Workshop SUMMARY

Maintaining and Enhancing Representativeness of State Health Surveys: Lessons for the California Health Interview Survey (CHIS)

Presented

November 12, 2009—Bethesda, MD

Executive Summary

Surveys providing local or small-population estimates are often administered by telephone but, in the last 20 years, obtaining survey responses has required increasing effort. For a number of reasons, U.S. survey telephone response rates have decreased from about 50 percent to about 20 percent. Low response rates undermine confidence that estimates accurately represent the study populations and provide unbiased estimates of associations of interest.

A November 2009 workshop, *Maintaining and Enhancing Representativeness of State Health Surveys: Lessons for the California Health Interview Survey (CHIS),* was designed to discuss methods that CHIS is using to maintain response rates and evaluate sample representativeness. Organized by the Applied Research Program (ARP) within the Division of Cancer Control and Population Sciences (DCCPS) of the National Cancer Institute (NCI), the invited workshop attracted a broad range of researchers because these issues are central to all telephone surveys. Survey methodologists explored three pressing issues confronting CHIS and other random-digit-dial (RDD) telephone surveys:

- Estimates for local areas
- Ethnically diverse populations
- Nonresponse bias and coverage bias

The California Health Interview Survey (CHIS) is an RDD household telephone survey. CHIS is the largest population-based state health survey in the U.S. With its focus on health care access, health insurance coverage, health behaviors, chronic health problems, cancer screening, and other key health issues, CHIS provides comprehensive data that serve a wide range of research and policy purposes. CHIS aims to produce high-quality, population-based data that are representative of California's geographic and demographic diversity. Therefore, the CHIS project team has an ongoing interest in meeting response rate, noncoverage, and related methodological challenges that potentially affect the representativeness of CHIS data.

Estimates for Local Areas. CHIS provides estimates for the entire state and most counties, as well as sub-county estimates for the two most populous counties, Los Angeles and San Diego. The estimates enable counties to compare themselves with other counties and the state as a whole. CHIS frequently receives requests from policymakers and others for estimates of increasingly smaller or alternative geographic areas that are not easily accommodated by the CHIS geographically stratified sample design.

One option for generating health estimates for small geographic areas is to use statistical models such as small area estimation. Modeling supplements survey data from the target area/domain of interest with other data. CHIS produces direct estimates and, to a lesser extent, modeled estimates. Key questions that workshop participants raised were whether methods for obtaining small-area estimates could contain costs, address the methodological requirements of small geographic areas, and maintain data quality. While useful and sometimes very effective for generating estimates at the zip code level, for example, small area estimation is complicated, resource intensive, difficult to replicate, and potentially challenging for policymakers and health advocates to embrace due to its complexity. For CHIS, small area estimation will likely continue as a useful, but limited, supplementary tool.

Ethnically Diverse Populations. Collecting information from racial/ethnic and other small populations is essential for addressing health care access and other disparities. Collecting these data has multiple operational challenges, including inconsistencies in racial/ethnic categories across surveys and over time, inaccurate linguistic translation of survey instruments, and difficulties in hiring, training, and retaining qualified bilingual interviewers.

Workshop participants discussed several approaches to efficiently identifying members of small populations for oversampling. One option is to stratify telephone exchanges linked to areas with a high density of a target ethnic group, but this method may be inefficient or biased if all strata are not sampled. Obtaining auxiliary lists from organizations with ethnic ties or ethnically associated surnames is also possible, but this technique requires that the frame and the list share a common identifier and that the list be relatively complete, with broad coverage of persons in the population. Also, using government administrative data sets as auxiliary lists tends to pose a variety of data accessibility and quality challenges. A third option, network sampling, relies on the social networks of persons in an identified group; however, because the selection probability is typically unknown, this approach does not generate a probability sample of small population groups.

CHIS combines surname list sampling and geographic oversampling of areas with high proportions of the target population to produce small-group estimates (i.e., Koreans and

Vietnamese). After each data collection cycle, CHIS evaluates and modifies its sampling procedures for the next cycle to optimize the efficiency of oversampling small populations. These methods have been presented at survey methods conferences and are described in detail in the CHIS methodology reports.[1,2]

Coverage and Nonresponse Bias. Surveys rely upon the representativeness of their samples to ensure that estimates describe the target population. Low response rates and poor coverage are the two major sources of bias that can undermine sample representativeness in RDD surveys. Studies of bias in CHIS data have found that coverage bias is a greater threat to their representativeness than nonresponse bias.[3,4]

Coverage Bias. Coverage bias may occur when population members do not appear in the sample frame. A traditional RDD frame includes only households with land-line telephones; because such samples exclude households that use cell phones, coverage bias may occur. Research from the National Center for Health Statistics (NCHS) has demonstrated that "cell phone only" households are likely to be younger, have lower income, lack health insurance, and report being in good or excellent health. CHIS has developed methodologies for including cell phone samples, piloting them in 2005 and implementing them in 2007 and 2009;[5-7] however, cell samples are not suited well for small areas or small populations.

An alternative to drawing a sample of telephone numbers is to use an address-based sample (ABS). The frame consists of mailing addresses of residential housing units within the geographic area of interest. A sample for each area can be drawn based on population density. This approach has challenges, due to multilingual deployment and questionnaire complexity. Specifically, telephone numbers would be unavailable for about 40% of addresses and for almost all cell-only households. Switching to an ABS frame would require mail recruitment, which could introduce different errors that are not well defined.

Nonresponse Bias. Like other telephone surveys, response rates for CHIS are low and have declined significantly over time. Intensive follow-up analyses and benchmark comparisons suggest that little significant, systematic bias is present for sociodemographic variables, but no benchmarks exist to evaluate bias for most of the local-level health variables measured by CHIS. Consequently, CHIS has conducted several experiments to test methods for improving its response rates. Methods shown to improve response rates have been incorporated in the next survey administration, reported at survey method conferences, and described in the CHIS methodology reports.[2,8]

An encouraging finding is that small prepaid cash incentives improve response rates, or at least help to stem their decline, and CHIS has been using them since 2005. CHIS may benefit from additional research to understand the optimal dollar amount for an incentive and the impact of incentives at different stages of the interview process (screener, extended, refusal). Workshop participants urged the CHIS project team to further explore incentive options with the Office of Management and Budget (OMB) and funders to build on previous findings.[2,9-10]

Mixed-mode administration approaches (e.g., mail survey followed by telephone interviews with nonrespondents) have shown promise for improving response rates. However, it is unclear what kinds and how much bias may be introduced by changing from a single to a mixed-mode survey. Paper-based instruments do not permit the complex skip patterns possible with computer-assisted telephone interview (CATI)-administered surveys, nor are they conducive to multilanguage survey administration. Because information about mixed-mode data collection and bias is limited, workshop participants suggested embedding small-scale tests into CHIS data collection procedures to ascertain whether this approach can improve response rates.

Exploring New Pathways to Strengthen and Sustain CHIS. As discussed, CHIS has tested ways to improve response rates and documented promising approaches. The CHIS pilot test of a cell-phone-only sample in 2005 addressed the feasibility of conducting an omnibus health survey by cell phone and explored methods to integrate cell phone and land-line samples in the weighting process. CHIS 2007 included an area probability sample to systematically evaluate and quantify potential coverage bias and nonresponse bias.

Workshop participants identified methodological research needs in the following areas:

- Creating and testing alternatives to response rates to better measure representativeness
- Monitoring and evaluating alternatives to RDD sampling to sample small geographic areas
- Identifying the optimal combination of lists and geographic oversampling to produce representative estimates for specific ethnic subgroups

- Determining optimal incentives for reducing potential nonresponse bias
- Assessing the representativeness of survey data obtained through mixed-mode vs. single-mode RDD methodologies
- Exploring options to reduce coverage bias, including RDD with cell and land-line samples, address-based sampling, and other/multiple frames
- Exploring cost trade-offs to maintain or increase response rates, such as whether the costs of intensive follow-up needed to modestly improve response rates are offset by savings in follow-up costs when incentives are paid to respondents.

Maintaining and Enhancing Representativeness of State Health Surveys: Lessons for the California Health Interview Survey (CHIS)

November 12, 2009—Bethesda, MD

Agenda

This workshop brings together leading survey methodologists to discuss the most pressing issues confronting the California Health Interview Survey (CHIS) and other telephone surveys. The discussion focuses especially on the representativeness of population-based surveys with large, ethnically diverse populations and alternative approaches to random-digit-dial landline surveys given the continued growth of cell-phone-only households and declining response rates. Participants explore alternative data collection options for the future.

9:30 a.m. **Introduction**: Rachel Ballard-Barbash, Associate Director, Applied Research Program, NCI

9:45 a.m. Panel 1: The evolution of CHIS in a rapidly changing survey environment

CHIS goals, approach to content and sample design, survey methods, challenges to representativeness, and methodological responses to challenges

E. Richard Brown, Principal Investigator, CHIS *(Moderator)*

Michael Brick, Vice President, Westat *(Speaker)*

David Grant, Director, CHIS *(Speaker)*

Brian Harris-Kojetin, Statistical and Science Policy Office, OMB *(Discussant)*

11:15 a.m. **Break**

11:30 a.m. Panel 2: The growing challenges to representative inclusion of ethnically diverse populations and small geographic populations

Methodological and sampling strategies for dispersed ethnically diverse population groups and for localities

Sunghee Lee, Survey Methodologist, CHIS *(Moderator)*

Stephen Immerwahr, Survey Unit, Bureau of Epidemiology, New York City Department of Health and Mental Hygiene *(Speaker)*

David Takeuchi, Associate Dean for Research, University of Washington School of Social Work *(Speaker)*

James Jackson, Director, Institute for Social Research, University of Michigan *(Discussant)*

1:00 p.m. Keynote speaker and Lunch

Robert Groves, Director, U.S. Census Bureau

2:30 p.m. Panel 3: Multiple, supplemental, and alternative: Exploring new pathways to strengthen and sustain CHIS

Exploration of multiple and supplemental modes, alternative sampling frames, and other steps to strengthen representativeness and operational efficiency of telephone surveys

Michael Brick, Vice President, Westat *(Moderator)*

Mick Couper, Research Professor, Institute for Social Research, University of Michigan *(Speaker)*

Nancy Mathiowetz, Professor of Sociology, University of Wisconsin-Milwaukee *(Speaker)*

Karol Krotki, Senior Research Statistician, RTI *(Discussant)*

4:00 p.m. Wrap-up and next steps for CHIS

Nancy Breen, Economist, Applied Research Program, NCI

E. Richard Brown, Principal Investigator, CHIS

4:15-5:00 p.m. **Reception**

NCI–CHIS Methods Workshop
November 12, 2009
Bethesda, MD
Participant List
Planning Committee Members are starred (*)

Rachel Ballard-Barbash, M.D., M.P.H.
Associate Director, Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, EPN 4005
Bethesda, MD 20892
Phone: 301-402-4366 Fax: 301-435-3710
E-Mail: barbashr@mail.nih.gov

David Berrigan, Ph.D., M.P.H.
Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, Room 4095B
Bethesda, MD 20892
Phone: 301-451-4301 Fax: 301-435-3710
E-Mail: berrigad@mail.nih.gov

Partha Bhattacharyya, Ph.D.
Division of Behavioral and Social Research
National Institute on Aging
7201 Wisconsin Avenue
Gateway Building, Suite 533
Bethesda, MD 20892
Phone: 301-496-3131 Fax: 301-402-0051
E-Mail: bhattacharyyap@mail.nih.gov

Nancy Breen, Ph.D. *
Economist, Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, EPN 4094
Bethesda, MD 20892
Phone: 301-496-4675 Fax: 301-453-3710
E-Mail: breenn@mail.nih.gov

J. Michael Brick, Ph.D. *
Vice President
Westat
1600 Research Boulevard
Rockville, MD 20850-3129
Phone: 301-251-1500
E-Mail: mikebrick@westat.com

E. Richard Brown, Ph.D. *
Director, UCLA Center for Health Policy Research
Principal Investigator
California Health Interview Survey
10960 Wilshire Boulevard
Los Angeles, CA 90024
Phone: 310-794-0812 Fax: 310-794-2686
E-Mail: erbrown@ucla.edu

Mick Couper, Ph.D.
Research Professor, Survey Research Center
Institute for Social Research
University of Michigan
426 Thompson Street 4030 ISR
Ann Arbor, MI 48104
Phone: 734-647-3577 Fax: 734-764-8263
E-Mail: mcouper@isr.umich.edu

Marcie Cynamon, M.A.
Project Director, SLAITS Team Division for Health Interview Statistics
National Center for Health Statistics, CDC
3311 Toledo Road, Room 2113
Hyattsville, MD 20782
Phone: 301-458-4001 Fax: 301-458-4020
E-Mail: Marcie.cynamon@cdc.hhs.gov

W. Sherman Edwards
Vice President
Organizational Management Research Group
Westat
1650 Research Boulevard
Rockville, MD 20850-3129
Phone: 301-294-3993 Fax: 301-294-3928
E-Mail: ShermEdwards@westat.com

David Grant, Ph.D. *
Director, California Health Interview Survey
UCLA Center for Health Policy Research
10960 Wilshire Boulevard, Suite 1550
Los Angeles, CA 90024
Phone: 310-794-0916 Fax: 310-794-2686
E-Mail: dgrant@ucla.edu

Alyssa Grauman, M.P.H. *
Public Health Advisor
Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, EPN 4113
Bethesda, MD 20892
Phone: 301-594-4148 Fax: 301-453-3710
E-Mail: graumana@mail.nih.gov

Robert Groves, Ph.D.
Director, U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233
Phone: 301-763-2135
E-Mail: Robert.M.Groves@census.gov

Brian Harris-Kojetin, Ph.D. *
Senior Statistician
Statistical and Science Policy
Office of Management and Budget
725 17th Street, NW, Room 10201
Washington, DC 20503
Phone: 202-395-7314 Fax: 202-395-7245
E-Mail: bharrisk@omb.eop.gov

Stephen Immerwahr
City Research Scientist, Survey Unit
Bureau of Epidemiology Services
New York City Department of Health and Mental Hygiene
125 Worth Street, Room 315, CN-6
New York, NY 10013
Phone: 917-664-7728
E-Mail: simmerwa@health.nyc.gov

James Jackson, Ph.D.
Director and Research Professor
Institute for Social Research
University of Michigan
426 Thompson Street 5010 ISR
Ann Arbor, MI 48106
Phone: 734-763-2491 Fax: 734-764-2377
E-Mail: jamessj@umich.edu

Jennifer Kincheloe, Ph.D., M.P.H.
Research Scientist
UCLA Center for Health Policy Research
10960 Wilshire Boulevard
Los Angeles, CA 90024
Phone: 310-909-8725 Fax: 310-794-2686
E-Mail: jkinchel@ucla.edu

Carrie Klabunde, Ph.D.
Epidemiologist, Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, EPN 4005
Bethesda, MD 20892
Phone: 301-402-3362 Fax: 301-435-3710
E-Mail: klabundc@mail.nih.gov

Kevin Konty, M.S.
Director
Methods Bureau of Epidemiology Services
New York City Department of Health and Mental Hygiene
125 Worth Street, Room 315, CN-6
New York, NY 10013
Phone: 917-664-7728
E-Mail: kkonty@health.nyc.gov

Karol Krotki, Ph.D.
Senior Research Statistician
Statistics and Epidemiology
RTI International
701 13th Street, NW, Suite 750
Washington, DC 20005-3967
Phone: 202-728-2485 Fax: 202-728-2095
E-Mail: kkrotki@rti.org

Sunghee Lee, Ph.D. *
Survey Methodologist
California Health Interview Survey
UCLA Center for Health Policy Research
10960 Wilshire Boulevard
Los Angeles, CA 90024
Phone: 310-794-2399 Fax: 310-794-2686
E-Mail: slee9@ucla.edu

Benmei Liu, Ph.D. *
Mathematical Statistician
Surveillance Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6116 Executive Boulevard, Room 5045
Bethesda, MD 20892
Phone: 301-594-9114 Fax: 301-480-2046
E-Mail: liub2@mail.nih.gov

Laura Lund
Science Advisor
Center for Health Statistics
California Department of Public Health
PO Box 997377, MS 7214
Sacramento, CA 95814-7377
Phone: 916-552-8113
Fax: 916-650-6889
E-Mail: llund@dhs.ca.gov

Nancy Mathiowetz , Ph.D.
Professor
Department of Sociology
University of Wisconsin-Milwaukee
2200 E. Kenwood Boulevard
Bolton Hall 778C
Milwaukee, WI 53201-0413
Phone: 414-229-2216
E-Mail: nancym2@uwm.edu

Richard Moser, Ph.D.
Research Psychologist
Behavioral Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, EPN 4052
Bethesda, MD 20892
Phone: 301-496-0273 Fax: 301-435-7547
E-Mail: moserr@mail.nih.gov

Ernest Moy, M.D., M.P.H.
Medical Officer
Center for Quality Improvement and Patient Safety
Agency for Healthcare Research and Quality, DHHS
540 Gaither Road
Rockville, MD 20850
Phone: 301-427-1329 Fax: 301-427-1341
E-Mail: Ernest.Moy@ahrq.hhs.gov

Ann Nakamura, M.P.H.
Research Program Specialist II
Health Information and Strategic Planning
California Department of Public Health
PO Box 997426, MS 7700
Sacramento, CA 95899-7426
Phone: 916-650-6871
E-Mail: Ann.Nakamura@cdph.ca.gov

William Nicholas, Ph.D., M.P.H.
Director of Research
The California Endowment
1000 North Alameda Street
Los Angeles, CA 90012
Phone: 213-928-8944 Fax: 213-253-0971
E-Mail: wnicholas@calendow.org

Jean Olson, M.D., M.P.H.
Deputy Branch Chief for Field Studies
Epidemiology Branch
Division of Cardiovascular Sciences
National Heart, Lung, and Blood Institute
6701 Rockledge Drive, Suite 10018
Bethesda, MD 20892
Phone: 301-435-0397 Fax: 301-480-1667
E-Mail: olsonj@nhlbi.nih.gov

Van Parsons, Ph.D. *
Mathematical Statistician
Office of Research and Methodology
National Center for Health Statistics, CDC
3311 Toledo Road, Room 3219
Hyattsville, MD 20782
Phone: 301-458-4421 Fax: 301-458-4031
E-Mail: vparsons@cdc.gov

Jennifer (AJ) Scheitler, M.Ed.
Development and Grants Manager
UCLA Center for Health Policy Research
10960 Wilshire Boulevard
Los Angeles, CA 90024
Phone: 310-794-8286 Fax: 310-794-2686
E-Mail: ajscheitler@ucla.edu

Salma Shariff-Marco, Ph.D., M.P.H.
Cancer Prevention Fellow
Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, EPN 4009C
Bethesda, MD 20892
Phone: 301-435-4986 Fax: 301-435-3710
E-Mail: shariffs@mail.nih.gov

Edward Sondik, Ph.D.
Director
National Center for Health Statistics, CDC
3311 Toledo Road, Room 7204
Hyattsville, MD 20782
Phone: 301-458-4174 Fax: 301-458-4020
E-Mail: ESondik@cdc.gov

Shobha Srinivasan, Ph.D.
Health Disparities Research Coordinator
Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, EPN 6126
Bethesda, MD 20892
Phone: 301-435-6614
E-Mail: Sriniva2@mail.nih.gov

David Takeuchi, Ph.D.
Professor
School of Social Work
University of Washington
4101 15th Avenue, NE, Room 354900
Seattle, WA 98105-6299
Phone: 206-543-5133
Fax: 206-221-3910
E-Mail: dt5@u.washington.edu

Frances Thompson, Ph.D., M.P.H.
Applied Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
6130 Executive Boulevard, EPN 4095A
Bethesda, MD 20892
Phone: 301-435-4410 Fax: 301-435-3710
E-Mail: thompsof@mail.nih.gov

Elaine Zahnd, Ph.D.
Senior Research Scientist
Public Health Institute
55 12th Street, 10th Floor
Oakland, CA 94607-4046
Phone: 925-283-6432 Fax: 925-283-6432
E-Mail: ezahnd@phi.org

Table of Contents

1.0 Introduction

On November 12, 2009, the National Cancer Institute (NCI) sponsored a workshop, *Maintaining and Enhancing Representativeness of State Health Surveys: Lessons for the California Health Interview Survey (CHIS)*. The workshop addressed the most common and complex methodological issues facing CHIS and other population-based surveys fielded in the United States today. This document provides an overview of CHIS, describes the purpose of the workshop, and summarizes the workshop's presentations and discussions in key topic areas. Before the workshop, presenters provided background papers and PowerPoint presentations describing their understanding of methodological issues to be addressed during the workshop. A large portion of the workshop was videorecorded and is available at http://videocast.nih.gov/Summary.asp?File=15743.

1.1 Description of CHIS

Since its inception in 2001, CHIS has been a leader in providing rigorously collected, population-based health survey data that are carefully documented and widely disseminated. The NCI has supported CHIS since its first administration. This workshop is part of NCI's longstanding commitment to identify and implement cutting-edge methods to improve the quality and cost-effectiveness of CHIS and, in so doing, advance the field for development of other sub-national surveys.

CHIS is a household survey conducted by telephone using a random-digit-dial (RDD) methodology. Administered every two years by the University of California, Los Angeles (UCLA) Center for Health Policy Research (the Center) and modeled after the National Health Interview Survey (NHIS), CHIS is the largest and most comprehensive state health survey in the United States—the 2007 survey sample included more than 53,000 households. The purpose of CHIS is to produce high-quality population-based health data that are representative of California's geographic and demographic diversity. The survey focuses on access to health care, health insurance coverage, health behaviors, chronic health problems, cancer screening, and other health issues. CHIS data are used for a wide range of policy purposes. CHIS provides statewide, regional, and county-specific information.

Unlike most large-scale health surveys, CHIS applies community-based participatory research principles as each data collection cycle is developed. More than 145 individuals from more than 60 state and local policymaking bodies, public health agencies, advocacy groups, research organizations, and health care organizations serve on advisory boards, technical advisory committees, and work groups. These organizational units interact with CHIS research staff in an accountable advisory process to shape survey topics, measures, and sample design and select the languages for translation of the survey instrument.[11]

CHIS data are of interest to a broad range of users with varying levels of statistical sophistication, and the Center is committed to making the survey data usable and accessible. About one-fifth of the CHIS operating budget is devoted to dissemination. CHIS data and supporting documentation are widely distributed through reports, an online query system (*Ask*CHIS), and public use data files. Federal and state policymakers, local health departments, health care providers and health plans, state health agencies, community-based organizations, health researchers, and the media use CHIS data extensively. Multiple funders support each iteration of the survey, with specific counties or organizations providing funding for oversampling areas of interest or for including specific additional topics. CHIS's broad user base supports ongoing efforts to measure and document health status and service use, and would benefit from improvements in data validity.

1.2 Representativeness of Survey Data

Surveys rely upon the representativeness of their samples to ensure that estimates and other findings accurately describe the target population. In theory, a sample is said to be representative if key characteristics of the sample are similar to characteristics of the target population.

According to sampling theory, samples drawn at random from a population will be representative except for the variation that is attributable to chance. A confidence interval around an estimate delineates how the estimates from different samples might vary if random chance alone were to influence the representativeness of the samples.

In practice, certain factors that systematically interfere with random selection may compromise sample representativeness. Two important examples of these problematic factors are incomplete coverage (when some members of the target population are systematically excluded from the sampling frame) and nonresponse (when some randomly sampled units do not provide information). Both situations may potentially lead to biases that cause the estimates to differ systematically from the population parameters being measured.

Coverage Bias. Coverage bias occurs when a sample frame (a list of all possible units that might be sampled) includes less than 100 percent of the target population, when the excluded individuals have characteristics different from those who are included on the frame, and when this difference is systematically related to variables being studied in the survey. Thus, coverage bias exists only if the exclusion factor (the characteristic shared by the excluded individuals) is related to variables being measured. For example, a survey that collects information on substance abuse may be biased when a sampling frame of respondents with residential addresses is used *if* there is evidence that homeless and institutionalized individuals are more likely to have involvement with substance abuse. In order to avoid biased estimates, these individuals would need to be included.

Nonresponse Bias. Nonresponse bias occurs when less than 100 percent of the selected random sample responds to the survey, when individuals who respond to the survey have characteristics different from those who do not respond, and when these differences are systematically related to variables being studied in the survey.

For example, people who respond to a community survey may be more committed individuals who are more invested in the community's future than are nonrespondents. However, not all survey estimates will be adversely affected by differences in the propensity to respond. If only the most committed persons provide data to the survey, estimates related to opinions about future plans may be biased. In this instance, because the factor of commitment is related both to opinions about future plans and to the propensity to respond, estimates for the community's opinions could be biased. In contrast, if commitment is not correlated with a variable of interest (e.g., ratings of community services), estimates of the community's ratings may be unaffected by the absence of nonrespondents' data.

Thus, because bias is a function of the relationship between propensity to respond and each variable measured by the survey, higher response rates are not necessarily associated with low levels of nonresponse bias. Indeed, similar response rates from different surveys have been associated with different levels of nonresponse bias, and different sample estimates within the same survey can be subject to different nonresponse biases.[12] Groves and Peytcheva[13] examined response rates and nonresponse bias in 59 surveys and found no direct association between response rates and nonresponse bias. In fact, some surveys with response rates over 70% had levels of nonresponse bias that were similar to surveys with response rates under 20%, depending on the variable being studied. Although response rates have been used as a convenient, single indicator of a survey's representativeness and data quality, research indicates that low response rates do not necessarily indicate the presence of nonresponse bias and that nonresponse bias is a property of a particular variable, not of a survey.

Determining the Presence of Bias. Surveys sometimes conduct studies to assess the representativeness of the survey. Most of these studies focus on nonresponse bias. A study investigating potential bias involves at least two components. The first involves *identifying* the dependent (outcome) variables that are most important in the study and examining models containing the variables (predictors) that are likely to be correlated with these primary dependent variables. If such predictors can be identified, then a variety of methods are available to assess the potential for bias. One such method is to examine the frequency of the predictor in the respondent set and in the population. The development of such a model may require multiple research efforts over time. For example, if household surveys measure persons with cancer, the estimates could be compared with the number of people on various cancer registries.

The second factor involved in investigating potential bias relates to the *availability* of comparable data from both the sample and the population. Unfortunately, the range of information available at the population level is somewhat limited and may not correspond to the information required by the theoretical model. For example, lists of people who have diseases that are not covered by comprehensive registries are generally not available or reliable.

Although a more sensitive and accurate measure than coverage rate or response rate is clearly needed for evaluating the representativeness of samples and the validity of estimates, no such measure has been fully developed and accepted by the scientific community.[14] In the meantime, CHIS has conducted carefully designed studies to document the representativeness of its sample and assess coverage and nonresponse bias.

1.3 Description of the Problem

CHIS is facing the same challenges as other telephone surveys conducted in the United States, and these coverage and nonresponse challenges could result in bias in at least some important estimates from the survey.

One growing challenge for telephone surveys is the adoption of cell phones as the only or primary means of telephone communication for a growing proportion of households. Cell phone numbers have not traditionally been included in RDD surveys. However, about 1 in 5 households use cell phones as their only telephone connection (“cell-phone-only” households), and an additional 14% of households have landlines but use cell phones for most communication (“cell-phone-mostly” households).[15] This situation is particularly problematic because individuals who rely solely or primarily on cell phones tend to have different characteristics related to health status and behaviors—variables of primary interest in CHIS. For this reason, CHIS has tested (in 2005) and implemented (in 2007) a sampling frame with both landlines and cell phones to reduce potential coverage bias.

Over the last 20 years, more and more effort has been required to obtain survey responses from a public that is increasingly reluctant to participate in surveys. Despite greater expenditures, response rates for telephone surveys, in particular, have declined. Between CHIS 2005 and 2007, total interviewer hours increased by 35% while response rates fell.[4] The adult landline overall response rate in 2007 was 18.7%, similar to that of the California 2007 Behavioral Risk Factor Surveillance System (a large, national telephone survey sponsored by the Centers for Disease Control and Prevention).

The increasing complexity and diversity of the U.S. population and of communication technologies have likely contributed to the declining response rate/increasing cost challenge. Consequently, the "one size fits all" approach used in the past to recruit potential respondents is no longer feasible. Differences in factors such as ethnicity/culture, language, response mode preference, literacy, and telephone usage must be considered in order to collect data from persons with diverse backgrounds at comparable rates and, in so doing, reduce nonresponse.

Despite these substantial challenges to maintaining the representativeness of its sample and containing costs, CHIS has continued to deliver high-quality data, and demand for its data remains strong. In fact, demand is increasing and California users are requesting data from smaller and smaller units. In order for survey estimates to be socially relevant and useful for policy and research purposes, CHIS is exploring methodological options that provide cost-effective solutions to complex issues involving representativeness of small area/small group estimates. In summary, while CHIS is facing the challenges of coverage and nonresponse typical of telephone surveys, it is also being called upon to provide estimates for increasingly smaller population subgroups.

1.4 Purpose of the Workshop

The purpose of the NCI-sponsored workshop was to bring together leading survey methodologists to discuss the challenges confronting CHIS and other telephone surveys. The CHIS Methods Workshop Planning Committee asked speakers to focus on the most pressing issues related to the representativeness of population-based surveys conducted with large, ethnically diverse populations. Speakers were also asked to discuss alternative approaches to landline RDD surveys given the continued growth of cell-phone-only households and declining response rates. Three main topic areas were addressed during the workshop's presentations and discussions:

(1) Sampling and producing estimates for small areas
(2) Ethnically diverse populations
(3) Coverage and nonresponse bias related to:
   (a) Steep declines in RDD response rates
   (b) Increased use of cell phones to replace or supplement landlines in households (the primary sampling unit for RDD studies).

Discussants were asked to help distinguish true threats to data quality from the mere appearance of bias, consider what alternatives might be most cost-effective, and discuss short- and long-term plans and possible solutions. The following sections summarize the discussions related to each main topic area.

2.0 Methods Supporting Small Area Estimates

Interest in CHIS data has increased as policymakers, program planners, and advocates have used local data to inform decisions. Researchers have used CHIS to assess the need for interventions and to appropriately design, monitor, and evaluate them. All have recognized the utility of having local evidence. Initially, CHIS focused on the county level as the smallest geographic unit, because many health decisions in California are made at that level. However, because they value the utility of CHIS to inform their decisions, policymakers and others request estimates from even smaller geographic areas, particularly within large counties such as Los Angeles and San Diego. In addition, some relatively sparsely populated counties initially grouped together for estimation purposes have requested their own county estimates.

Creating estimates for small geographic areas is challenging from both financial and practical standpoints. Significantly increasing sample sizes to produce direct estimates is often not financially feasible, while matching a telephone frame to a small geographic area is becoming increasingly difficult. At present, CHIS produces both direct and modeled estimates and is interested in exploring other methods to contain costs, address the methodological requirements of small geographic areas, and maintain data quality. Section 2 of this report describes the challenges related to estimates for small areas in greater detail, along with potential strategies for meeting those challenges.

2.1 Probability Sampling With Telephone Samples and the Problem of Small Areas

The sampling plan for CHIS has been carefully designed to maximize the efficiency of producing county-level estimates. CHIS uses a geographically stratified sampling design that samples telephone numbers for each geographically defined stratum. During implementation, sample is released in replicates (multiple small samples with the same statistical properties) to efficiently obtain the target number of interviews in each sampling stratum. This strategy has generally been successful but is threatened by interest in sampling smaller geographic communities and the increasing use of cell phones, which are not easily associated with geography.

Equal Probability and Stratified Sampling Plans. The simplest plan for creating the CHIS sample would involve drawing a statewide random sample, where every residential telephone number in the state would have an equal probability of inclusion in the sample. Under this scheme, densely populated areas would generate a large number of cases, permitting precise estimates, whereas sparsely populated areas would generate fewer cases, yielding less-precise estimates. Using an equal probability approach, in order to obtain enough cases to meet the required precision for sparsely populated areas, far more cases than necessary would be drawn from the high-density areas; this approach is neither cost-efficient nor statistically efficient.

A more efficient alternative would be to implement a stratified sample, where small geographic areas are sampled at higher rates in order to meet common precision targets. In CHIS, this entails matching the RDD frame to specific geographic areas and then sampling at a different rate within each geographic area in order to provide no more cases than are needed for meeting precision targets.

Replicate Samples. The release of sample in identical sample replicates is a frequently used strategy for efficiently obtaining a targeted number of completed interviews despite uncertainties regarding sample productivity. Rather than drawing a single sample to complete a target number of interviews in a given county, for example, multiple smaller samples with the same statistical properties are drawn. In this way, the sample can be worked in smaller pieces and the productivity of the released sample can be assessed to evaluate the need for release of subsequent replicates.

Replicate samples are also used to address uncertainties when the sample is targeted to a relatively small geographic area, such as a neighborhood or zip code. For RDD samples, telephone exchanges are only modestly associated with geography. Replicate samples for a geographically targeted area can be released and closely monitored to determine which exchanges are most and least productive. This information can then be used to release additional replicates for productive exchanges while suppressing unproductive replicates. Thus, ongoing experience informs data collection procedures, thereby achieving greater efficiency. Also, as experience is gained with populations in different areas and with a particular survey instrument, sample management becomes simpler and it becomes possible to release fewer replicates after the main sample.

The use of replicates is most successful when researchers anticipate that a high proportion of initially selected cases will later be excluded as ineligible. This may occur when statisticians are unable to map telephone numbers accurately to the geographic area of interest (e.g., in small neighborhoods in New York City or in cell phone samples) or when extensive screening is needed to identify members of a rare group. In such cases, the strategy helps researchers tailor the size of the released sample to the real-time completion rate and precision targets.

2.2 An Alternative—Address-Based Sampling

In order for CHIS to meet the demands for local-level (county and sub-county) data, the RDD sample frame requires that telephone numbers be matched to geography. The geomapping of a telephone frame is complicated by (1) increased telephone number portability, leading to less precision in the matching of telephone numbers to geography; and (2) cell phone samples, which have very limited geographic information available. Particularly for cell phone samples, the area code and exchange often no longer indicate the current residential location of the individual.

An alternative to mapping a sample of telephone numbers is to use an address-based sample (ABS) based on a frame provided by the U.S. Postal Service. The frame is formed by mailing addresses of residential housing units within the geographic area of interest, and a sample for each area can be drawn based on population density. This appears to be a promising alternative to geomapping an RDD frame and fielding a separate effort that captures cell phone users.

There are some complexities involved in using an ABS for telephone surveys. Vendors are able to match most addresses with telephone numbers, so telephone calls for these “matched” sampled units can proceed in a manner similar to that used with an RDD sample. However, in 2010, 30%–40% of addresses in the ABS frame could not be matched to telephone numbers. In order to conduct a telephone survey with potential respondents at these “unmatched” addresses, an initial recruitment step, such as mail or in-person contact, is needed. Recruitment mailings to unmatched addresses could request that potential respondents call in to a telephone center or return a postcard with a telephone number. If there is no response to the mailing, a recruiter could make an in-person visit to the unmatched address. At the visit, the respondent could be interviewed using a cell phone provided by the recruiter. Thus, using an initial recruitment step for unmatched addresses retains the telephone as the primary data collection mode for respondents from both matched and unmatched addresses in an ABS, thereby reducing the chance of response mode bias.

There are methodological issues associated with implementing the strategy just described. Because the unmatched sample would be recruited differently than the matched sample, there is high potential for different response rates and systematic bias that is not well understood or quantified. More research is needed in this area. In addition, in-person recruitment is expensive, and a sampling plan would need to be carefully developed to minimize the number of recruitment visits while retaining precision.

2.3 Small Area Estimation

Policymakers and analysts often wish to estimate characteristics for geographical areas smaller than the sampling frame (e.g., a city within a sampled county) and/or have inadequate sample to produce “direct” estimates. The sample size required for direct estimation may exceed the number of cases available from an existing survey, or when planning future data collection the cost of targeting these special domains may far exceed financial resources. In cases where direct estimation is not feasible, model-based methods such as *small area estimation* may represent a viable alternative.

Small area estimation involves creating statistical models that use available survey data from the target area of interest supplemented with data from other surveys and external sources. The model produces an estimate for the area of interest by “borrowing strength” from areas outside the target. The small area estimate uses such supplemental data from areas similar to the target to model the estimate. While direct estimation is considered to be a robust, globally applicable procedure that works well with large sample sizes, small area models tend to be variable-specific and require some model validation before being accepted as producing a sufficiently accurate and precise estimate. Rao[16] provides a comprehensive (and somewhat technical) review of this topic.

An example of small area estimation was published in 2007. Yu et al.[17] described using patterns of associations from the Current Population Survey (CPS) to generate health insurance estimates for legislative districts in California. The CPS includes items assessing insurance status, but only 50,000 cases are fielded nationwide—too few to generate estimates for smaller areas. Moreover, California legislative districts do not match the sampling boundaries utilized in the CPS. Using CPS California data, the researchers modeled age, gender, race/ethnicity, and income in logistic regressions predicting insurance status. To produce the synthetic estimates for legislative districts, the modeled probabilities were merged with 1990 decennial U.S. Census tract data containing information on the predictor variables, which were updated with information from a private marketing firm. Finally, as a check, the new estimates were combined and compared with direct estimates from the CPS in 2000.

2.4 Summary

Generating estimates for small areas presents numerous challenges. RDD sampling strategies have been modified and adapted to collect data to meet CHIS goals. However, additional and possible alternate strategies must be considered as the social context for data collection changes and new technologies develop.

3.0 Methods Supporting Representative Inclusion of Ethnically Diverse Populations

Disparities in health care access and health status are addressed at the local level. This underscores the need for local information about racial/ethnic and other small subpopulations. Because non-English-speaking subpopulations are critical to collecting accurate racial/ethnic data in California, CHIS spends substantial resources to address methodological issues associated with including an ethnically diverse population. This section describes the challenges associated with collecting representative data for rare subgroups within California's population, as well as some of the solutions already implemented by the CHIS team and others suggested by workshop participants.

3.1 Survey Development and Administration

Multiethnic survey development and administration pose multiple challenges, and these challenges have cost implications. First, racial and ethnic categories are not universally defined and frequently change. There are multiple categories of race and ethnicity based on genes, culture, institution, and socioeconomic status (SES). Moreover, the social construction of racial and ethnic categories changes over time in response to current sociopolitical issues; and items, response codes, and methods for collecting data on race/ethnicity have changed and continue to change. For example, Census questions regarding race have changed substantially since 1890, when "white, black, mulatto, quadroon, octoroon, Chinese, Japanese, or Indian" were the precoded categories and the interviewer provided the response based on visual inspection. Today, the Office of Management and Budget (OMB) requires that respondents report their own race separate from ethnicity (self-identified), with hundreds of combinations possible across all Federal surveys.

In order to ensure that racial/ethnic data accurately reflect the California population, CHIS staff have worked with a multicultural technical advisory committee to develop and define appropriate racial/ethnic subcategories. Because CHIS receives some Federal funding, it also must conform to OMB requirements.

Second, to maintain data quality, careful translation and review must be conducted to develop instruments for respondents who are not fluent in English. Survey items require review and testing to ensure that respondents who are not native English speakers understand the concepts being measured. CHIS is administered in five languages and, to the greatest extent possible, surveys in all languages including English undergo multiple reviews.

Third, additional costs are involved in training and hiring multilingual interviewers. It is difficult to find, train, and retain interviewers for these positions, and bilingual interviewers are paid more than interviewers who speak English only.

3.2 Identifying Members of Small populations.

Depending on a survey's overall sample size and the proportion of a group of interest within the sampling frame, random sampling alone may not provide sufficient cases to produce a precise estimate for the group. A large survey effort such as CHIS is able to support direct statewide estimates for many racial and ethnic groups in California without undertaking special methods. Estimates with adequate precision can be produced for Latino, African-American, Chinese, Filipino, Japanese, and South Asian subgroups without extra effort. However, to obtain precise estimates for additional smaller population subgroups (e.g., Korean and Vietnamese), oversampling is required. Oversampling can be implemented using geographic targeting, auxiliary lists, network sampling, or administrative data.

Geographic Targeting. In order to minimize costs, researchers may use auxiliary information to enhance the sample's yield. In CHIS, telephone exchanges are stratified according to geographic areas with relatively high concentrations of targeted oversample ethnic groups, such as Koreans and Vietnamese. However, some groups, such as South Asians in California, are dispersed throughout the state, so geographic targeting is not always a productive oversampling method. In addition, if oversampling of high-yield geographic areas is the only method used, factors associated with the subgroup could be confounded with the characteristics of persons living in that area. For example, the attributes of Hmong living in Fresno, an area with high Hmong density, may be quite different from attributes of Hmong dispersed across California.

Auxiliary Lists. Lists from a variety of sources can be used to identify and survey rare populations of interest; such lists may include membership lists for organizations with ethnic ties, lists of ethnically associated surnames, or lists from agencies serving a particular subpopulation. The quality and characteristics of auxiliary list samples vary widely, and their appropriate use depends on the survey's goals and which alternative sampling frames are available. If more than one frame is used for sampling, both samples need to have a common element, such as a telephone number, to appropriately account for selection probabilities. Another consideration is the completeness of the list; it must contain a large proportion of the subgroup of interest. Lists that are more complete make screening and sample yield more efficient. Finally, the list should cover a relatively broad spectrum of the target group's members; lists in which some important characteristic is underrepresented may introduce bias.

Early CHIS experience with using surname-listed samples to target racial/ethnic groups revealed that if lists are used exclusively or primarily to target a group, the sample becomes less efficient, due to large design effects. "Design effect" is a term used in comparisons of specific sample design alternatives. The design effect is defined as the ratio of the sampling variance of an estimate for the total population under a complex sample design (such as one involving lists) to the variance of the estimate for a simple random sample (SRS). For a given number of respondents, an SRS often has higher precision for overall population estimates. However, under an SRS, the sample size for rare groups is small and hence the precision of estimates is low. Deviating from an SRS to increase the number of cases in rare groups using lists decreases the variance for rare group estimates but may increase the variance in the overall estimate (hence increasing its design effect). In general, the quality and characteristics of the data in the list, the balance of priorities between overall and subgroup estimates, and the resources available for additional sample should be weighed carefully when considering the use of surname-listed sample.

Since CHIS 2003, a combination approach has been used to oversample Korean and Vietnamese households from both surname lists and geographically targeted RDD samples. Lists used or considered for other populations, such as American Indians, Alaska Natives, and Pacific Islanders, may not provide adequate coverage. Other strategies will need to be investigated for identifying members of these groups.

Network Sampling. Another alternative to identifying members of a subgroup of interest is network sampling, which taps into the social networks of persons who belong to the target group and know other individuals in the same group. Network sampling comprises several techniques, including respondent-driven sampling. This approach can save substantial resources if the persons supplying the names have a replicable, accurate understanding of the target group's characteristics *and* this referral follows a random fashion (i.e., is free from any systematic patterns).

However, this option has features that make it problematic for producing samples that are suitably representative for scientific research. The largest problem is that the probabilities of selection cannot be determined for most network sampling techniques, resulting in a nonprobability sample. Additionally, this approach often requires that the social networks have several characteristics in order to avoid systematic biases. For example, the structure of the social networks of the persons supplying additional names must have one component, meaning that there is one path between every person and every other person in the network. The network is also assumed to provide adequate coverage of the target subgroup. Third, referrals are assumed to be made randomly, without any personal judgment. These assumptions are difficult to verify, and therefore network sampling has not been employed in CHIS.

Administrative Data Sets. Government administrative data (such as the California Department of Motor Vehicle Registration database) hold promise as auxiliary lists for identifying rare groups. They may also provide information about telephone numbers associated with sampled addresses or used as an address-based sampling (ABS) frame. However, at this time, impediments to their use include barriers to data sharing, inconsistent data categories and quality, and limited technology for accurate linking with other data sources.

First, obtaining permission to access government data sources often requires substantial time and effort, and may require major cultural changes in the way Americans think about their data and privacy. Currently, substantial barriers limit access to this information.

Second, variables in administrative databases would need to match the survey's variables of interest. For example, the database would need to utilize racial and ethnic categories that are relevant to the survey researchers. As previously mentioned, interest in and classification of particular groups and subgroups vary considerably across U.S.-based surveys.

Third, the quality of data entered in administrative databases varies widely. Data entry may be inconsistent in assigning racial or ethnic categories. For example, some may enter a category based on observation only; others, by asking a single question; and still others, by asking a series of questions to obtain maximum detail.

Fourth, if used as an auxiliary list, the method for linking a respondent's administrative data record to an individual respondent in the sample would need to be developed. A simple match of name and address may not provide adequate information, particularly if the target group is somewhat transient. Tracing can be conducted to verify matches between addresses and names, but this is expensive. Ideally, a unique identifier such as Social Security number could be used, but this information is closely guarded due to concerns about privacy and potential exposure to fraud.

In conclusion, the use of government administrative data for sampling has multiple implementation challenges. The accessibility and quality of the administrative data should be weighed against alternative methods for obtaining subgroup lists.

3.3 Summary

CHIS has considered and tested many methods for collecting representative data to support estimates for subpopulations and small geographic areas. These methods are evaluated after each data collection cycle and modified as needed, creating opportunities for advancing the survey field and optimizing the efficiency of the sample for these subgroups. Currently, CHIS uses a combination of lists and geographic oversampling to increase the sample size for Koreans and Vietnamese.

4.0 Methods for Addressing Coverage Issues Related to Random-Digit-Dial Methodology

RDD survey frames have typically lacked telephone numbers from cell phones. This lack of coverage is becoming increasingly problematic as more households are using cell phones only; the proportion of households using only cell phones increased from less than 8% in 2005 to more than 20% in 2008. More importantly, cell-phone-only status is more highly concentrated in households known to differ from the overall population with respect to health and health care access. The National Health Interview Survey (NHIS) found that persons living in households with only wireless telephones were more likely to be younger, have lower income, lack health insurance, and report being in good or excellent health.[15] This section reviews the implications of cell phone bias for CHIS as well as possible solutions.

4.1 CHIS Investigation of Cell Phone Bias

CHIS conducted a cell phone pilot test during the 2005 data collection. The pilot test explored the feasibility of covering cell phone households in future CHIS administrations by examining issues that had never been explored in dual-frame telephone surveys. Identified issues included:

- Completion of a screener that includes random selection of an adult
- Identification of cell-phone-only as well as cell-phone-mostly households
- Use of reimbursement for cell phone minutes
- Completion of a lengthy telephone interview via cell phone
- Assessment of the safety and privacy of the respondent's immediate environment
- Handling of mismatches between cell phone numbers and geographic areas
- Combining of the two samples to produce estimates.

When comparing the landline RDD sample with cell-phone-only and combination households, CHIS found that the cell-phone-only sample was younger, less likely to be married or own homes, and more likely to be male, employed, and living in a single-person household. The landline sample showed lower rates of risky health behaviors and better health care utilization and health insurance coverage. Nonresponse weighting reduced but did not eliminate this coverage bias. As a result, cell phone samples (both cell phone only and cell phone mostly) were included in 2007 and 2009 sample designs.

4.2 Alternatives

Discussants found no simple alternative to CHIS's RDD and cell phone sample frames. As previously mentioned, an address-based sample (ABS) frame would provide contact information for both landline and cell phone households, but telephone numbers would be unavailable for about 40% of addresses. Switching to an ABS frame would require additional recruitment methods, such as mail, which could affect trends and introduce biases that are not well defined.

Opt-in Panels. Another option is to use an online survey with an opt-in (volunteer) panel. However, these are typically nonrepresentative, especially of minority populations. Furthermore, the panels are reaching saturation—these panels are not large enough to yield sufficient numbers of respondents from California, especially for key subgroups. Moreover, because they are not probability based, these panels cannot generate estimates with known confidence intervals.

Probability-based Panel. Another alternative is a probability-based panel, such as that run by Knowledge Networks. This panel has recently switched from an RDD to an ABS frame, so biases against cell phone users are abating. However, the panel's size is insufficient to permit reliable inference for key subgroups in CHIS, and response rates tend to be below 20%, which is similar to the trend in CHIS response rates.

4.3 Summary

Evidence from multiple sources indicates that coverage issues related to cell phones could bias many CHIS estimates. CHIS is addressing these issues through the use of cell phone surveys. Although these are expensive and difficult to conduct, discussants found no readily available alternatives.

5.0 Methods for Addressing Possible Biases Associated With Low Response Rates

Although response rate is not a direct reflection of nonresponse bias, past attempts to minimize nonresponse bias have focused on maximizing response rates. This approach may no longer be feasible due to the 20-year decline in survey response rates. The decline in response to telephone surveys has been particularly steep, perhaps because technologies now allow a potential respondent to selectively screen out the survey. Other reasons include increased telemarketing, use of home fax machines and multiple landlines.

CHIS has responded to decreasing response rates by: (1) testing multiple methods for improving these rates; (2) attempting to quantify bias associated with nonresponse; and (3) continuing to incorporate nonresponse adjustments to the weights to mitigate potential nonresponse bias. Alternative methods for addressing low response rates are under discussion. The pertinent issues are described in this section.

5.1 Methods for Improving Telephone Survey Response Rates

CHIS has embedded experiments designed to test strategies to improve response rates. Randomized experiments have been conducted to investigate the initial presentation of the survey (i.e., mailing a notification, mailing a letter of support), interview techniques, local sponsorship, refusal conversion, and follow-up methods. The most effective practices have been incorporated into the next round of data collection.

Recent studies have suggested that even small incentives for survey respondents improve response rates (e.g., Brick et al, 2005).[18] Given the intensive follow-up that is necessary to modestly improve response rates, the incentive's costs could be partially offset by the savings in effort. Curtin et al. 2007 suggest this hypothesis but do not present data to address it; it would be useful to study this further. Some considerations in offering incentives are OMB's guidance on Federal agencies' use of survey incentives as well as researcher reluctance to use government or foundation funding to pay for them. More specifically, OMB requires agencies to justify the use of survey incentives and has encouraged experimental research on the effects of incentives on response rates and nonresponse bias. Workshop participants suggested that studies be conducted to assess whether costs for incentives would be smaller than costs for follow-up. They also suggested that these findings be presented to OMB and funders.

To summarize, because low response rates have become the norm, CHIS has compared estimates from subsamples for which standard and intensive participation strategies were used; little to no evidence of nonresponse bias has been revealed.[3,19] A number of embedded experiments have been performed and successful strategies, including the use of prepaid incentives, have been incorporated into CHIS procedures. Workshop participants encouraged CHIS to explore incentive options in greater detail.

5.2 Quantifying Bias Using Intensive Follow-Up

Working with experts in the survey research and methodology community, CHIS 2007 examined the effects of nonresponse in an address-based probability sample of persons who completed interviews in 908 Los Angeles County households. The sample of residential addresses was matched to telephone numbers, and those that were matched were contacted and interviewed by telephone using standard CHIS procedures. Two groups were targeted for in-person recruitment: (1) addresses without matching telephone numbers; and (2) addresses with matched telephone numbers that did not respond to telephone recruitment. Households in these two groups were recruited to participate by an in-person visit, but all interviews were conducted by telephone. The study yielded three groups: (1) persons recruited solely by telephone; (2) persons recruited by a household visit alone; and (3) persons who did not respond to initial telephone calls but were recruited through a household visit.

Comparisons of the telephone recruitment group with the initial nonresponse group (groups 1 and 3) revealed that nonrespondents were more likely to be younger, be Latino, live in households with children, and report lower levels of educational attainment and income. Statistical analyses revealed that these two groups differed with respect to one of nine health behavior indicators and with respect to four of 21 health care access and utilization indicators. After logistic regressions were performed for standard demographics (age, gender, race, ethnicity, education, home ownership, children) typically used during the weighting of CHIS data, only one difference remained statistically significant: tested for a sexually transmitted disease in the past 12 months. Researchers concluded that there was no evidence of significant, systematic nonresponse bias in CHIS. For additional details on this study, see http://www.chis.ucla.edu/pdf/dataquality3_doc.pdf.

5.3 Quantifying Bias Using Benchmarks

Benchmark comparisons allow researchers to verify that the characteristics of the sample that completed the survey are similar to the characteristics of individuals in a more inclusive database. Theoretically, if the survey in question provides estimates similar to estimates from other surveys or databases, bias is minimal.

In order to evaluate its representativeness, researchers have conducted benchmark comparisons of CHIS data with several sources, including the U.S. Census, Medi-Cal (the Medicaid program in California), the National Health Interview Survey (NHIS), and the Medical Expenditure Panel Survey (MEPS).

In comparisons of CHIS data with Census data at the neighborhood level, there was little to no substantial difference in neighborhood characteristics between respondents and nonrespondents. The response propensity of the CHIS sample was similarly distributed across these characteristics measured by the Census, so the projected nonresponse bias appeared to be very small.[3]

Findings from data comparisons between CHIS and NHIS, and between CHIS and MEPS, are more complex. First, the comparisons were difficult to implement due to differences in sampling, data collection mode, and weighting procedures. Second, large sample sizes allowed researchers to detect small, statistically significant differences, the implications of which need to be more fully understood. For example: How much of the difference is a “real” difference, indicating significant bias? What are the implications for the relationships between variables, if some variables appear to be more “biased” than others? Third, there were some statistically significant differences between NHIS and CHIS estimates, and more research is needed to understand why these differences exist.

CHIS's benchmark studies are building a foundation for development of methodologies to quantify bias and representativeness in survey samples. Such methodologies are important for improving the quality of survey data, particularly in an increasingly challenging data collection climate. Over time, such approaches can be used to develop more effective data collection strategies.

5.4 Mixed Mode—A Solution With Multiple Challenges

Methodologists are exploring multimode data collection as a way to improve quality and response rates and control costs. During the workshop, participants discussed which modes have been found most effective in improving response rates and containing costs.

Leading With a Mail Survey. Leading with a mail survey may be the most cost-effective approach for obtaining higher response rates using multimodal data collection, although the effect on nonresponse bias is less well understood.

Adding a mail survey to the Behavioral Risk Factor Surveillance System (BRFSS) telephone survey improved the response rate by 20%, although the respondents who returned the mail survey had characteristics similar to those of telephone respondents. On that basis, some researchers have suggested that the additional cases may have provided little protection against bias and may even have increased it. Adding a Web survey to BRFSS improved the response rate to a lesser degree.[20]

The Health Information National Trends Survey (HINTS), which collects nationally representative data about the American public's use of cancer-related information, tested a dual-frame RDD and area-based sample design. The test revealed higher initial response rates by mail when compared with initial telephone response rates. Interviewer follow-up after the mailing improved the response rate by several percentage points, whereas interactive voice response follow-up did not improve the telephone response rate at all.

A mixed-mode survey by Beebe et al[21] showed that a telephone follow-up interview after an initial mailing was particularly effective for obtaining complete interviews from members of ethnic subgroups.

Because mail and/or Web-based surveys are less expensive than telephone interviews (which typically require multiple call attempts), obtaining completed surveys initially using mail or the Web could decrease costs, allowing researchers to focus resources on recruiting reluctant respondents.

Complications Related to Mixed Modes. The manner in which multiple modes are presented to potential respondents affects response rates. Ample evidence across surveys indicates that providing respondents with an initial choice of mode is less effective than sequential mode assignment; with the latter approach, nonrespondents to one mode are given the opportunity to respond via another mode. The American Community Survey (ACS), for instance, follows the latter procedure.

Although mixed-mode methods appear to hold promise for improving low response rates, changing from a single-mode to a mixed-mode survey involves complex factors, including the coverage properties relative to the target population, sampling design effects, response rate characteristics, essential measurement properties, and mode effects. Thus, data collection modes are not simply alternative communication media; they also involve decisions about sampling frames, recruitment protocols, and instrument design.

For example, recruitment for telephone interviews typically requires a telephone number, which is not available for all address-based cases. Similarly, addresses are not universally available for telephone numbers on an RDD frame. This means that for a multimode survey, the portion of the sample with information supporting both modes has a greater opportunity to respond than other parts of the sample. Adding another mode could increase bias if individuals with dual information sources are unusual compared with the rest of the sample. More information is needed about “dual mode” bias and how it compares with coverage and nonresponse bias in single-mode studies.

Similarly, the random selection of household members that is standard in RDD is difficult to implement in online or mail surveys. More information is needed about whether the lack of randomization creates significant bias, particularly when compared with coverage and nonresponse bias associated with RDD studies.

In terms of instrument design, paper-based instruments do not permit the complex skip patterns that can be easily incorporated into computer-assisted telephone interview (CATI)-administered surveys. For example, after a negative response to a general CATI question, skip patterns allow interviewers to omit irrelevant questions about details. In contrast, paper surveys with skip patterns are often misread by respondents, creating extra burden and frustration for respondents and difficulties in interpreting data. Thus, mixed-mode surveys would need to be redesigned to minimize and/or simplify skip patterns. The resulting loss of detailed information should be weighed against the benefits of an improved response rate when deciding to switch to multiple-mode administration. Also, paper-based instruments do not easily accommodate multiple-language surveys such as CHIS.

Bias Related to Mode Effects. Little is known about bias related to mode effects. The presence of interviewers is known to affect responses to items that are socially sensitive, and self-administered modes (mail, online) tend to produce more accurate information for sensitive questions. Similarly, items presented in modes that are primarily visual (such as online or mail surveys) are likely to generate bias toward initial response categories, whereas items presented in aural modes (spoken interviews) are more likely to be biased toward the last-mentioned response category. To what degree will these biases affect the validity of resulting data? Dillman et al[22] and others have recommended strategies for developing survey items that are resistant to mode effects. Workshop discussants stated that there appears to be little between-mode difference in validity for nonsensitive factual questions, but more research is needed.

Understanding mode effects in mixed-mode survey data is difficult because individual differences and mode effects are confounded when response rates are low. Respondents appear to differ in their preferences for responding to visual vs. audio vs. face-to-face modes. When response rates are low, comparisons of individuals randomly assigned to groups with different data collection modes cannot distinguish the influence of mode from group differences related to self-selection. For example, individuals who are randomized to a telephone interview group and respond may have lower literacy than the nonrespondents in the telephone group. Alternatively, individuals who are randomized to an online survey and respond to it may have higher literacy than nonrespondents in the online group. Significant differences between the groups on variables related to health literacy and access may be attributable not to the influence of different modes, but instead to differences in the characteristics of respondents in the two groups.

To summarize, although mixed-mode data collection appears promising for improving low response rates, it is difficult to recommend until the bias introduced by mixed-mode designs is more fully understood. Bias associated with different modes should be evaluated, quantified, and compared with bias introduced by low response rates.

5.5 Reducing Survey Length to Increase Response Rates

It is well established that, in general, reducing a survey's length improves response rates. Workshop discussants briefly explored two main strategies for decreasing the length of the core survey. The first was a split-ballot sample design, in which participants receive different subsections of items from the full inventory of survey questions. However, this approach may dampen statistical power due to decreased sample sizes per item. Other implications of this alternative were not discussed.

Another option for shortening surveys is to supplement survey data with data obtained on individuals from other sources. For example, health care utilization data from CHIS core questions could be linked with government health care administrative data. As discussed in Section 3.2, this would require a profound shift in the U.S. public's attitude toward sharing private information, including medical information, as well as the resolution of significant barriers to data linking.

5.6 Imputing Data to Reduce the Effects of Nonresponse

In the event that missing data result in poor data quality, discussants suggested imputation as a way to increase sample sizes available for analysis. There are two types of missing data: (1) item-level missing data, where some but not all survey answers are provided; and (2) unit-level missing data, where the sampled person has provided no answers at all.

Item-level missing data in CHIS is generally very low (in the range of 1%–4%) and is routinely imputed for nearly every variable in the CHIS data files. Imputation for item-level missing data is commonly used in CHIS and other surveys and relies upon information provided across survey responses. In contrast, imputation is used relatively rarely to address unit-level nonresponse because it requires significant common information about nonrespondents and respondents. Researchers typically do not attempt imputation at the unit level unless extensive data are available through the sample frame or by matching with an administrative database. CHIS does not currently have information about unit-level nonrespondents, so it cannot carry out imputation for these missing units. If some type of administrative record becomes available for survey operations as indicated in Section 4.5 and can be linked to both respondents and non-respondents, imputation may become an option to consider.

5.7 Summary

Like other telephone surveys, CHIS is challenged by low response rates. Studies with CHIS and other surveys to investigate nonresponse bias have shown that these low response rates do not necessarily indicate significant systematic bias. More research is needed to understand the implications of statistically significant differences between CHIS data and benchmark survey data, as well as findings regarding characteristics of telephone survey nonrespondents. Initial findings from mixed-mode surveys indicate that higher response rates from these methods may not prevent bias. It is unclear how much bias mode effects may introduce; potential bias from mixed modes should be weighed against potential bias from low response rates. One distinct advantage of mixed-mode designs is that they may conserve funds, allowing targeted reallocation of resources to improve response rates in subpopulations that are difficult to reach. Research is still quite limited with respect to mixed modes, and it may be advisable to embed small-scale studies in CHIS to determine whether they result in improvements. Given the multiple methods for improving response rates reviewed at the workshop, discussants agreed that additional incentive options should be considered and could prove quite cost-effective.

Workshop participants briefly discussed using a split-ballot survey design with a rolling sample to reduce the survey's length, which might improve response rates. Using this design strategy would require careful attention to the handling of missing data and a larger sample size. Shortening surveys by supplementing them with data obtained from other sources (e.g., government health care administrative data) for those individuals is associated with substantial privacy and data linkage issues. Imputation is often used in CHIS and other surveys for item-level missing data, but CHIS is not able to impute unit-level missing data because information about unit-level nonrespondents is not available.

6.0 Conclusions

There are multiple factors influencing the representativeness of CHIS data. The CHIS Methods Workshop participants explored new pathways to strengthen and sustain CHIS. For example, CHIS and its data collection contractor, Westat, have systematically worked to test and document methods for improving response rates and evaluating possible bias.

In the past, researchers have viewed response rate as the only appropriate measure of representativeness. In this era of decreasing response rates for telephone surveys, additional evidence may need to be considered in determining whether survey estimates are based on an adequately representative sample. CHIS and Westat have been leaders in distinguishing types of bias and have identified coverage bias as the most important concern in terms of population representation for RDD surveys. CHIS and Westat have also explored innovative ways to survey respondents via cell phone and have assessed methods for integrating cell phone and landline data to produce valid and reliable statistical estimates.

More information is needed before best practices can be determined for surveys designed to collect data on specific racial/ethnic groups and small areas. Workshop participants encouraged UCLA and Westat to continue to embed methodological experiments in CHIS and carefully track results in published studies in order to systematically explore ways to resolve important issues.

Methodological research needs were identified in the following areas:

- Creating and testing alternatives to response rates for measuring representativeness
- Monitoring and evaluating possible alternatives to RDD sampling to sample small geographic areas
- Identifying the optimal combination of lists and geographic oversampling to produce representative estimates for specific ethnic subgroups
- Determining optimal incentives for reducing potential nonresponse bias
- Assessing representativeness of survey data obtained through mixed-mode vs. single-mode RDD methodologies
- Exploring options to reduce coverage bias, including RDD with cell phone and landline samples, address-based sampling, and other/multiple frames
- Exploring cost trade-offs to maintain or increase response rates, such as whether the costs of intensive follow-up needed to modestly improve response rates are offset by savings in follow-up costs when incentives are paid to respondents.

7.0 References

1. Edwards WS, Brick JM, Flores-Cervantes I, DiSogra C, Yen W. Sampling race and ethnic groups in RDD surveys. *Proc Surv Res Methods Sec Am Stat Assoc.* 2002;893-898.

2. California Health Interview Survey. CHIS survey design and methods: methodology reports. 2008. Available at: http://www.chis.ucla.edu/methodology.html. Accessed October 19, 2010.

3. Lee S, Brown ER, Grant D, Belin TR, Brick JM. Exploring nonresponse bias in a health survey using neighborhood characteristics. *Am J Public Health.* 2009;99(10):1811-1817.

4. California Health Interview Survey. CHIS 2007 methodology series report 2: data collection methods. Los Angeles, CA: UCLA Center for Health Policy Research; 2009.

5. Brick JM, Flores-Cervantes I, Lee S, Norman G. Nonsampling errors in dual-frame telephone surveys. *Surv Methodol*. 2010 (in press).

6. Lee S, Brick JM, Brown ER, Grant D. Growing cell-phone population and noncoverage bias in traditional random digit dial telephone health surveys. *Health Serv Res.* 2010;45(4):1121-1139.

7. Brick JM, Edwards WS, Lee S. Sampling telephone numbers and adults, interview length, and weighting in the California Health Interview Survey cell phone pilot study. *Public Opin Q.* 2007;71:793-813.

8. Edwards WS, Brick JM. Effects on response rates of multiple sponsors on advance letters for an RDD survey. Presented at the Annual Meeting of the American Association for Public Opinion Research, Montreal, Quebec, Canada; 2006.

9. Singer E, Van Hoewyk J, Maher MP. Experiments with incentives in telephone surveys. *Public Opin Q.* 2000;64(2):171-188.

10. Curtin R, Singer E, Presser S. Incentives in RDD telephone surveys: a replication and extension. *J Off Stat.* 2007;23(1):91-105.

11. Brown ER, Holtby S, Zahnd E, Abbott GB. Community-based participatory research in the California Health Interview Survey. *Prev Chronic Dis.* 2005;2(4):A03.

12. Groves RM. Nonresponse rates and nonresponse bias in household surveys. *Public Opin Q.* 2006;70(5):646-675.

13. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias. *Public Opin Q.* 2008;72(2):167-189.

14. Groves RM, Brick JM, Couper M, Kalsbeek W, Harrris-Kojetin B, Kreuter F, Pennell B, Raghunathan T, Schouten B, Smith T, Tourangeau R, Bowers A, Jans M, Kennedy C, Levenstein R, Olson K, Peytcheva E, Ziniel S, Wagner J. Issues facing the field: alternative practical measures of representativeness of survey respondent pools. October 30, 2008. Available at: http://surveypractice.files.wordpress.com/2008/11/survey-practice-october-2008.pdf. Accessed February 8, 2010.

15. Blumberg SJ, Luke JV. Wireless substitution: early release of estimates from the National Health Interview Survey, July - December 2008. May, 2009. Available at: http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200905.htm. Accessed January 21, 2010.

16. Rao JNK. *Small Area Estimation*. Hoboken, NJ: John Wiley & Sons, Inc.; 2003.

17. Yu H, Meng YY, Mendez-Luck CA, Jhawar M, Wallace SP. Small-area estimation of health insurance coverage for California legislative districts. *Am J Public Health.* 2007;97(4):731-737.

18. Brick JM, Montaquila J, Hagedorn MC, Brock Roth S, Chapman C. Implications for RDD design from an incentive experiment. *J Off Stat.* 2005;21(4):571-589.

19. California Health Interview Survey. CHIS 2007 area probability sample to assess nonresponse bias. *CHIS Working Paper Series*. Los Angeles, CA: UCLA Center for Health Policy Research; 2008.

20. Link MW, Mokdad AH. Can Web and mail survey modes improve participation in an RDD-based national health surveillance? *J Off Stat.* 2006;22(2):293-312.

21. Beebe TJ, Davern ME, McAlpine DD, Call KT, Rockwood TH. Increasing response rates in a survey of Medicaid enrollees: the effect of a prepaid monetary incentive and mixed modes (mail and telephone). *Med Care.* 2005;43(4):411-414.

22. Dillman DA, Smyth JD, Christian LM. *Internet, Mail, and Mixed-Mode Surveys: the Tailored Design Method*. Hoboken, NJ: John Wiley & Sons, Inc.; 2008.