Special Sampling Issues

National Probability Samples in Studies of Low-Prevalence Diseases. Part I: Perspectives and Lessons from the HIV Cost and Services Utilization Study

Martin F. Shapiro, Marc L. Berk, Sandra H. Berry, Carol-Ann Emmons, Leslie A. Athey, David C. Hsia, Arleen A. Leibowitz, Carl A. Maida, Marvin Marcus, Judith F. Perlman, Claudia L. Schur, Mark A. Schuster, J. Walton Senterfitt, and Samuel A. Bozzette

Objective. To examine the trade-offs inherent in selecting a sample design for a national study of care for an uncommon disease, and the adaptations, opportunities and costs associated with the choice of national probability sampling in a study of HIV/AIDS.

Setting. A consortium of public and private funders, research organizations, community advocates, and local providers assembled to design and execute the study.

Design. Data collected by providers or collected for administrative purposes are limited by selectivity and concerns about validity. In studies based on convenience sampling, generalizability is uncertain. Multistage probability sampling through households may not produce sufficient cases of diseases that are not highly prevalent. In such cases, an attractive alternative design is multistage probability sampling through sites of care, in which all persons in the reference population have some chance of random selection through their medical providers, and in which included subjects are selected with known probability.

Data Collection and Principal Findings. Multistage national probability sampling through providers supplies uniquely valuable information, but will not represent populations not receiving medical care and may not provide sufficient cases in subpopulations of interest. Factors contributing to the substantial cost of such a design include the need to develop a sampling frame, the problems associated with recruitment of providers and subjects through medical providers, the need for buy-in from persons affected by the disease and their medical practitioners, as well as the need for a high participation rate. Broad representation from the national community of scholars with relevant expertise is desirable. Special problems are associated with organization of the research effort, with instrument development, and with data analysis and dissemination in such a consortium.

Conclusions. Multistage probability sampling through providers can provide unbiased, nationally representative data on persons receiving regular medical care for uncommon diseases and can improve our ability to accurately study care and its outcomes for diseases such as HIV/AIDS. However, substantial costs and special circumstances are associated with the implementation of such efforts.

Key Words. Probability sampling, health services research, low-prevalence sample designs

---

Health services research must assess the effects of changes in the healthcare system on specific populations in terms of their impact on access to care and care efficiency (Andersen 1995), on the extent to which clinical, social, and resource needs are being met (Marx et al. 1997), and on health outcomes (Lehman 1995). In undertaking a national study of care for persons with HIV, we evaluated a series of approaches that have quite different implications in terms of cost and ability to inform public policy.

For the reasons detailed in this article, the HIV Cost and Services Utilization Study (HCSUS), funded through a cooperative agreement with

---

The HIV Cost and Services Utilization Study is being conducted under cooperative agreement HS08578 (M. F. Shapiro, Principal Investigator; S. A. Bozzette, Co-Principal Investigator) between RAND and the Agency for Health Care Policy and Research. Substantial additional support for this agreement was provided by the Health Resources and Services Administration, the National Institute for Mental Health, the National Institute for Drug Abuse, and the National Institutes of Health Office of Research on Minority Health through the National Institute for Dental Research. Additional support was provided by the Robert Wood Johnson Foundation, Merck and Company, Glaxo-Wellcome, and the National Institute on Aging.

Martin F. Shapiro, M.D., Ph.D. is Chief, Division of General Internal Medicine and Health Services Research, Department of Medicine, the University of California, Los Angeles; and a Senior Natural Scientist at RAND Health. Samuel A. Bozzette, M.D., Ph.D. is Professor, Department of Medicine, University of California, San Diego; Health Services Research and Development Senior Research Associate of the Department of Veterans Affairs and the Veterans Affairs San Diego Healthcare System; and a Senior Natural Scientist at RAND Health. Address correspondence and requests for reprints to Drs. Shapiro and Bozzette, HCSUS Office, RAND, 1700 Main Street, Santa Monica, CA 90407-2138. Marc L. Berk, Ph.D. and Claudia L. Schur, Ph.D. are from the Project HOPE Center for Health Affairs. Sandra H. Berry, M.A., Arleen A. Leibowitz, Ph.D., Judith F. Perlman, M.A., and Mark A. Schuster, M.D., Ph.D., and J. Walton Senterfitt, M.P.H. are from RAND Health. Carol-Ann Emmons, Ph.D., Dr. Leibowitz, Carl A. Maida, Ph.D., Marvin Marcus, D.D.S., M.P.H., and Dr. Schuster are from UCLA. Leslie A. Athey, M. S. and Dr. Emmons are from the National Opinion Research Center. David C. Hsia, J.D., M.D., M.P.H. is from the Agency for Health Care Policy and Research. This article, submitted to *Health Services Research* on January 6, 1998, was revised and accepted for publication on December 14, 1998.

the Agency for Health Care Policy and Research (AHCPR), adopted a national probability sampling design with the intent of producing a nationally representative sample of HIV-infected persons receiving regular medical care. The nationally representative cohort of about 3,000 patients has the potential to produce data that are valuable in their own right and provide a context for interpreting data obtained using other designs. Because of this, the HCSUS model is relevant to the study of other diseases of policy importance.

The experience of conducting the study has allowed us to identify much of what is gained and lost by this approach. In the following sections, we review the major design options for national studies of medical care for specific diseases. We discuss the strengths and limitations of the convenience samples usually employed by such studies and attempt to familiarize health services researchers with the strengths and limitations of national probability sampling. We describe some of what we have learned about conducting such a study. We also discuss other issues relevant to the design we selected: the desirability of involving both a national network of investigators and the community affected by the disease, and the need to collect, report, and disseminate data from sick patients in a very large number of locales. A companion article (Frankel, Shapiro, Duan, et al. 1999) documents the specifics of the HCSUS sample design and reviews the essential features that made implementation of the design feasible.

POTENTIAL SAMPLE DESIGNS FOR HEALTH SERVICES RESEARCH

The HCSUS is a prospective observational study that is collecting data through serial interviews with patients; interviews with providers and caregivers; and abstraction of medical, dental, pharmacy, and billing records. As outlined in Table 1, the aims of the HCSUS are broad. Accordingly, the HCSUS consortium considered a broad range of approaches to acquiring data on the experience of care in different health systems. Table 2 summarizes the advantages and disadvantages of each approach.

Data collected by health plans (Harris, Hanes, Jimison, et al. 1997; Iezzoni et al. 1996; Welch and Welch 1995), in response to regulatory demands (Grimaldi 1997, O'Malley 1997), has great value for answering some questions of importance to the plans. It can provide insights quite efficiently into whether much variability exists in compliance with guidelines by plan or by patient characteristics. The major problems with such studies are the uncertain or uneven quality of data (National Committee for Quality Assurance

Table 1: Key Aims of the HIV Cost and Services Utilization Study

1.	To provide national estimates of the population under care for HIV, and to examine the utilization and costs of medical and nonmedical services and variations in utilization and costs across geographic locations, across healthcare systems, and by patients' clinical and demographic characteristics
2.	To examine variations in access to care, unmet needs for services, health-related quality of life, social support, quality of care, and satisfaction with care by geographic location, type of healthcare system, patient characteristics, and the relationship of these variables to utilization, service mix, costs, and survival
3.	To examine the relationship of severity and stage of illness to other patient characteristics and to determine how these relate to costs and utilization
4.	To determine the settings in which HIV care is delivered, to examine the transitions in principal provider of care for HIV infection, and to evaluate differences in treatment between HMOs and other types of care
5.	To conduct special studies of HIV in rural areas; of the need, use, and adequacy of care for mental health and oral health problems; of access to experimental and costly drugs; of formal and informal caregivers; of medical providers; and of the use of alternative HIV therapies

[NCQA] 1997), the limited information on specific diseases, and the lack of generalizability to other settings.

Multistage probability-based samples of households such as those used by the National Health Interview Survey (NHIS) produce nationally-representative data, but are unlikely to provide sufficiently large samples of persons with most disease of interest and are not designed to collect detailed data about specific diseases. For these reasons, even the NHIS, which interviews more than 100,000 persons per year, cannot meet the needs of policymakers interested in HIV, autism, or any number of other specific diseases. A modification of this approach, using a household survey to screen for a specific disease, is a theoretically sound idea but would be fantastically expensive. The NHIS would need to screen about one million persons in an unstratified household sample to identify 1,000 persons with a diagnosis of AIDS. This approach may be further complicated by excessive noncooperation or even political difficulty in the case of a stigmatizing disease such as HIV.

A third approach is to study a convenience or purposive sample of subjects. In many "outcomes" studies, these sites are selected to be demographically diverse at locations known to the investigators, or at sites that submit successful proposals for participation in the study. As with studies based on data generated by health plans, the sample size is usually sufficient

Table 2: Advantages and Disadvantages of Alternative Sample Designs for Health Services Research

Type of Study	*Advantages*	*Disadvantages*
Data provided by plan	Inexpensive Provide insights into plan performance	Uncertain data quality May lack detail about specific diseases
Household probablilty sample	Data are representative	Unlikely to provide adequate cases for studies of specific diseases
Studies of convenience/purposive samples	Sample size usually adequate for key analyses within specific desease entities Can be fielded rapidly, relatively inexpensively Useful for comparisons across sites, hypothesis generation, studies of very small/rare populations	Representativeness is unknown, making inference problematic, even if sites are selected to reflect national demographics Selected sites may be more familiar with standardized treatments, social services, etc.
Probability sampling through randomly selected providers	Data are representative	See text

to allow comparison of subgroups of patients across categories such as cost, utilization, and outcomes (Ware, Bayliss, Rogers, et al. 1996). They can be implemented more easily, more rapidly, and less expensively than probability samples, and they are very useful for making comparisons within and across sites. Consequently, they are an invaluable means of developing hypotheses about phenomena that can later be studied in experiments, more generalizable populations, or both. They also may produce the most adequate sample when the population of interest is small or when the number of sites that can be sampled is very small.

However, inferences to broader populations of interest are limited because the study population may not be representative of these larger populations (Thomas, Stoyva, Rosenberg, et al. 1997). Moreover, the degree to which they are unrepresentative is unknown and is often masked by other elements of the design. The enrollment of a large number of subjects in a convenience sample can produce a narrow confidence interval around the estimate of a mean, while the relation of the sample mean to the population mean is unknown. Attempts to be more reflective of the national demographics of the disease under study by purposely selecting providers in multiple cities,

regions, or delivery systems and assuring that key subpopulations are included may yield a sample that appears to be representative. However, it is likely to be quite biased if the range of included areas and practices is biased. That is, from the perspective of total survey error, this is not accuracy but rather an illusion of accuracy: the lack of a probabilistic link between the study and the target populations means that the difference between the study values and the values in the target population is unknown and possibly large. Randomly sampling within sites assures that subjects are representative of the sites, but it does not address the issue of the site's representativeness.

The AIDS Cost and Services Utilization Study, or ACSUS (Hellinger, Fleishman, and Hsia 1994), which randomly sampled HIV patients at 26 purposely sampled providers in ten large cities, provided much important information, but ACSUS investigators concluded that coverage bias was probably the most important source of error in the study because many segments of the HIV population had no chance of being selected into the study (Berk, Maffeo, and Schur 1993). A convenience sample of providers is likely to be quite different from the increasingly broad spectrum of such providers. For example, they may be more likely to be aware of standardized clinical or available research protocols, obscuring important relationships between patient characteristics and medical care, and leaving unresolved the issue of whether any phenomena observed among a particular subgroup of patients (e.g., HMO enrollees or African Americans) reflected more general phenomena or were particular to the participating sites.

A fourth approach is to use probability-based sampling to identify cases through randomly selected providers rather than households and to enroll patients with the disease of interest rather than residents screening positive. This approach, which was adopted by HCSUS, uses multistage scientific sampling to randomly select geographic locations in the first stage, providers in those locations at the second stage, and patients of those providers in the third stage from comprehensive lists of areas, providers, and patients, respectively. The central notion of the HCSUS design is that this can be accomplished in a way that ensures that (a) coverage problems are solved because all persons in the reference population had some theoretical chance of selection, and (b) representativeness problems are solved because knowing the probability of selection for all included subjects allows the construction of sampling weights. The essential elements of this approach are that investigators must be able to identify providers and cases, and to construct valid sampling weights. The latter are the direct quantitative link to the population of interest and ensure that the weighted data from the sample are representative of that entire reference population.

POTENTIAL VALUE AND LIMITATIONS OF THE NATIONAL PROBABILITY SAMPLE MODEL

When designing the HCSUS, we concluded that the large public investment planned for the project[1] could be justified only if it produced the unbiased information needed to make difficult policy decisions. It became obvious that such unbiased information could not be produced at lesser cost: nearly 80 percent of the cost of HCSUS went to activities associated with assembling the sample, developing the instruments, collecting and processing data, and getting buy-in from relevant groups.

At the same time, this approach has important limitations. First, in a probability sample of a large enough overall size, adequate numbers of subjects from less common subpopulations may not be present. This is particularly true in the case of "self-weighting" or Equal Probability of Selection Method (EPSEM) samples (Kish 1965). In some cases, increasing the overall sample size to increase the numbers of subjects in key subpopulations is sensible, but often the concomitant increases in other subgroups will increase costs unreasonably. The alternative is stratified sampling with the application of differential probabilities of selection, but increasing the probability of selection for elements in a small subpopulation or stratum increases the variance of the "overall" estimate for the entire population. The statistical implications of this approach are well discussed in the basic sampling literature (Kish 1965). For still less common subpopulations, it may be possible to oversample heavily from strata within which the subpopulations of interest are not sparse, but studies purposely focused on sites where the subjects can be readily identified may make more sense.

In the case of the HCSUS, there was a particular desire for the study to include adequate numbers of populations such as (1) injection drug users, (2) women, (3) persons receiving care through managed care organizations, (4) rural residents, and (5) racial/ethnic minorities. Injection drug users, African Americans, and Latinos were well represented in the reference population and required specialized sampling procedures. Representation of women and patients of staff-model HMOs was enhanced by doubling the probability with which they were sampled. In order to ensure an adequate number of rural cases for analysis, we constructed a separate national rural sampling frame. But not even these strategies could be successfully implemented at a tolerable cost for certain ethnic minorities, such as Asians and Pacific Islanders, because of their extreme scarcity in the population under care.

Second, a sampling frame needs to be constructed. Comprehensive rosters of patients with specific diseases do not exist in the United States or in most other countries. The HCSUS decision to limit the sampling frame to persons receiving ongoing medical care for their condition, as detailed in the accompanying paper by Frankel, Shapiro, Duan, et al. (1999), makes it possible to build a sampling list from a comparatively small number of doctors, but it does limit the generalizability of the findings. It does not represent persons who are not receiving care, but it does capture a group that is highly relevant to policy: those using the healthcare system. However, those outside of the system have poor access by definition, and should be taken into account in studies in which access is a key variable. This is less of a problem for other research topics such as cost: the amount of utilization that is missed in sampling the population under care is small because those missed are low utilizers.

The characteristics of the population that is missed by provider-based sampling is likely to vary across chronic diseases. In the case of a chronic progressive disease such as HIV/AIDS, persons with early asymptomatic disease are much more likely to be out of care than are those in the advanced stages of the illness. Very infrequent users of care are an intermediate group. They have some chance of selection, but it is much smaller than that of more frequent users. Data on those who are selected can be used to generate estimates that include other, comparable patients. In HCSUS, we used data from patients seen in a two-month period to generate estimates of the numbers and characteristics of subjects who would have been sampled over a longer period (Bozzette, Berry, Duan, et al. 1998).

Third, even if sampling is limited to a population in care, the expense of assembling the sampling frame, then of enrolling a disparate sample, is an important practical limitation. Investigators must obtain estimates of the size of the population of such persons in locales of interest, establish a careful and complex sampling operation, convince providers and patients who are not in the habit of volunteering for studies to participate, obtain approval from a large number of Institutional Review Boards, and undertake a nationwide data collection effort. The additional expense of a probability sample can be justified only if the resulting information is of high quality.

Fourth, investigators must be prepared to spend a great deal of time on the critical task of convincing providers to participate and on encouraging them to provide patient lists for sampling. At least in the case of stigmatizing diseases, human subjects' concerns are likely to prevent investigators from directly obtaining the names of persons in the sampling frame (Alonzo and

Reynolds 1995). Under these circumstances, investigators must also depend on providers, at a minimum, to contact sampled patients and present the study to patients in a way that maximizes their likelihood of being willing to talk to project staff. Even when providers elicit agreement from patients to be contacted by the study, there is no certainty that they will be located, enrolled, and interviewed. For this reason, studies employing provider-based national probability designs simply cannot be executed nearly as quickly as convenience studies in which providers are chosen because they wish to cooperate with the study. In the case of HCSUS, 15 months elapsed from the day on which the first patient was sampled to the day on which the last sampled patient was interviewed.

Fifth, there is no point in conducting a probability sample study without an adequate response rate, in the absence of which the study degenerates into a convenience sample study of those willing to be interviewed. Accordingly, investigators undertaking such a study should expect to devote substantial resources to recruitment. The challenge is much greater than in provider-based convenience studies, because there is no guarantee that providers who are sampled as part of a probability sample will provide information on their practices or allow investigators to obtain access to their patients.

The assessment of response or coverage rate for the population must take into account providers from whom no data can be obtained on eligible numbers of patients, providers who refuse to enroll, providers who enroll but who fail to enroll their patients, and patients who enroll but are not interviewed. For example (assuming a study in which all providers and patients have equal weight), if eligibility is successfully ascertained for 95 percent of eligible providers, 70 percent of selected eligible providers agree to participate in a study, 75 percent of selected patients agree to enroll, and 80 percent of enrollees actually complete the interview, the overall participation rate is (0.95)(0.7)(0.75)(0.8), or about 40 percent of all persons who *should* have been interviewed.

The HCSUS made use of extensive physician-to-physician contacts and a willingness to adjust operational plans to local needs to minimize provider nonparticipation. We undertook a three-pronged strategy to minimize subject nonresponse: (1) very intensive follow-up of all sampled individuals, including the provision of resources to sites to do outreach; (2) use of a small number of specially trained interviewers, who discussed consent with individuals who were undecided about participation; and (3) availability of members of the HCSUS Community Advisory Board (see further on) to address the concerns of reluctant participants. To minimize the loss of data, we offered

a shorter interview to subjects who could not complete the full interview because of illness. We also sought similar data from providers or proxies for subjects who could not be interviewed at all because they were dead, too ill, or unlocatable. Finally, we sought aggregated and de-identified data from providers for patients who declined participation.

IMPLICATIONS OF NATIONAL PROBABILITY STUDIES FOR RESEARCH ON SPECIFIC DISEASES

A probability design for the study of specific diseases offers enormous opportunities for answering questions of considerable policy relevance–but at great expense. The assembly of a randomly selected study population, design and programming of the computer-assisted survey instruments, and basic data cleaning and weight construction are all very expensive in a study such as the HCSUS. The size of the investment in a single project creates opportunities and obligations.

Use of a National Probability Study as an Opportunity for Multiple, Parallel Efforts

The decision to undertake such a study of a disease should be made only if compelling national policy questions need to be addressed using unbiased national data. Once a commitment is made to undertake such a study, it makes sense to try to find ways to study additional questions of policy relevance that are beyond the scope of the original study or less central to the interests of the primary funder. For example, once the fixed cost of assembling the HCSUS sample was met, it became attractive to sponsors to fund relatively efficient supplemental studies of rural HIV and early disease (HRSA), mental illness (NIMH), drug use (NIDA), oral health (NIH Office for Research in Minority Health through NIDR), older HIV patients (NIA), providers of care (the Robert Wood Johnson Foundation), and HIV viral load and resistance (NIAID, Bristol-Myers/Squibb, and several other pharmaceutical and biotechnology firms) as well as of studies of the use of antiretroviral therapy (Glaxo-Wellcome and Merck and Company). Thus, HCSUS became something of a "program project," organized around the essential feature of a national probability sample. This model may be relevant to similar studies of other diseases.

The Need for Buy-in from Persons Affected by the Disease

When a major national study of a disease is undertaken, it is absolutely essential that a very strong voice be given to the patients. The HIV activist community's tradition of involvement in policy and research made it a natural step for the request for proposal to mandate their involvement for these reasons. First, such studies come along infrequently. It is important that the full range of interests be present as project policies and priorities are formulated. Second, participants in these studies often carry valuable insights into important and emerging issues, as well as into ways in which they ought to be addressed. Third, patients can promote participation in the study, both by having a presence in the study and by direct intervention.

In HCSUS, we involved the patient community very seriously. We hired a community coordinator who selected a national Community Advisory Board. Community representatives were appointed to every policy body, task group, and research team on the project. There is no doubt that their presence shaped the discussion in all of these settings and that they enhanced the quality of the project. Such high-level community participation might well prove equally valuable in studies of other diseases.

The Need for Buy-in from Practitioners and Investigators Who Control Access to Patients

Provider cooperation is critical to the success of a probability sample study like HCSUS, but a compelling reason needs to exist for providers, some of whom are involved in other clinical investigations, to make their patients available to such a study. The value of the data to be produced probably is an insufficient incentive on its own. The experience of HCSUS has been that it is necessary to involve both physicians in the project leadership and physicians in the community to recruit providers, using a range of incentives that include monetary compensation for their efforts as well as participation as investigators in the project.

Recruitment of local investigators or "captains" to represent the study in selected communities was aided by the argument that a nationally representative sample study can be useful to all investigators in the relevant disease area. It can provide a context for related studies using convenience or purposive samples or subsamples. Potential collaborators can learn how compatible the characteristics of smaller samples in their studies are to the probability sample population and, therefore, to what extent relationships that are observed in smaller studies might generalize to the population as a

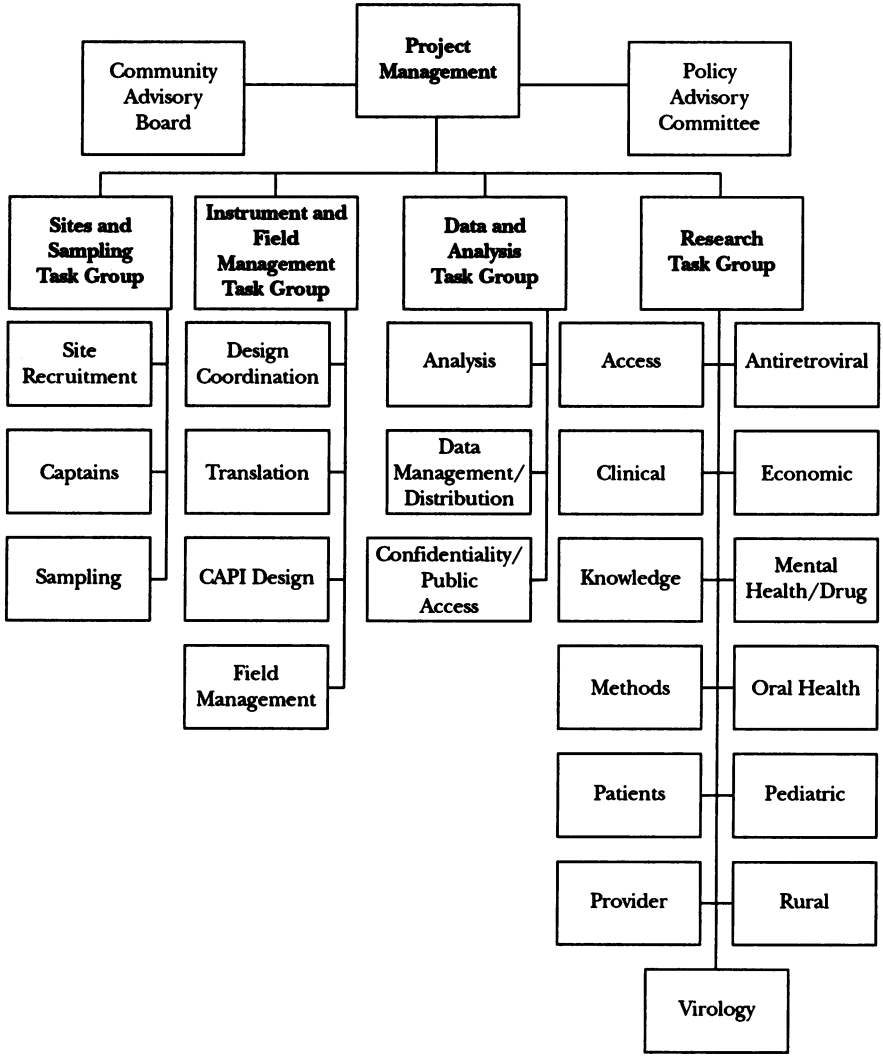
whole. At the same time, a probability study can identify a range of policy and research issues that need to be addressed in more focused studies. Thus, nationally representative cohorts should be viewed as public resources that not only provide useful information, in their own right, but also enhance the ability of others to conduct meaningful scholarship. If a project clearly commits itself to this principle, it should be possible to foster an environment of cooperation with such studies.

The Virtual Research Organization. The assembly of a nationally representative sample is a large and expensive proposition. Broad participation of members of the national research community with interests in relevant areas helps ensure that the data collected will be relevant and important. Accordingly, AHCPR mandated the involvement of a broad team in HCSUS. The study team we assembled had very active participants across more than a dozen major institutions and occasional participants at a score of others. To be able to work effectively in this context, we established what amounts to a virtual research organization that relies heavily on electronic mail and conference calls. This organization functions through five operationally oriented task groups: Sampling, Instrument Design and Data Collection, Data Management and Analysis, Rapid Response, and Research. The Research task group consists of 13 topical research teams (see Figure 1). After an initial in-person meeting, nearly all subsequent contact between and among the research teams has been through electronic mail and weekly or biweekly conference calls. In general, the teams and task groups have worked quite effectively and cohesively.

Instrument Design. As alluded to earlier, the broad involvement of researchers, community members, and sponsors increases the difficulty of managing the instrument design process: the inclusion of a large number of investigators and several supplemental projects could lead to endless instrument revisions and a huge subject burden. To combat this, the HCSUS instrument design task group required research teams to develop research questions and candidate items. Prioritization and revision of the research questions then led directly to the prioritization and revision of candidate data items. The candidate items were then fashioned into draft instruments, which were piloted. The results of the pilot were fed into models to estimate the impact of further changes on timing.

One additional issue that our Instrument Design team and investigators faced was whether or not to put effort and resources into the derivation of new instruments to assess health status, social experiences with HIV, patterns of utilization, patient perceptions of quality, adherence, and the like. We rapidly

Figure 1: Functional Organization of the HCSUS



concluded that a national probability sample is not the place for extensive instrumental innovation: the time with randomly selected patients is simply too valuable. Because we considered the value of the information to be produced as relating largely to the study's representativeness and comprehensiveness, we used existing or lightly adapted measures wherever possible.

Data Management and Analysis. The importance of producing a quality

data set quickly for use by a large number of researchers with potentially overlapping research objectives necessitated a highly organized effort from the study's Data and Analysis task group. The HCSUS has devoted a great deal of effort to finalizing weights and performing necessary imputations; other researchers undertaking national probability designs should not underestimate the effort required. While those efforts are ongoing, we have asked researchers to write analysis plans for each research question they anticipate pursuing, including at least five components: unit of analysis, applicable sample, dependent variable(s), independent variable(s), and analysis methods.

Members of the task group assisted in writing these individual plans, which were a required component of proposals for publication. The Data and Analysis task group subsequently produced an overall analysis plan that allowed the project leadership to map the order and relationship between individual plans according to global project objectives. A core set of variables (e.g., demographics such as gender and age or utilization measures such as numbers of emergency room visits or outpatient visits) were common to many analyses. They were derived by the internal project staff for use by all research teams. This task group addressed methodological challenges that all research teams face and provided training in these techniques to the HCSUS Research teams. Such centralization helps achieve timeliness and uniformity of quality of data and methods.

Dissemination of Data and of Findings. One of the main purposes of fielding large projects with nationally representative samples is to shape health policy. Clearly, fulfilling this mandate means that even a very large-scale project such as the HCSUS must be able to get out important information quickly and broadly. Accordingly, we have instituted a program of rapid or advance data releases of some key findings of particular interest to policymakers in the manner common to "in-house" government studies. In addition, HCSUS coinvestigators at the AHCPR will maintain the capability for a rapid response to requests from policymakers, in collaboration with the project and research team leadership. Finally, to ensure that persons outside of the project will have opportunities to conduct scholarship on HCSUS data, the HCSUS data are broadly available at a repository at AHCPR on a short timeline relative to the standards in the field.

Ensuring that papers and reports are published in an orderly yet timely fashion and that topics are allocated fairly is a high priority for the project management. Accordingly, a Publications Committee reviews all analysis and reporting proposals and has promulgated timelines for individuals to produce approved papers.

Doing It Better Next Time. We learned a great deal from our efforts in implementing HCSUS that should inform future efforts to conduct a similar study. First, future investigators should take care not to underestimate the expense and effort of assembling a national probability sample. We ended up devoting far more resources to this effort than anyone on our team anticipated. This resulted in some required trade-offs. Eventually, we sacrificed a fourth round of planned interviews and a second wave of planned record abstraction to cover the additional cost of obtaining the sample and baseline interviews. We also offset some of these costs through efficiencies achieved by adding supplemental projects at marginal cost.

Among our costliest decisions was to use list-based sampling at most sites, in which providers periodically gave us complete lists of eligible patients seen. The alternative was "real-time sampling," in which subjects were sampled as they were identified. Emphasizing list-based sampling saved on the up-front costs because we did not need to station study personnel at the site. However, it added enormously to the expense of eliciting subject participation because we were dependent on providers to contact patients whom they saw episodically. We spent a great deal of time and effort promoting these contacts. In retrospect, emphasizing "real-time sampling" would have saved much time and expense in the long run.

Second, studies of sparse populations consume enormous time and resources. In our study, the effort to enroll a rural oversample distracted us at times from the business of assembling the core national probability sample. Third, it is important to stay closely attuned to evolving policy concerns. Fortunately, we were aware of the therapeutic revolution that occurred during our study and adapted our interview instruments accordingly. We were less attentive to emerging issues in insurance coverage, such as the spread of Medicaid managed care and limitations in private coverage. We have had to compensate with alternative data collections that were unfunded or required supplemental funding.

Fourth, the weighting process is quite complex and has the potential to greatly slow the production of policy-relevant data. For example, weights can take account of interview data, proxy data, nonresponse data, record data, and national death index data, all of which may become available at different times. Trade-offs are necessary between the complexity of the weighting activities and the need to produce timely information.

Finally, we found it very challenging and expensive to meet the specifications of the Office for the Protection from Research Risks (OPRR). Some of these specifications appeared to be responses to bureaucratic imperatives

in going beyond the clear need to protect confidentiality and obtain informed consent, and some did not appear to be designed primarily to protect subjects from harm (Berry et al. 1998). This, too, proved to be a distraction from other project tasks.

CONCLUSION: IMPLICATIONS FOR OUTCOMES RESEARCH

The promise and costs of national probability studies present an interesting dilemma to the field of health services and outcomes research. We believe that a need exists for at least some outcomes studies of specific diseases to use such designs if our field is to achieve its maximum potential policy relevance. Even though such studies are expensive, the cost of doing most disease-specific analyses using general health studies such as the National Health Interview Survey or the Medical Expenditure Panel Survey would be far greater. Such studies would have to be greatly expanded to include sufficient samples of persons with uncommon diseases. When the federal government and the research and policy communities determine that it is important to collect unbiased national data on specific diseases, the paradigm used in HCSUS makes such studies more practical.

Of course, the field of health services research must balance the potential benefits of very large projects and their attendant "big science" teams against the need to encourage creativity at all levels of scholarly enterprise. If resources are not preserved for smaller investigator-initiated efforts, large projects could come to dominate scholarship in the manner of modern particle physics. Little room would remain for the separate nimble efforts from which come many of the most creative, helpful, and enriching ideas. However, noting that the HCSUS used less than 3 percent of AHCPR's annual budget at its peak, we believe that there should be room for both kinds of efforts on the national health services research agenda. This will clearly be the case if the practical value of studies like the HCSUS can be used as an argument to infuse substantially more resources into health services research or to convince the NIH to support research efforts on diseases of interest.

In summary, the national probability sampling model offers an opportunity to represent accurately the population under care for specific diseases, thereby enhancing the value of existing research and providing a vehicle for the production of unbiased data on a full range of issues that matter to clinical and social policy. The model used by HCSUS does not account for people

not receiving care, is quite expensive, and is logistically very challenging. Still, more frequent and successful adoption of national probability designs should help make health services research an irresistible force in the tumultuous discourse about the healthcare system.

NOTE

1. The original award from AHCPR for the project provided for total costs of $15 million.

REFERENCES

Alonzo, A. A. and N. R. Reynolds. 1995. "Stigma, HIV and AIDS: An Exploration and Elaboration of a Stigma Trajectory." *Social Science and Medicine* 41 (3): 303–15.

Andersen, R. M. 1995. "Revisiting the Behavioral Model and Access to Medical Care: Does It Matter?" *Journal of Health and Social Behavior* 36 (1): 1–10.

Berk, M. L., C. Maffeo, and C. L. Schur. 1993. *Research Design and Analysis Objectives: AIDS Cost and Services Utilization Survey (ACSUS) Reports, No. 1.* AHCPR Pub. No. 93-0019. Rockville, MD: Agency for Health Care Policy and Research.

Berry, S. H., J. F. Perlman, S. Nederend, and T. K. Bikson. 1998. "Variations in IRB Response to the HIV Cost and Services Utilization Study (HCSUS)." Presented at the 15th Annual Meeting of the Association for Health Services Research, Washington, DC, 21–23 June.

Bozzette, S. A., S. H. Berry, N. Duan, M. R. Frankel, A. A. Leibowitz, D. P. Goldman, R. D. Hays, J. Keesey, D. Lefkowitz, J. A. McCutchan, J. Perlman, and M. F. Shapiro. 1998. "The Care of HIV-infected Adults in the United States: Results from the HIV Cost and Services Utilization Study." *The New England Journal of Medicine* 339 (26): 1897–1904.

Frankel, M. R., M. F. Shapiro, N. Duan, S. C. Morton, S. H. Berry, J. A. Brown, M. A. Burnam, S. E. Cohn, D. P. Goldman, D. F. McCaffrey, S. M. Smith, P. A. St. Clair, J. F. Tebow, and S. A. Bozzette. 1999. "National Probability Samples in Studies of Low-Prevalence Diseases. Part II: Designing and Implementing the HIV Cost and Services Utilization Study Sample." *Health Services Research* 34 (5): 969–92.

Grimaldi, P. L. 1997. "New HEDIS Means More Information about HMOs." *Journal of Health Care Finance* 23 (4): 40–50.

Harris, D. M., P. Hanes, H. Jimison, D. Jones, J. Bryan-Wilson, and M. R. Greenlick. 1997. "Physician and Plan Effects on Satisfaction of Medicaid Managed Care Patients with Their Health Care and Providers." *Journal of Ambulatory Care and Management* 20 (1): 46–64.

Hellinger, F. J., J. A. Fleishman, and D. C. Hsia. 1994. "AIDS Treatment Costs During the Last Months of Life: Evidence from the ACSUS." *Health Services Research* 29 (5): 569–81.

Iezzoni, L. I., M. Shwartz, A. S. Ash, and Y. D. Mackiernan. 1996. "Using Severity Measures to Predict the Likelihood of Death for Pneumonia Inpatients." *Journal of General Internal Medicine* 11 (1): 23–31.

Kish, L. 1965. *Survey Sampling.* New York: J. Wiley.

Lehman, A. F. 1995. "Measuring Quality of Life in a Reformed Health System." *Health Affairs* (Millwood) 14 (3): 90–101.

Marx, R., M. H. Katz, M. S. Park, and R. J. Gurley. 1997. "Meeting the Service Needs of HIV-infected Persons: Is the Ryan White CARE Act Succeeding?" *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 14 (1): 44–55.

National Committee for Quality Assurance. 1997. "NCQA Implements New Outcomes Audit Standards." *Healthcare Benchmarks* (4): 84–85.

O'Malley, C. 1997. "Quality Measurement for Health Systems: Accreditation and Report Cards." *American Journal of Health-System Pharmacy* 54 (13): 1528–35.

Thomas, M. R., J. Stoyva, S. A. Rosenberg, C. Kassner, G. E. Fryer, A. A. Giese, and S. L. Dubovsky. 1997. "Selection Bias in an Inpatient Outcomes Monitoring Project." *General Hospital Psychiatry* 19 (1): 56–61.

Ware, J. E., Jr., M. S. Bayliss, W. H. Rogers, M. Kosinski, and A. R. Tarlov. 1996. "Differences in Four-Year Health Outcomes for Elderly and Poor, Chronically Ill Patients Treated in HMO and Fee-for-Service Systems: Results from the Medical Outcomes Study." *Journal of the American Medical Association* 276 (13): 1039–47.

Welch, W. P., and H. G. Welch. 1995. "Fee-for-Data: A Strategy to Open the HMO Black Box." *Health Affairs* (Millwood) 14 (4): 104–16.