Sampling Patients within Physician Practices and Health Plans: Multistage Cluster Samples in Health Services Research

John L. Adams, Steven L. Wickstrom, Margaret J. Burgess, Paul P. Lee, and José J. Escarce

Objective. To better inform study design decisions when sampling patients within health plans and physician practices with multiple analysis goals.

Study Setting. Chronic eye care patients within six health plans across the United States.

Study Design. We developed a simulation-based approach for designing multistage samples. We created a range of candidate designs, evaluated them with respect to multiple sampling goals, investigated their tradeoffs, and identified the design that is the best compromise among all goals. This approach recognizes that most data collection efforts have multiple competing goals.

Data Collection. We constructed a sample frame from all diabetic patients in six health plans with evidence of chronic eye disease (glaucoma and retinopathy).

Principal Findings. Simulations of different study designs can uncover efficiency gains as well as inform potential tradeoffs among study goals. Simulations enable us to quantify these efficiency gains and to draw tradeoff curves.

Conclusions. When designing a complex multistage sample it is desirable to explore the tradeoffs between competing sampling goals via simulation. Simulations enable us to investigate a larger number of candidate designs and are therefore likely to identify more efficient designs.

Key words. Clustering, design effect, multistage, sampling studies, simulation

Most surveys or data collection efforts in health services research have multiple goals. Frequently several researchers or institutions will jointly conduct a study, each with a different primary analytical interest. Even within a single researcher's data collection effort there may be several analysis priorities of differing importance with each suggesting a different study design. Multiple goal surveys are a sensible compromise in an era of limited research budgets.

Studies also may face design constraints. Beyond the obvious constraint of a limited budget, there are often practical constraints on portions of the

study design. Limitations on the availability of the surveyed institutions or participants' time are typical examples. The typical sampling design in these multiple goal situations is to consider the optimal design (Cochran 1977), if known, for each of the goals and develop a hybrid of the individual optimal designs. This can be difficult if the optimal designs for the competing goals are very different. Real but nonstandard constraints can also complicate the design problem.

In this article we develop a simulation-based method to design samples that increase the efficiency of the study design. There are two main ideas underpinning the simulation sampling method:

1. When confronted by a study with competing goals and/or complicated structure many alternative designs should be considered. In some complicated study design problems there is no obvious optimal solution among the well-known sampling designs. Even if there is a known solution for one analysis goal, it may be quite different from the best approach for another goal. Many designs should be considered and evaluated, including designs that are nonstandard but that might capitalize on unique features of the specific problem.
2. There is no longer a compelling reason to restrict our attention to the easily evaluated designs from classical sampling texts. Almost any design that can be conceived can be evaluated via simulation. Sampling probabilities can be approximated even in cases where the multiplicity of ways that a unit could be sampled would stymie analytical calculation. Features of the sample that are difficult to calculate in closed form can be simulated even for complicated statistics.

This article introduces a sampling methodology for multistage sampling of health services users. The approach described here can be applied to any complex sampling problem where the levels of the structure are nested or can be redefined as nested.

Address correspondence to John L. Adams, Ph.D., RAND Health, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-3297. Dr. Adams is a Statistician. Steven L. Wickstrom, M.S., and Margaret J. Burgess are with the Center for Health Care Policy and Evaluation, UnitedHealth Group, Eden Prairie, MN. Paul P. Lee, M.D., J.D., is with the Department of Ophthalmology, Duke Eye Center, Durham, NC. José J. Escarce M.D., is with RAND, Santa Monica.

BACKGROUND

A common study design in health services research is the multistage cluster sample (Kish 1965) to abstract data from medical records. The typical example would be the sampling of practices from a geographic area or a health plan, followed by the sampling of physicians associated with those practices, followed by the sampling of patients treated by the sampled physicians. The sampling plan for the study must address which practices, physicians, and patients to select. The ultimate goal is a sample of patients, but the human subjects approvals process and the costs of access to medical records compel following the hierarchy of the service delivery system to the patient. Fixed costs associated with recruiting and visiting practices and physicians can make the cost of simple random sampling of patients prohibitive.

Consider the simple and common two-stage problem of sampling patients within physician practices to estimate a population average (e.g., mean blood pressure). The reference design would be to sample practices proportional to size and then take equal numbers of patients per sampled practice up to the budget constraint. This design minimizes the design effect (the design effect specifies the loss or the gain in precision of a design relative to simple random sampling) by simultaneously minimizing cluster effects and the variability in the sampling weights. But what is the right design if there are two or more subpopulations (e.g., men and women) for whom separate estimates are required? With this simple expansion of the design goals there is already no standard sampling strategy.

The primary goals for study design should be to maximize the analytical information while controlling data collection costs. The example above is instructive. The naïve approach might be to sample patients from both subpopulations at random and then go to all of their providers to review medical records for patient characteristics or care received. This approach can result in “snowballing” data collection costs as abstracters go from office to office to find the required charts. To keep costs down, the design should capitalize on the lower marginal costs of data collection at providers already being visited. Care must be taken, however, to ensure that the sample does not end up concentrated only in the high-volume providers or overclustered in a way that hurts the effective sample size.

As our example suggests, complex designs often have different competing goals. Statisticians consider small design effects for several quantities of interest and low or fixed cost of the study as worthy design goals. Sampling proportional to size at the first stage might best support one

goal of the study, whereas another goal might be best supported by equal probability sampling. For example, if the characterization of the first stage units is a goal of the study then equal probability sampling is a design that should be considered.

Our proposed method falls in the general category of using optimization techniques to improve the sample. Several authors have explored variations on this approach in the literature. Typically this approach results in a nonlinear, but perhaps convex, optimization problem. Kish (1974) suggests a fixed-cost approach to sample optimization. The idea is to minimize the variance of quantities of interest subject to a cost constraint. Chromy (1987) discusses sample optimization for multiple goals by minimizing costs subject to variance constraints that correspond to the goals. This approach has the attractive feature of spending no more than the minimum needed to achieve the stated goals. Folsom, Potter, and Williams (1987) discuss the construction of size measures to achieve a self-weighting sample when sampling proportional to size. We adopt a similar idea to use various size measures as a way to define a universe of potential sampling strategies. Biemer (1983) uses simulations to evaluate various dual-frame sample allocations. We also adopt the use of simulations to compare alternative designs.

In the next section, we outline a more general approach. By calculating the design effects or effective sample sizes for multiple goals under different design parameters we enable the user to examine the tradeoffs among design alternatives to make an informed decision. We use simulation to avoid some of the challenging optimization techniques other authors have used. We suggest a simple graphical device for exploring design tradeoffs to avoid the need to prespecify the exact values of the goals. Our method also avoids the need to sequentially alter constraints until a solution is found like earlier methods. Rather we try to facilitate the designer's exploration of the available options.

As an example we will focus on the design of the Chronic Eye Care Study (CECS). The three-year project is an observational study of the impact of varying organizational and financing arrangements under managed care on the quality of eye care for working-age patients with diabetic retinopathy (background and proliferative) or primary open-angle glaucoma, the two leading causes of blindness in the United States. The project represents a collaborative effort between investigators at RAND and the Center for Health Care Policy and Evaluation (CHCPE), the research institute commissioned by UnitedHealth Group, a leading managed care organization that includes UnitedHealthcare health plans. The project is currently analyzing data collected from patients in six geographically dispersed health plans to

examine the impact of varying managed care features on process of care, intermediate outcomes of care (e.g., control of intraocular pressure in glaucoma), patient satisfaction, and utilization of eye care services.

Managed care features assessed by the project include type of eye care provider, provider payment method and practice arrangement, and utilization and quality management systems.

Six UnitedHealthcare managed care plans (three from the Midwest, one from the Northeast, one from the Southeast, and one from the West) were selected based on diversity of organizational and financing arrangements for providing eye care, geographic diversity, and availability of high-quality administrative data. The plans used administrative data to identify patients with the study conditions and assign them to providers and practices.

Data sources include patient survey, medical record abstraction, a survey of providers, and administrative data from the study plans. Both the medical record and the provider survey costs suggested a multistage cluster design.

METHOD

The method is based on simulating sample draws from a large collection of potential sampling designs to search for a design that maximizes the information in the sample while meeting the constraints. We consider all the typical design alternatives including sampling with equal probability and proportional to size as special cases. Further, we explore alternative definitions of the importance of a particular sampling unit. We consider different design goals and use simulations to evaluate them and make the tradeoff between the competing goals explicit. Our approach to picking a sampling design is the blueprint for our plans to develop a more general sampling design development tool. The five steps are:

1. List a set of sensible sampling designs (usually parameterized)
2. List the constraints (economic or organizational)
3. Simulate sample draws from the design set that meet the constraints
4. Evaluate the draws with respect to multiple goals
5. Pick the design that provides the best compromise among the goals.

We now explain these steps in detail, using the Chronic Eye Care study to illustrate the concepts. We employ a three-stage design that samples practices within health plans, followed by the sampling of physicians associated with

those practices, followed by the sampling of patients treated by the sampled physicians.

1. List a Set of Sensible Sampling Designs (Usually Parameterized)

Having a choice among several sampling designs is at least as good, and often better than, having just one sample design. We consider several designs and then choose the design that yields the best compromise among all goals of a particular study. To elicit a candidate list of sensible designs, it is helpful to identify design parameters. In the CECS we sampled practices proportional to size. The improvement we made over the standard approach was to consider an expanded family of size measures:

$$\text{Size} = \alpha * (W_B * N_B + W_G * N_G + W_P * N_P)^\lambda / (\text{cost}^{1/2})$$

Where α is a tuning constant to control the number of practices drawn; N_B , N_P , and N_G are the numbers of background, proliferative, and glaucoma cases in the clinic; and W_B , W_P , and W_G are weights we use to assign greater value to rarer conditions. And λ is an exponent between 0 and 1. Cost is the fixed cost of including the clinic in the sample.

If two of the three weights in this size measure are set to 0, λ is set to 1, and costs are equal across all practices, this size measure is the standard sampling proportional to size measure for one of the conditions. The square root of costs is used since it is the way costs enter the optimal sampling probability calculation in the standard case (Cochran 1977). The additional parameters allow us to explore a range of compromises among the samples for the three conditions. A similar size measure with its own weighting parameters was used for the sampling of physicians within practices. This size measure is different from the size measure used by Folsom, Potter, and Williams (1987). There the goal was to use the size measure to achieve a self-weighting sample. With recent widespread availability of statistical software that uses weights properly, self-weighting is a relatively less important goal.

The final stage of sampling is the selection of patients within practices. At this stage we are free to select the sampling probabilities by condition separately. Patient sampling probabilities by condition may be set independently; they are adjustable parameters themselves like the W s in the previous stages. The constraint of no more than 10 sampled cases per condition per

physician does affect the sampling probabilities:

$$P_{\text{patient}|\text{physician}} = P_{\text{condition}}$$

Unless $P_{\text{condition}} * (\# \text{ cases available}) > 10$, then

$$P_{\text{patient}|\text{physician}} = 10 / (\# \text{ cases available})$$

where $P_{\text{condition}}$ is set separately for the three conditions.

This stage samples each condition with its own sampling rate up to a cap of 10 cases per condition. Capping the sample draw in this way reduces the cluster design effect as well as satisfies the physician-level burden constraint (see below). Note that it would be straightforward to do stratification by patient characteristics within-physician. For example, stratification by race or gender would frequently be useful. The within-physician sampling probabilities would then differ by race and gender and these probabilities could then be adjusted to change the resulting race and gender mix of the sample.

In the CECS we did not anticipate any differential nonresponse. If differential nonresponse was anticipated, for example, by age or geographic area it could be addressed in the patient-level sampling probabilities. The relationship would then become: $P_{\text{patient}|\text{physician}} = P_{\text{response}} * P_{\text{condition}}$ where the response term would be a function of patient characteristics associated with nonresponse. This would incorporate design effects due to differing response rates into the sample choice. Alternatively, you could incorporate differential costs per case for the additional effort required to equalize the expected response rates.

2. List the Constraints (Economic or Organizational)

The principal constraint in many studies is cost. Unfortunately, the cost function in all its detail is seldom available. Ideally the study designer would have fixed and variable costs for each design decision. At best the designer may be able to get the costs for a handful of alternative designs from the data collection subcontractor. Typically, costs end up expressed as a list of constraints, for example: "We can afford 280 cases per site as long as we don't need to visit more than 30 practices."

The CECS had six health plans. The constraints on the sampling in each of the six plans were:

- A maximum of 280 cases total in each market (to control data collection costs).

- As many proliferative cases as possible. These are the most infrequent and severe cases and therefore the most interesting cases for some analyses.
- A maximum of 60 background retinopathy cases in each market (this is the more abundant and less interesting condition).
- No more than 10 patients from each condition from any physician (an administrative constraint imposed to limit the burden on participating physicians and minimize cluster design effects).
- No more than 30 practices per market (to limit the burden on the data collection staff).

These constraints may (and will) be modified in response to the reality of the distribution of the available patients in the six markets.

3. Simulate Sample Draws from the Design Set That Meet the Constraints

As part of our collaboration with the CHCPE, they developed software in *SAS* that implements the simulation sampling approach for the three-stage cluster-sampling problem. The software has three main parts; the sampler, the evaluator, and the loopier. The sampler takes a particular set of sampling parameters (α , β , λ , W_P , W_B , W_G , P_P , P_B , P_G) and performs a sample draw from the population. The evaluator then takes this sample draw and extracts properties of the sample statistics. This includes numbers of practices drawn, number of patients drawn in each of the three conditions, effective sample sizes for each of the three conditions as well as design effects and total cost estimates. The loopier then reruns the sample draw for other values of the sampling parameters until all possible designs have been evaluated. This would usually be done for more than one replicate of each set of design parameters to assess the variability of the sample properties.

4. Evaluate the Draws with Respect to Multiple Goals

The downside to more elaborate sampling designs is the design effect. Intuitively, the design effect is a consequence of trying to represent a population while concentrating the sample in subsets of the population. The design effect is expressed as a number, usually greater than 1. This number is the number of cases in your sample required to achieve the same precision as one case in the simple random sample. The sample size can be divided by the design effect to produce an effective sample size. The effective sample size is

the size of a simple random sample that would achieve the same precision as this more complex sampling design.

The true design effect can be a difficult calculation in some cases. In this project we used a simple approximation. We approximate the overall design effect with two design effect factors. They are the probability effect and the cluster effect (Gabler, Haeder, and Lahiri 1999). These design effects can be readily calculated for any given sample. It would be possible in some situations to use analysis software that does properly adjusted standard errors (e.g., *SUDAAN*, or *SAS* version 8) to calculate more precise design effects.

The probability design effect is the effect due to differential sampling probabilities. The probability design effect is most frequently seen calculated in terms of the analysis weight, $w = 1/Prob_{patient}$ where $Prob_{patient} = P_{clinic} * P_{physician|clinic} * P_{patient|physician}$. The probability design effect is then: $Deff_p = N * \Sigma w^2 / (\Sigma w)^2$ where N is the total sample size. This design effect is a consequence of the greater variability of the weighted analysis than the variability of the unweighted analysis from a simple random sample. There is a different probability design effect for each of the three conditions of interest.

The cluster design effect accounts for the stronger similarity of cases within a cluster. Presumably patients within a single physician are more alike than patients selected at random from a different physician. This similarity of patients within a single physician is frequently characterized by the correlation within physician, ρ . If this correlation is greater than 0 it reduces the effective sample size. This is simply a consequence of the lower value of information of a second patient from a particular physician's practice than the value of a first patient from a physician's practice that has not yet been sampled. A simple approximation to this design effect is: $Deff_c = 1 + \rho * (m - 1)$, where ρ^1 is the intraclass correlation and m is the mean cluster size (within physician). Typically one would not know the value of ρ a priori. The standard approach is to plug in a value of ρ thought to be appropriate from experience in similar settings. The mean cluster size is a consequence of a given sampling design.

These two design effects, probability and clustering, both participate in the effective sample size calculation. Our effective sample size for each condition is then: $ESS = N / (Deff_p * Deff_c)$ where N is the actual number of patients sampled. We have three effective sample sizes, one for each of the three conditions. Effective sample size for a given analysis is the ultimate measure of the information that a sample contains for estimating population statistics. Note that other properties of the sample statistics (e.g., regression coefficient standard errors) could be of interest in other problems and could be calculated with the approach presented here.

5. Choose the Design That Is the Best Compromise among the Goals

To select the design that best compromises among the competing goals we use a manufacturing analogy that motivated the approach. The key idea that we borrow from elementary economics is the idea of a “production possibilities frontier (PPF).” The PPF concept is that a factory (or economy) can produce differing quantities of various goods by changing the production parameters but there is a tradeoff if cost or inputs are constrained. Here we have experimented with a range of input parameters and are examining how many cases of each of the three conditions of interest are in the resulting sample. Our sampling system can be thought of as a factory that produces three fundamental but competing products; the effective sample sizes for background, proliferative, and glaucoma cases.

In the classic introductory economics text example only two goods are produced, typically “guns” and “butter.” This results in a simple PPF where the possible quantity pairs can be seen. Visualizing the PPF for the three conditions requires a three-dimensional PPF graph. Figure 1 illustrates the PPF for the trade-off among background, proliferative, and glaucoma cases in

Figure 1: The Production Possibility Frontier for the Tradeoff Between Proliferative, Glaucoma, and Background Cases

terms of effective sample size within one of the six health plans (SE). Many of the sample designs fall in the interior of this PPF, dominated by designs on the frontier, and are not shown in this figure. As you move to the upper right on this graph you slowly reduce the effective sample size for proliferative cases while gaining effective sample size for background and glaucoma. As you move along one of these equal proliferative curves you trade off between background and glaucoma cases. The entire surface plotted here is one collection of equal cost design options. Increasing one kind of case decreases another kind. The study designer must decide how to make the tradeoffs. One thing is clear from the flat area in the lower left of the graph: no matter what design parameters you pick you cannot produce a proliferative effective sample size much above 58. The “X” on this graph corresponds to the set of design parameters that were ultimately selected for the SE sampling. This graph may be contrasted with the conceptually similar approach in Chromy (1987). Here, rather than set minima for the three dimensions and then minimizing costs, we facilitate visualizing potential solutions. This could result in rethinking the goals as the designer gains intuition into what is possible or difficult.

This contour graph was produced by taking the three effective sample sizes from each run of the simulation model, aggregating the results by the background and glaucoma effective sample sizes, and selecting the maximum proliferative effective sample size from the aggregated cells.²

Although contour plots of this type are a useful way to visualize the tradeoffs in the study design, they can be difficult to produce. Cell sizes and smoothing or modeling choices must be selected with care. Frequently the study designer will simply not have the time and resources to produce these graphs. It is also difficult, if not impossible, to produce PPF graphs in more than three dimensions. In most cases tradeoffs can be examined by selecting subsets from the design database. For example, taking the simulation results and applying restrictions like: “Are there any cases where I have an effective sample size of at least 60 glaucoma and 60 proliferative cases while drawing no more than 280 cases total?” If a query of this form returns more than one possible design, further restrictions may be applied. (“How about 70 proliferative?”) If no cases are returned from such a query the least important constraint can be relaxed. (“How about only 50 glaucoma cases?”) In more than three dimensions it may also be useful to look at several PPF graphs examining three variables at a time. It may become clear that some dimensions are “low leverage” on the design decision. One can quickly weed out the easy to achieve and impossible to achieve dimensions of the design.

Usually all design goals cannot be optimized simultaneously; therefore, the design of choice is a compromise. However, we prefer a conscious compromise in which we are aware of the tradeoffs. What tradeoffs are acceptable depends on the context of the problem at hand. Once a design has been chosen, a single final sample is drawn.

RESULTS

The goals of the CECS are to have an adequate sample in each of the six plans to characterize the three conditions as well as adequate samples in each of the three conditions when combined across the sites. Initially it was hoped that a balanced design across all three conditions at each of the six sites would be possible. The reality of the distribution of the conditions across plans and providers makes such balance impossible to achieve. The relatively rare but most interesting proliferative cases are difficult to sample in quantity due to their dispersion across providers. Ultimately we needed to try and get more proliferative cases at the sites where more were available. This modification of the design goals in the middle of the design process shows the power of the simulation method to inform coping with unfortunate realities that emerge in the study design process between the abstract design in the proposal and the ultimate design that will be fielded.

Table 1 presents the final design parameters used to draw the sample. We ultimately did not find a compelling reason to use different λ s, clinic, or physician weights across the plans. The practice and clinic weights of .7 for

Table 1: Final Parameters Used to Draw the Samples

<i>Plan</i>	<i>SE</i>	<i>MW1</i>	<i>MW2</i>	<i>NE</i>	<i>West</i>	<i>MW3</i>
Alpha	35	11	20	7	40	7
Beta	30	25	40	20	15	15
Lambda	0.25	0.25	0.25	0.25	0.25	0.25
Practice weight for Background	0.15	0.15	0.15	0.15	0.15	0.15
Practice weight for Glaucoma	0.15	0.15	0.15	0.15	0.15	0.15
Practice weight for Proliferative	0.7	0.7	0.7	0.7	0.7	0.7
Physician weight for Background	0.15	0.15	0.15	0.15	0.15	0.15
Physician weight for Glaucoma	0.15	0.15	0.15	0.15	0.15	0.15
Physician weight for Proliferative	0.7	0.7	0.7	0.7	0.7	0.7
Member probability for Background	0.96	0.4	0.4	0.35	1	0.5
Member probability for Glaucoma	0.7	0.22	0.22	0.13	1	0.25

proliferative and .15 for the two more common conditions show the effort to concentrate the sample in practices that are relatively rich in proliferative cases. We did substantially vary the patient probabilities from site to site to control the fraction of patients in each condition. In all sites the patient probabilities for proliferative cases were one.

Table 2 shows the results and reallocation across the plans. In MW1, MW2, and NE we took more proliferative cases. This is where they could be obtained economically. We took the extras in these plans from the less interesting glaucoma cases. West was scaled back across the board due to the limited number of health plan members available. The extra cases in the budget from the West scale back were reallocated to the larger SE, MW1, MW2, and NE plans.

Table 3 presents the design effects as well as the design effects for a simple probability proportional to size design for comparison. In this case PPS uses just the unweighted patient count as the size measure. The design effects are relatively low for a design of this complexity. Background design effects are all 1.16 or lower. The glaucoma design effects get as high as 1.28 in MW1. This is acceptable since we have much higher sample sizes for glaucoma. The largest proliferative design effect was 1.37 in MW2. These are the plans where we attempted to get the largest number of proliferative cases because they had larger numbers available. Despite the constraints, relatively high effective sample sizes were obtained relative to the actual number of cases per condition.

Following each design effect row in Table 3 is the design effect for a simple PPS design. This design has the same constraints as the final design but lacks the freedom to reweight the size measure and adjust the λ . After sampling proportional to size the sampling, fractions for each of the three conditions

Table 2: Properties of the Final Sample

	<i>SE</i>	<i>MW1</i>	<i>MW2</i>	<i>NE</i>	<i>West</i>	<i>MW3</i>	<i>Total</i>
Background Cases	60	40	46	44	40	50	280
Glaucoma Cases	192	120	166	169	48	169	864
Proliferative Cases	60	137	89	82	27	69	464
Total Cases	312	297	301	295	115	288	1,608
Number of Practices	27	30	26	32	20	39	174
Number of Distant Practices	3	1	3	0	3	1	11
Number of Physicians	52	55	46	60	31	66	310
Background Effective Sample Size	54.4	35.1	40.9	39.8	38.3	43.0	251.4
Glaucoma Effective Sample Size	151.8	93.6	141.2	149.0	44.9	147.1	727.5
Proliferative Effective Sample Size	55.8	102.0	65.0	70.3	26.1	65.7	385.0

Table 3: Design Effects Compared to Probability Proportional to Size Design Effects

	<i>SE</i>	<i>MW1</i>	<i>MW2</i>	<i>NE</i>	<i>West</i>	<i>MW3</i>
Background Design Effect	1.10	1.14	1.13	1.11	1.04	1.16
Background Design Effect-PPS	1.19	1.09	1.16	1.76	1.04	1.08
Glaucoma Design Effect	1.26	1.28	1.18	1.13	1.07	1.15
Glaucoma Design Effect-PPS	1.21	1.18	1.55	1.37	1.03	1.79
Proliferative Design Effect	1.07	1.34	1.37	1.17	1.03	1.05
Proliferative Design Effect-PPS	1.10	1.37	1.35	1.13	1.04	1.04

were set to achieve the same actual sample sizes as the final design. The constraint of no more than 10 patients per condition per physician was maintained. This would be a typical design for this problem. In SE and West this design would not produce much difference in effective sample sizes from our final design. Note that in SE we even selected a design with a somewhat larger design effect for glaucoma in order to decrease the proliferative design effect. In MW1 we improved the proliferative design effect from 1.34 to 1.17 with this method. In MW2, NE, and MW3 we avoided some fairly severe, as large as 1.79, design effects with improved design.

The software that simulated and evaluated these designs was written in SAS, version 8. It is available by request from the authors at John_Adams@rand.org.

DISCUSSION

We have presented a simulation-based approach for examining multiple sampling goals for a multistage study design. In evaluating the sampling goals, we use simulations to explore a more expansive collection of possible designs than the typical study considers.

The potential significance of our research is in its contribution to more economical study design. These economies are threefold. First, it is possible to produce sampling plans that may produce more of everything in some cases. The removal of the constraint that the researcher must use one of a relatively small collection of standard approaches can yield better designs for all of the objectives simultaneously in some cases. Second, these methods allow the design team to consider more options more thoroughly. Third, automating the process facilitates adapting to evolving information about the population

available to the study. We believe that good tools that help in the design process will prove to be very useful to the researcher.

Another useful feature of this approach is that it can be easily adapted to problems that arise in the field. For example, samples must occasionally be reduced because of cost overruns; an unexpectedly large number of practices or physicians may refuse to participate or additional funds may make sample expansion possible. Dealing with all these problems while still maintaining a probability sample is straightforward. The simulations are merely expanded to include the additional actions being considered and new sampling probabilities, design effects, and effective sample sizes are calculated.

One potential criticism of this approach is that it can be time-consuming to develop the simulations and analyze them. Whether the investment in this approach is worthwhile depends on the scale of the project. Larger studies can more readily bear the costs of more elaborate design efforts since design costs do not increase as fast as study size increases. In the CECS we estimated the marginal cost per case for data collection as \$175–\$200. Considering only the four condition-plan combinations with the greatest improvement in design effects (MW2, MW3, and NE glaucoma and NE background) the value of the additional information that results from improved design is \$36,000 to \$42,000. This is greater than 10 percent of the medical records abstraction costs. The larger the study, the more attractive the savings are likely to be.

NOTES

1. The intraclass correlation is the correlation between two patients sampled from the same physician's practice. It is a consequence of patients within a physician practice typically being more similar than patients from different practices.
2. In order to smooth these maxima by bins we fit a generalized additive model (Hastie and Tibshirani 1990) with the proliferative effective sample size as the dependent variable and the background and glaucoma effective sample sizes as predictors including nonparametric smoothed main effects and various parametric interactions. This smoothing would not be necessary if a fine enough grid of design parameters were used in the simulations. The *S-plus* software package was used to produce the contour plots.

REFERENCES

- Biemer, P. 1983. "Optimal Dual Frame Sample Design: Results of a Simulation Study." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 630–5.

- Chromy, J. R. 1987. "Design Optimization with Multiple Objectives." *Proceedings on the Research Methods of the American Statistical Association*, 194–9.
- Cochran, W. G. 1977. *Sampling Techniques*, 3d ed. New York: Wiley.
- Folson, R. E., F. J. Potter, and S. R. Williams. 1987. "Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains." *Proceedings on the Research Methods of the American Statistical Association*, 792–6.
- Gabler, S., S. Haeder, and P. Lahiri. 1999. "A Model Based Justification of Kish's Formula for Design Effects for Weighting and Clustering." *Survey Methodology* 25 (1): 105–6.
- Hastie, T. J., and R. J. Tibshirani. 1990. *Generalized Additive Models*. New York: Chapman and Hall.
- Kish, L. 1965. *Survey Sampling*. New York: Wiley.
- . 1974. "Optimal and Proximal Multipurpose Allocation." *Proceedings of the Social Statistics Section, American Statistical Association*, 111–18.
- SAS, version 8. Cary, NC: SAS Institute.