Evaluation of Statistical Methods for Data Collection and Analysis on Racial and Ethnic Minorities and other Hard-to-Reach Populations

ABSTRACT

The ability to assess health disparities among relatively small race/ethnic groups is a major area of emphasis for the Department of Health and Human Services. Our project attempts to address this concern, having carried out a systematic literature review which suggests that while field methods may have limited payoff—in terms of increasing effective sample size—the largest and potentially most useful payoffs may result from sampling and analytic techniques. As such, the feasibility study phase of our project will consist of several sampling techniques including: geographic over-sampling (individual-based, micro-clusters, and/or macro-clusters) and overlapping frame designs (surname lists). We will also pilot analytic strategies such as "borrowing strength" from contiguous years to improve the efficiency and accuracy of estimates. While our ultimate goal is to increase effective sample size, careful attention will be paid to trade-offs with cost, end-user capability, and feasibility for current national health data sets.

INTRODUCTION

Health disparities are often defined as differences in incidence, prevalence, and mortality rates between disadvantaged and advantaged population groups. Elimination of known disparities is a major United States policy goal. However, most national health surveys have limited to no ability to adequately assess the health of major race/ethnic subgroups other than Mexican-Americans. The nation is in dire need of a comprehensive assessment of disparities in various subpopulations.

This paper addresses the relatively poor accuracy or complete lack of national prevalence estimates of the nation's smaller racial/ethnic subgroups--defined as those constituting approximately 1% each of the U.S. population. We have established a focus on Asian subpopulations and American Indians/Alaska Natives (AI/AN) given their relative neglect in national health surveys. While groups that are smaller than 1% of the population may be of interest, we believe that attaining adequate precision for groups constituting less than ½% of the U.S. population involves too difficult a leap to make from where we stand currently in terms of practically available sample size. As such, our goal is to increase precision for approximately 1% groups to approximately the same level of precision currently attainable for Asians as a whole (approximately 5% of the population), through a combination of design and analytic approaches. Simply scaling up the sample size of current surveys proportionately to obtain needed health information on hard-to-reach subpopulations is costly and inefficient given that most of the added sample size is wasted on large groups for which accuracy is already sufficient and where additional sample size produces very little benefit. The intent of this project is to assess the feasibility and potential application of various sampling strategies to the nationally representative health assessment of Asian and AI/AN subgroups in a cost-effective and accurate manner. We

also consider what implications these approaches might have for researchers seeking to analyze data collected using these techniques.

Section 1 discusses our systematic review of the literature. Section 2 describes the design of our feasibility study in which we will seek to test promising methods identified in the literature review. Section 3 [placeholder] explains the results of the feasibility study and Section 4 [placeholder] provides initial recommendations on cost-effective and statistically reliable approaches for deriving estimates of prevalence of various health behaviors, risk factors, and outcomes in small racial/ethnic and other hard-to-reach populations.

LITERATURE REVIEW

Methodology and Execution

In order to assess the means by which subgroup accuracy might be increased cost-effectively, we considered all three stages of the estimation process: sampling design methods; field methods; and analytic techniques. We employed a systematic review approach in gathering documents on the general topic of sampling and analyzing hard-to-reach populations. We searched a variety of library databases, surveyed experts in the field via a snowball sample to identify additional documents, and also gathered documents from project staff personal libraries. We hand-searched reference lists of documents for additional relevant citations. Several of our literature sources included "gray literature," that is unpublished documents or those that have limited distribution. We screened titles and abstracts, and ordered any documents that appeared relevant. We abstracted data on each relevant document via a short screener form. We then reviewed documents in detail that pertained to certain topics of interest.

Our searches produced 1866 titles of possibly relevant documents. The screening for relevance identified 453 documents to be obtained in full, of which we were able to obtain 441 (97%). Based on the in-depth screening, we were able to determine that of these 441 documents, 264 were relevant.

Summary of Design, Field, and Analytic Approaches

We deliberately cast a broad net with respect to sampling alternatives. We did not want to exclude any fruitful alternatives; on the other hand, we had to make several difficult decisions with respect to the potential payoff of some alternatives. We assume that the average end-user can successfully implement statistical software that accounts for aspects of complex survey design. We do note any additional burden on end-users when techniques assume additional knowledge.

Our discussion requires the definition of two key constructs that guide our decision-making: accuracy and cost-effectiveness. We describe accuracy in terms of effective sample size (ESS), as compared to standard analytic techniques applied to a simple random sample (SRS). Cost will be measured in dollars, so that our final assessment of the cost-effectiveness of a given technique will be based on the cost per ESS (CPESS), relative to standard analytic techniques applied to an SRS.

We discuss the major methods we encountered in sample design, field methods, and analytic techniques respectively. Since these three phases are independent, one could change or improve practice in more than one of these areas at once. The gains produced by improved sampling design will generally be applicable to all outcomes, but may involve trade-offs between accuracy in subpopulations and overall accuracy. In addition, some forms of complex sample design

require sophistication on the part of the end-user. On the other hand, the gains produced by improved field methods are typically applicable to all outcomes, do not typically involve trade-offs with the accuracy for the overall population, and typically do not complicate analysis by the end-user. Finally, the gains produced by improved analytic techniques typically vary by outcome and almost always require sophistication on the part of the end-user.

Conclusions of the Review of Sampling Design Methods

Some form of rotating over-sample, in which a rotational schedule targets specific subgroups for over-sampling on a periodic basis, is probably among the most promising approaches that we have identified. Screening, in which individuals are quickly interviewed to determine race/ethnicity, can be a cost-effective tool for lengthy health surveys when used in combination with other techniques. Expert and neighbor ratings to determine where to over-sample a particular subgroup are probably most effective in local communities. Their potential for national implementations is probably limited.

Over-sampling of geographic clusters containing high proportions of a target subpopulation is a very common and effective technique, which is often referred to as "geocoding." Over-sampling at the macro-cluster level may be somewhat helpful for both Asians and AI/AN. Over-sampling at the micro-cluster level is not likely to be very useful for Asians and AI/AN as a whole, but might be useful for Asian subgroups with greater micro-clustering. One exception that may be worth investigating is sampling multiple eligible individuals within households. Adaptive techniques, which over-sample via within-cluster allocation on the basis of information that is not available a priori may be helpful if micro-clustering in units for which data is not available a priori can be identified (e.g. apartment complexes).

Overlapping sampling frames that are generated a priori have potential if they have very good sensitivity and positive predictive validity. Surname lists can be somewhat useful in this regard for some subgroups, such as Chinese, Japanese, and Vietnamese, especially if they are applied to administrative lists with good coverage of the population. Network sampling, which begins with a "seed" probability sample of appreciable size ("nodes"), and each "node" provides "links," names of others with defined relationships to the node, is a promising technique that has been shown to be very cost-effective in some applications, including some involving race-ethnic subgroups.

Snowball sampling and multiplicity sampling are unlikely to achieve the goals of this project.

Conclusions of the Review of Field Methods

Independent of sampling issues, there may be unique methodological difficulties in surveying racial/ethnic minorities. Low response rates, high mobility, language obstacles, and the cross-cultural validity of survey instruments are among the most common encountered difficulties. Special considerations of the socio-cultural characteristics of the minority group or community while planning and implementing a survey may mitigate many of the difficulties that are commonly encountered. Both identifying community leaders and involving the community as a whole in the research are important steps toward this end. Culture-specific field methods have the potential to substantially improve cost-effectiveness independent of other improvements. On the other hand, their applicability to an ongoing, national survey may be limited.

Conclusions of the Review of Analytic Methods

The most relevant analytic method for this project is small area estimation (SAE) and the various inferential methods that are used in connection with SAE, such as empirical Bayes estimation. This method seeks to provide estimates for a variety of domains, which can be geographic areas or socio-demographic groups or other subpopulations. The expression small area (or local area) traditionally denotes a small geographical area, such as a county, or municipality, but can also describe small domains that are not geographic (e.g. a specific age-sex-race group in a large geographical area).

Aggregation over time, including using SAE to combine over time may be useful in combination with oversampling. SAE may also be helpful in producing estimates for small regions. Shrinking, smoothing, or pooling sampling weights, may increase cost-effectiveness, but these techniques are likely to be outcome-specific and dependent on sophistication of the end-user. Weighted surveys might be examined for instances in which weights can be appropriately collapsed or shrunk before reaching end-users.

FEASIBILITY STUDY: DESIGN

The literature review revealed a range of sampling design, field, and analytic methods that could help make data collection and national health estimates for very small racial/ethnic subgroups more precise. We will now attempt to demonstrate the feasibility of the most promising approaches identified in our systematic literature review. The choice of subgroups in this study is primarily motivated by both feasibility and the extent to which subgroups represent sampling and analytic challenges that are representative of other such subgroups, rather than being motivated by an assessment of the relative health needs of these subgroups.

Data and Methods

Our analysis will focus exclusively on secondary data analyses and data simulations that we can undertake through the use of publicly available data sets. We will conduct a portion of our pilot analyses using several years of the National Health Interview Surveys (NHIS). Our choice of this data set is based on the fact that the NHIS is used as the primary sampling frame for many of the other NCHS health data sets and contains the largest sample of several small race/ethnic groups over a long period of time. Since 1957, the NHIS has continuously conducted nationwide household interviews to collect information concerning the health of the U.S. civilian non-institutionalized population, making it the best dataset available for national estimates of health by race/ethnicity. The survey collects information on race/ethnicity, socioeconomic characteristics, and self-reported health status.

Demonstration Racial/Ethnic Groups for Feasibility Study

We chose two promising groups for the feasibility study, based on their overall generalizability to other groups and the availability of data for the present quantitative exercise. In short, if we can demonstrate viable methods for one race/ethnic group, we want this group to have the largest and broadest applicability to other groups. This general applicability is based on several field and sampling factors, including: the level of micro-geographic clustering, the level of macro-geographic clustering (at roughly the census tract or larger areas), the level of mobility, whether or not surname analysis is a viable tool, population size, and barriers to high response rates (including language barriers and legal status). Table 1 displays a number of sampling-related characteristics of selected small racial-ethnic subgroups. As a result of considering both

potential generalizability and feasibility based on these characteristics, we have chosen to focus on Chinese-Americans and American Indians to demonstrate our feasibility study.

Table 1: Sampling Characteristics of Selected Racial-Ethnic Subgroups

Racial/Ethnic Subgroup	NHIS Annual Sample Size (2000-2002 Average, except where noted)	Strong Macro-Geographic Clustering	Strong Micro-Geographic Clustering	Highly Mobile	Surname Analysis Viable	Proportion of US Population (YR 2000)	Language Barriers Likely
AI/AN	629	Yes: 30% live in reservations or blocks with an AI/AN concentration>60%	N/A	Yes	No	~1%	No
Chinese	684	Yes- Urban, certain states	No: 43% in census block with <% density	No	Yes	~1%	No
Filipino	598	N/A)	No: 50% in census block with <5% density	N/A	No	0.84%	No
Vietnamese	145*	Yes- Urban, certain states	No: 52% in census block with <5% density	N/A	Yes	0.43%	Yes
Korean	116*	Yes- Urban, certain states	N/A	N/A	Yes	0.44%	N/A
Asian Indian	589	N/A	N/A	N/A	Somewhat	0.68%	No

*1992-1993 average; only available from 1992-1995; substantially smaller samples 1994-1995 than 1992-1993.

Our choice of Chinese-Americans for the feasibility study is based on the observations that they represent approximately 1% of the population, are a tenable sub-population for surname analysis, exhibit both macro-geographic clustering in their residential patterns, and have been consistently identified as a separate race/ethnic subgroup for several years of the NHIS (1992-2002). American Indians also make up approximately 1% of the population, are macro-geographically clustered, and have been continuously identified in the NHIS. In addition, they are relatively mobile, a characteristic shared with other minority subpopulations (e.g. Hispanics and recent immigrants) that poses unique sampling challenges.

Cost Scenarios

We will explore at least two different cost-scenarios. A *supplemented cost* approach will add additional sample (targeted subpopulations) without any corresponding reduction in other sample sizes, making for a more expensive survey than a reference survey. In a *cost-equivalent* (or "carve-out") approach, the costs associated with these design changes will be balanced by a reduction in the non-Hispanic white sample size, trading off a small reduction in the excellent precision of both overall and non-Hispanic white estimates for a substantial improvement in estimates of disparities between non-Hispanic whites and these very small subgroups. It should be noted that such an approach would have the same cost as a reference survey, but that changes to an existing survey will also incur transition costs. Finally, we may explore a *third cost-neutral* approach that will estimate transition costs and accommodate both the supplemental sample and transition costs with even larger reductions of the non-Hispanic white sample.

Techniques for Feasibility Study

The current nominal sample size for NHIS is roughly 100,000 completed interviews, containing approximately 75,000 White, 14,000 Black/African-American, 700 Chinese (0.70%), 650 American Indian, and 5,600 others. Of these, approximately 21,000 have Hispanic origins. Our goal is to increase the effective sample size (ESS)—that is, the nominal sample size divided by the design effect—by a factor of five to ten for Chinese and American Indians.

In a "carve-out" approach (cost-equivalent or cost-neutral), costs for increasing the ESS for the two subgroups will come from reductions in the non-Hispanic white sample size. We wish to make this reduction as small as possible (as an additional goal we might limit the reduction size of this group to no more than 15%), which means we want to increase the ESS as cost-effectively as possible. In a supplemental cost approach, we want to minimize the cost of the supplement. Making these trade-offs requires a detailed understanding of NHIS cost structure, including fixed, marginal, and transition costs

The magnitude of increase in ESS that we seek is too great to be reasonably achieved by a single approach, or even by a combination of design approaches alone. We believe that a combination of design and analytic approaches will be necessary to achieve these goals.

Specific Techniques Considered

To reiterate, we have considered in greater depth several techniques identified as promising in our systematic review of the literature. We initially conceived of the estimation problem as having three parts: design (whom we target and how), field implementation (improving response rates or reducing costs within a given sampling design), and analysis (improving precision with a given set of data).

Field implementation

While there are clearly means to improve field implementation, such as contacts with local community leaders about the research and culturally sensitive approaches to administering the instrument, the literature provides little quantitative detail about the extent of improvements that might be expected in a national implementation. Furthermore, many of the techniques are not generalizable beyond a particular mode of administration. Finally, these techniques are not particularly amenable to simulation within the feasibility study. For these reasons, our approach does not include additional exploration of field implementation techniques.

Analysis

One analytic approach emerged as most promising: small-area estimation (SAE). In particular, using SAE to pool data over time, while not sufficient in the absence of design change, was identified as likely to be an important and perhaps necessary supplement to changes in design. In this approach, estimates for a given year will borrow strength from perhaps two to six adjacent years. We will use a composite estimator that includes adjacent years with weights determined by the extent to which data from adjacent years are correlated with data from the year in question. These weights will be assigned on an outcome-by-outcome basis. Outcomes that are very stable over time will result in weights for adjacent years that are nearly equivalent to weights in the "target" year, which in turn will mean that the contributions of these observations to the precision of the estimate will be nearly as great as the contributions of observations from the target year. If the adjacent years are fully equivalent to the target year, the reduction in

mean-squared-error (MSE) from using SAE to pool over three to seven years will be equivalent to a three to seven-fold increase in sample size.

We will use NHIS data to empirically determine where on the theoretical range of no improvement to three- to seven-fold multiplication of effective sample size (ESS) actual improvements lie for each of several NHIS health outcomes. Given current sample sizes for the target racial/ethnic subgroups, estimates of these and other parameters may need to borrow strength from estimates in other racial/ethnic groups for reliability in the present exercise.

We recognize that analytic techniques are ultimately implemented by end-users, but we feel that they may be an important part of the feasibility study for at least two reasons. First, knowing the extent of potential improvement informs the extent of design change that is needed. Second, we hope to provide a model for using these techniques in this context for end-users not currently familiar with this approach.

Sample design

Four techniques emerged as particularly promising within the area of design: oversampling and screening at the household level, oversampling at the macro-geographic level, oversampling from an incomplete frame with substantial coverage of the population, and network sampling.

Oversampling and screening at the household level implies a sampling strategy that will include all sampled households containing the target subgroup, but only portions of sampled households not containing members of target subgroups, once race/ethnicity has been determined. NHIS currently employs this approach to oversample Blacks and Hispanics.

Oversampling at the macro-geographic level implies allocating additional sample to census tracts or larger geographic units that contain particularly high densities of the target subgroups.

This is also an approach currently employed by NHIS with respect to Blacks and Hispanics. In the case of American Indians, reservations contain a substantial proportion of the total population and contain very few members outside the target group. In the case of Chinese, some urban census tracts contain relatively high proportions of the target group, but this technique holds less promise for American Indians and probably less promise than for Blacks. This technique requires the use of design weights to account for the overrepresentation of members of the target group who reside in areas where the target group is most prevalent relative to those members of the target group who reside in other areas.

An incomplete frame is a list that is not by itself representative of the target population. If an incomplete frame contains a substantial proportion of the target population, one may use it effectively in conjunction with a complete frame. In this application, we will supplement observations from the complete frame with observations from the incomplete frame. We will then define two strata: those members of the population listed on the incomplete frame and those members of the population not listed on the incomplete frame. This results in a disproportionate stratified random sample, oversampling the stratum corresponding to being listed on the incomplete frame. All observations that exist in the incomplete frame will be treated identically, regardless of whether they were obtained from the complete frame or the incomplete frame. In the present application, standard techniques produce a complete frame and commercial marketing lists may provide effective incomplete frames for Chinese and American Indians.

The best potential frame of which we are aware is the American Community Survey (ACS), a new survey collected by the Census Bureau that will soon take the place of the Census Long Form. To the extent that privacy concerns and regulations will allow the use of race/ethnic and address information, this will be a representative frame that can supplement certain race/ethnic

groups very effectively, given the possibility of relatively recent address information. (The Census Bureau will release information at the level of census tracts approximately 5 years after each ACS survey. We understand that NCHS is currently exploring this possibility in a single state.)

Our investigations have also identified other possible incomplete frames if the ACS is not available. The Bureau of Indian Affairs may be able to supply tribal rolls for Native Americans. We have also uncovered marketing lists (e.g. SSI) that claim to contain 35% of the entire Asian population in the US. We are investigating the extent to which these lists specifically identify Chinese. In the event they do not, surname list techniques can further refine these lists, since Chinese surname lists are among the most predictive.

Network sampling is a probability sampling technique that supplements members of the target subgroup obtained by traditional means with observations from defined networks associated with each of the original "seed" respondents. It is critical that networks be precisely defined, so that seed respondents know exactly how many people meet the network definition in their case. In our application, a network may consist of children, siblings, and parents. To the extent that this approach inexpensively identifies other members of the target subgroup not residing in the same household as the seed respondent, it has potential to be cost-effective.

We will use a combination of two design approaches for American Indians and a combination of three techniques for Chinese. We view oversampling and screening at the household level as a necessary components of oversampling both American Indian and Chinese. We will use Census data to estimate the extent of screening that will be necessary as a function of the prevalence of the target subgroup in the sampled areas. We will employ oversampling of reservations in the case of American Indians and oversampling of census tracts in the case of

Chinese, using Census data to estimate design effects. We will use marketing lists to oversample Chinese, as described above. We will estimate design effects by evaluating claims of coverage to the extent this is possible and will also employ published estimates of the performance of Chinese surname lists, if applicable.

Estimating the cost-effectiveness of network sampling is particularly difficult, as it is dependent upon many quantities that are difficult to estimate. These quantities include the proportion of the network not residing in the same household as the seed respondent, the average network size, the intra-class correlation of health outcomes for network members not residing together, and finally the comfort of seed respondents with network referrals and the resultant effects on response rates and accuracy of network listings. Although network sampling has been considered for several large applications, these uncertainties have ultimately hindered its implementation. Similarly, we feel that it will be difficult to provide estimates of the improvements that might result from network sampling with a great degree of precision and confidence and therefore will not pursue this technique in the feasibility study.

We will pursue an exploratory effort in which we build multivariate models of health outcomes within our target subgroups that include oversampling indicators corresponding to marketing lists, reservations, or oversampled census tracts as predictors, using NHIS data. This will clarify the extent to which the oversampled strata are unrepresentative with respect to the outcomes of interest, and will allow estimation of the bias present in unweighted estimates. We will build additional models that add covariates intended to account for differences between oversampled and undersampled members of the target population with respect to health outcomes. These variables will include measures of socio-economic status (e.g. family income, education level, employment and marital status) and other demographic characteristics (e.g.

nativity, urban versus rural residency). If these covariates rendered the oversampling strata unpredictive of certain health outcomes, this will lay the groundwork for considering observations exchangeable across strata, conditional upon the covariates. Exchangeability will in turn imply that design weights and the corresponding design effects were ignorable for specific models.

FEASIBILITY STUDY: RESULTS

[placeholder]

- Description of specific approaches for Chinese and American Indian subgroups
- Sensitivity analysis that conveys the circumstances under which the techniques are likely to be effective
- Estimates of the increase in precision expected for each of several outcomes
- Discussion of the applicability of these techniques to other racial/ethnic subgroups
- Estimates of the reduction in overall precision and precision within the non-Hispanic White subgroup, given a cost-neutral approach
- Cost assumptions and discussions of two to three cost scenarios

Techniques employed will include:

- multivariate modeling of NHIS and Census data, including SAE
- Monte-Carlo simulation
- power calculations

RECOMMENDATIONS

[placeholder]