed the researchers to use blocks or groups of blocks as the measure of neighborhood in subsequent analyses. Overall 439 blocks were selected averaging 6.6 non-zero blocks per census tract, with a range between 2 and 14 blocks. Although blocks were sampled with probability proportion to block population size, a floor was placed on the sampling probabilities to limit the sample weights, and very large blocks were sampled with certainty.

In the third stage, 50 households were selected within each tract, which produced a balanced design in which the same number of households per tract would be interviewed. It also minimized the effects of block- and tract-level clustering. Households were selected from the listing of dwelling units developed through extensive field operations, and households with children under 18 years of age were oversampled so that they made up 70 percent of the sample (rather than the 35 percent if they were selected with equal probability). Further, households where interviews could not be completed in either Spanish or English were excluded from the sample.

In the final sampling stage, one adult respondent was sampled at random in each selected household (this person was designated as the randomly sampled adult or RSA). In households with children, on child respondent was also selected at random and designated at the randomly...
sampled child (RSC). These two respondents were then followed throughout the longitudinal survey. The selection of one child per family at random as the focal child produced an over-representation of small families, which was later adjusted through weighting.

**Data Collection Procedures and Methods**

The field work for Wave 1 of L.A. FANS was completed between April 2001 and January 2002. Interviewers visited each household in order to administer a screener that asks whether any children under 18 years of age resided in the dwelling. Households selected for the sample were first asked to complete a roster. Then, individual respondents were selected from among all full-time household residents. One adult respondent was sampled at random in each selected household (designated the RSA or randomly sampled adult). In households with children, one child respondent was also selected at random and designated the RSC (randomly sampled child). In households with children, the mother of the randomly selected child was selected as a respondent and termed the Primary Caregiver (PCG). If the RSC’s mother did not live in the household or was unable to answer questions about the child, the child’s actual primary caregiver was selected as the respondent to provide information on the selected child.

Interviews were conducted in person using computer-assisted personal interviewing (CAPI) and computer-assisted self-administered interviewing (CASI). Separate questionnaire modules were developed for each subject area/respondent combination.

**Languages**

Interviews were conducted in English and Spanish.

**Cooperation Rate**

The cooperation rates were 85 percent for adult; 89 percent for primary caregiver; 87 percent for child; and 86 percent for sibling (calculation based on assumption that none of the cases of unknown eligibility are in fact eligible).

**Response Rate**

The response rates were 61 percent for adult, 63 percent for primary caregiver, 62 percent for child and 61 percent for siblings (calculation based on estimate of the proportion of cases of unknown eligibility that are actually eligible).

**Estimating Sampling Error and Weighting**

L.A. FANS provided sample weights for RSAs, RSCs, PCGs, all adult respondents (RSAs and PCGs combined), and all child respondents (all RSCs and SIBs combined). All of the weights adjusted for over-sampling by strata, for the household selection probabilities by tract, and for the tract-specific rates of over-sampling of households with children and of household nonresponse.
Making Connections Survey

Overview

The Making Connections Initiative is a place-based comprehensive community initiative of the Annie E. Casey Foundation (AECF; Coulton, Chan, and Mikebank, 2010). It was launched in 1999 and focused on 10 metropolitan cities throughout the country with a goal to improve outcomes for disadvantaged children by strengthening their families, improving their neighborhoods, and raising the quality of local services. Community residents were viewed both as the beneficiaries and the partners in improvements that are sought in their neighborhoods and in the systems that serve them.

Household surveys were conducted in 2002-2003 as part of Making Connections in low-income neighborhoods in 10 cities (Denver, Des Moines, Hartford, Indianapolis, Louisville, Milwaukee, Oakland, Providence, San Antonio, and Seattle/White Center). The purpose of the survey was to assess residents’ perceptions of their neighborhood space and identity. Survey respondents were asked to draw a map of their neighborhood and to provide a name for the neighborhood. Then GIS tools were used to analyze the maps and identify spaces that were commonly identified by the residents and those that were not included in the neighborhood descriptions. The results show lack of alignment between residents’ perception of their neighborhood and the boundaries for the target areas. Although 83 percent of residents were able to draw a map of their neighborhood and 69 percent were able to provide a name for their neighborhoods, only 25 percent identified their neighborhood with the official named used by the MC initiative. The median size of the neighborhoods drawn by the residents (0.35 square miles) was smaller than the initiative target area (2.23 square miles). Overall, residents agreed about on the neighborhood boundaries in 6 of 12 neighborhoods within each MC target area.

Sample Size

A total of 7,498 households were interviewed. The average sample size was approximately 750 for each site.

Sampling Plan

Random samples of households were selected within each of the 10 communities. Subareas were identified in five cities and the subareas with separate samples: Denver had four subarea samples; Des Moines had two; Indianapolis, two; Hartford, three; and Providence, three. Seattle/White Center and Milwaukee oversampled pre-specified blocks. San Antonio, Louisville, and Oakland did not have subareas.

Each site was defined by census tracts and block areas, with households within the designated target areas comprising the household survey population. The sampling frame consisted of the U.S. Postal Service (USPS) master list of delivery addresses, the Delivery Sequence File (DSF), a frequently updated (typically every six months) list of all residential delivery points in the United States. This frame serves as the basis for the Master Address File (MAF) used by the U.S. Bureau of the Census for the decennial census of population and to update the MAF for the continuing monthly American Community Survey (ACS). All zip codes that overlapped with any part of the MC areas were included in the sample frame. All the addresses that were within the sites were mapped using a geocoding software. Then, field checks were conducted with
blocks containing 4000-5000 addresses at each site, which validated the sampling methodology.

All households selected for the sample were included whether they had children or not. However, in households with children, a roster of all children in the household was compiled, and one child was selected at random. An adult most knowledgeable about the selected child responded to the survey. In households without children, an adult was chosen at random.

Data Collection Procedures and Methods

The interviews were conducted in-person at the residents’ homes.

Languages

The interviews were conducted in English, Spanish, and additional languages that were prevalent in the site (the additional languages were not specified).

Cooperation Rate

Unknown

Response Rate

The average response rate was 69 percent.

Estimating Sampling Error and Weighting

Unknown