Table 7-3. Tabular Summary of Estimated Retention Rates for Each Design Option

Years of Follow-up	Estimated Retention Rates			
	Probability-Based (National)[a]	Probability-Based (Center)[b]	Center-Based (Patient Lists)[c]	Volunteer Sample[d]
0	100%	100%	100%	100%
1	92.5%	94.0%	98.5%	98.8%
2	85.6%	88.3%	96.9%	97.7%
3	79.2%	83.0%	95.4%	96.6%
4	73.3%	78.0%	94.0%	95.4%
5	67.8%	73.3%	92.5%	94.3%
6	62.7%	68.9%	91.1%	93.2%
7	58.0%	64.7%	89.7%	92.2%
8	53.7%	60.8%	88.3%	91.1%
9	49.7%	57.1%	86.9%	90.0%
10	46.0%	53.7%	85.6%	89.0%
11	42.5%	50.5%	84.3%	88.0%
12	39.4%	47.4%	83.0%	86.9%
13	36.4%	44.6%	81.7%	85.9%
14	33.7%	41.9%	80.4%	84.9%
15	31.2%	39.4%	79.2%	84.0%
16	28.8%	37.0%	78.0%	83.0%
17	26.7%	34.8%	76.8%	82.0%
18	24.7%	32.7%	75.6%	81.1%
19	22.8%	30.7%	74.4%	80.1%
20	21.1%	28.8%	73.3%	79.2%
21	19.6%	27.1%	72.2%	78.3%

a. Calculated using a first-order exponential decay model with the rate of decline in the retention rate assumed to be five times higher than that of a Center-Based (Patient List) approach.
b. Calculated using a first-order exponential decay model with the rate of decline in the retention rate assumed to be four times higher than that of a Center-Based (Patient List) approach.
c. Calculated as the weighted average of parameter estimates derived from fitting a separate first-order exponential decay model to each study.
d. Calculated using a first-order exponential decay model with the rate of decline in the retention rate assumed to be 1.3 times lower than that of a Center-Based (Patient List) approach.

7.3.1 Limitations in the Data

In addition to not having any studies that have the same scope, size, and design of the NCS there are three primary limitations inherent in the data used to estimate the initial response rates and the long-term retention rates. First, it is important to note that only a limited number of studies were examined and information on response and retention rates for these studies was not always available. In some part the limited number of studies examined is a function of limiting the literature search to only those studies that were thought to be the most relevant. However, it is also due to the fact that only a limited number of large, longitudinal studies of children remotely similar to the NCS have been conducted. Potentially relevant studies were identified through a keyword literature search, by examining bibliographic references in published studies that were previously identified as potentially relevant, and though Internet searches. Nevertheless, the identified studies may not be truly representative of a larger population of

longitudinal studies. That is, there might be studies that were not considered that would suggest different assumptions on response rates or retention rates.

Many of the studies where long-term retention rates were available for use in this analysis were conducted in foreign countries. Thus, the response rates and retention rates observed in these studies may not be representative of a longitudinal study conducted in the U.S. For example, some European countries have socialized medicine, which may result in higher initial response rates and retention rates.

As previously discussed, the identified studies did not uniformly begin with a child's birth (or during the preconception and/or prenatal periods as being planned in the NCS). To include these studies in the modeling activity, it was necessary to assume that the retention rates were a function of the number of years of follow-up and not age of the study participant. However, this may not be a completely appropriate assumption because drop-out from participation in a longitudinal study such as the NCS may be impacted by many factors, including major life events such as birth, adoption, divorce, changing jobs, etc. Over time, as these issues are resolved by family members, participation in the study may again be an option for the household/child (Bender et al. 1997). Turner and Le Souef (2003) found that "…parents had refused to let their child participate at 6 years, but were agreeable five years later, suggesting that parents of younger children are less likely to allow participation in research studies." Thus, there is some evidence in the published literature that participation in a longitudinal study is related to the age of the participant.

Finally, despite the fact that over 20 studies were examined, only a limited amount of information is available that can be used to estimate initial response rates and retention rates over time. In particular, initial response rates were identified for only one Center of Excellence study. Retention rates that could be modeled over time were essentially obtained only for studies that employed a hospital/physician/center based sampling approach, requiring assumptions for applying the modeling results to other sample designs.

7.3.2 Alternative Modeling Approaches

As discussed in Section 7.2, a modeling approach consisting of fitting separate first-order exponential decay models to each study and then calculating a weighted average of the estimated model parameters was employed for this study. As in any analysis of this type, alternative modeling approaches could be employed. The following discusses two likely alternative approaches that could be employed and the impacts that each would have on the estimated retention rates.

One possibility for modeling the retention rates over time would be to treat all observations as if they came from a single study and fit a model to the combined data (e.g., a first-order decay model). This approach is attractive because it results in uniformly higher estimated retention rates (2% to 7% higher depending upon the number of years of follow-up) than the employed approach. However, this modeling approach includes an implicit weighting scheme where studies with more observations (i.e., more information available on retention rates over time) are weighted higher than studies with fewer observations, regardless of where in the

follow-up period the retention rates were observed. For example, a study with six observed retention rates at years 1 to 6 could be twice as influential as a study with three observed retention rates at years 18 to 20.

A second possible modeling approach would be to use a more complicated model that would facilitate the modeling of rates of attrition that change as a function of years of participation in the study. For example, one attrition rate may be applicable early in the study where the rate of attrition is the greatest, and a second, lower attrition rate could be modeled after a certain point in time to allow for "long-term" study participants to drop-out of the study at a lower rate (i.e., impose even more of a "leveling" effect on the rate of attrition after a certain point in the study is reached). A model with this level of flexibility might be attractive because it allows the attrition to level off significantly after a period of time, which increases the estimated retention rates over those estimated in Section 7.2 for later years of follow-up. Figure 7-3 illustrates how the retention rates can be significantly impacted by employing a more flexible modeling approach. In Figure 7-3, the same exponential decay model was used for the convenience sample and sample of patients already affiliated with the Centers. However, for the national probability based sample and the probability-based sample of MSAs surrounding the Centers, the following model was used:

$$\text{Retention Rate} = e^{-\beta * [\sum_{i=0}^{\min(Years, X)} (X - i) + \max(0, Years - X)]}$$

where β is the rate of exponential decay associated with the PBS of Center patients, and X is a factor that takes on the value of five for the national probability-based sample and four for the probability based sample of areas surrounding the Centers. The interpretation of this equation is that for the national probability based sample, the rate of attrition is five times greater than the rate of attrition for center patients in Year 1, four times greater in Year 2, three times greater in Year 3, two times greater in Year 4, and equal for the remainder of the study. As observed in Figure 7-3, this results in retention rates for probability-based approaches from unrestricted populations at the end of the study period that are much higher than those calculated using the methods described in Section 7.2. The higher estimated retention rates (shown in Figure 7-3) associated with the national probability-based sample and the probability-based sample of areas surrounding centers, while not supported by any specific study data, does have intuitive appeal to study planners who believe that the differences in rates of attrition associated with the mode of sampling will disappear as the study progresses. Therefore, these alternative retention rates are explored further in Chapter 10 (and in detailed appendices) with estimates of study costs and power associated with the higher retention rates associated with study participants recruited via probability-based sampling of unrestricted populations.

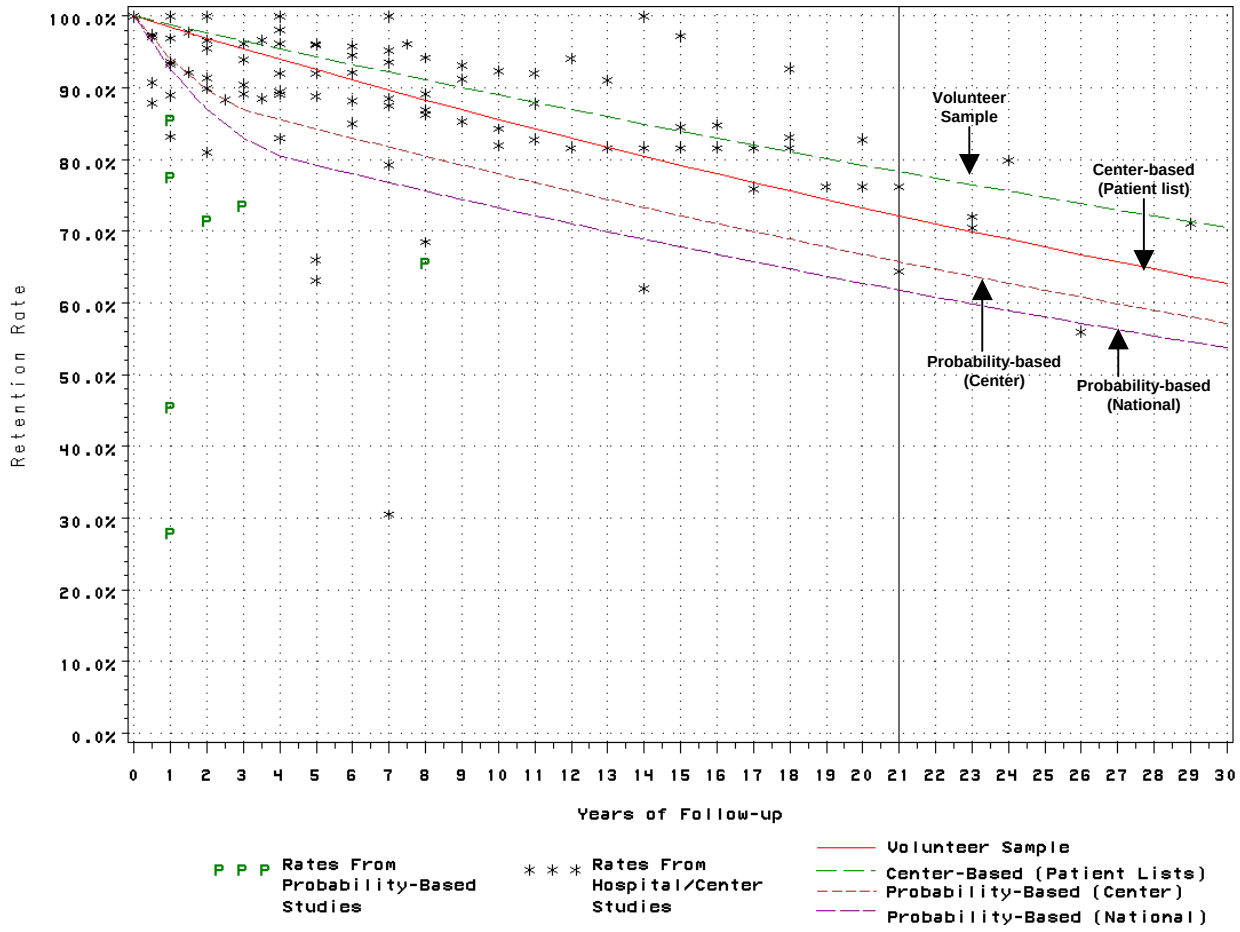


Figure 7-3. Illustration of the Impact on Retention Rates if the Attrition Rate is Assumed to Change as a Function of the Years of Follow-up

7.3.3 Factors Related to Study Approach

Each of the studies reviewed and utilized to estimate response and retention rates employed different methodologies for recruiting and retaining study participants. As discussed in Section 7.1 and Section 7.2, these methods can have a significant impact on both the initial response rates and the retention rates over time. Tracking/tracing study participants is one such factor that can have a significant impact as it may be the single most effective effort that can be conducted to reduce attrition. Additionally, it should be noted that retention rates do not always decline over time. In some cases, the retention rates may increase as participants re-enter the study. No explicit assumptions were made on the impact of tracking/tracing methods or re-entry of study participants for estimating retention rates. One possible assumption that could be employed would be to assume a "leveling" of the attrition rate after some point in the study, as previously discussed.

8 COST ESTIMATES FOR NCS SAMPLING DESIGN OPTIONS

A component of the decision regarding the sampling approaches used to conduct the NCS involves the cost of conducting the study under various different design scenarios. Potential cost differentiators among the four sampling approaches – a National PBS (NPBS), a PBS of the geographic area around a Center (area PBS), a PBS of Center patients, and a purposive sample of Center patients – will result in differing cost estimates based on the proportion of the sample drawn from each of these frames. This chapter of the report presents our thoughts on the cost differentiators among the four sampling approaches within each of seven major activity areas, the basis for estimates of fixed and variable costs within each of the seven areas, and two sets of cost estimates developed based on these assumptions. The first set presents cost estimates for 25 design options constrained to enrolling 100,000 live births. The second set presents estimates for the same 25 design options all constrained to a total budget of approximately $2.7 billion. For most design options, enrolling 100,000 live births leads to a total cost of more than $2.7 billion.

Caution should be used in interpreting these cost estimates. Many assumptions were made regarding retention rates, number and frequency of samples obtained from participants, and operational and management costs over a 25-year period. Changing one or more of these assumptions can have significant impacts on the overall cost of the strategies. On the other hand, given the assumptions, these estimates do allow cost impacts of various design strategies to be compared. Note that to produce these cost estimates we developed a model whose inputs can easily be revised to produce new estimates based on different assumptions.

Note also that these cost estimates were developed by Battelle, not the NCS nor the lead Agencies. As mentioned, a number of simplifying assumptions were made for the purpose of making general cost comparisons among designs. By releasing these cost estimates to inform discussions at the Sampling Workshop, Battelle and the NCS do not imply that these estimates and assumptions are accurate, nor do they represent decisions, assumptions, or budget projections made by NICHD.

8.1 COST DIFFERENTIATORS AMONG THE FOUR SAMPLING FRAMES

Although each of the four sampling frames contains features that could lead to significant cost differences, a significant portion of the costs will likely be similar regardless of the sampling frame used. For example, the resources required to conduct the work in the areas of Sample Analysis/Storage and Data Management are largely independent of the mix of sampling frames across the various design options. The costs for laboratory analysis of samples, for example, is independent of the sampling frame that led to those samples being collected assuming that relatively few laboratories will have responsibility for sample analysis. We define a "cost differentiator" as a cost element that is estimated to be significantly affected by the choice of a sampling frame. Thus, the identified cost differentiators determine the differences in costs between the different frames. As noted above, however, some study costs are fairly independent of the choice of sampling frame and therefore the cost differentiators do not necessarily result in large percentage differences in the overall study cost.

Cost differentiators associated with conducting the national probability-based sample include:

- Operation of local study offices in a large number of areas around the country,
- Operation in geographic areas that do not include participating research centers,
- Additional training and QA/QC costs because of the large number of PSUs,
- Recruitment of randomly selected participants who are not associated with a participating research center in any way, and
- Relatively low retention rates because of how those participants are recruited.

The last two bullets do lead to significant differences in the cost estimates provided later. Relatively low retention rates lead to less data collection, sample analysis, and data management and, thus, lower costs. It will be extremely important, however, to assess the impact of the reduced longitudinal sample size that would be associated with this reduced cost.

There are also a number of cost differentiators involved in conducting the study at selected research centers.

- Lower recruitment cost – Recruitment rates for the probability sample of the geographic area surrounding a center are estimated to be higher (and recruitment cost subsequently lower) than those for a national probability-based sample because of local familiarity with the institution coordinating the study in the area. Similarly, recruitment rates for the probability sample of center patients are estimated to be significantly higher because of participants' existing relationship with the center.
- Higher start-up cost – The process of selecting qualified centers and beginning their operations will involve additional cost from the selection process, establishing contracts or cooperative agreements with each organization, and potentially conducting separate human subjects reviews at each organization.
- Higher operating cost – It is likely that the operating expenses of a participating center will be higher than the operating expenses of an organization managing a study office in an area unassociated with a center. Centers associated with universities and hospitals will have overhead costs that increase overall operating expenses.

8.2 THE SEVEN MAJOR COST AREAS

Cost estimates were developed for seven major areas of work, which we will introduce below. For costing purposes, all study-related costs were assigned to one of these seven categories.

8.2.1 Study Design and Start-Up

The Study Design and Start-up area includes all costs involved with planning, designing, and implementing the study up to the point at which participant recruitment can begin. Activities include:

- Sampling strategy design,
- Questionnaire and data collection tool development including pilot studies and focus groups,
- Preparing IRB materials and participating in IRB reviews,
- Preparing OMB clearance materials and obtaining approval,
- Preparing Quality Assurance Project Plans and Standard Operating Procedures,
- Preparing training material and delivering training to all study personnel, and
- Selecting study centers, contractors, and other participating organizations.

8.2.2 Recruitment

Recruitment covers the work involved in selecting participants in each of the sampling frames, recruiting all participants, and operating study offices and centers over the period of study recruitment. Start-up costs associated with opening study offices and centers, including equipment purchase, are incorporated into the recruitment area. For the cost estimates provided in this report, we assumed that recruitment would last three years; however, the Business Plan estimated that recruitment could last three to five years.

8.2.3 Data Collection

The Data Collection cost area composes the majority of the resources required to conduct the study. This area includes operating study offices and centers for the approximately 22 years following recruitment of all participants; obtaining biological samples and other medical information; performing environmental assessment; conducting questionnaires and surveys; and performing QA/QC audits of data collection protocols. This includes the costs for data collectors to visit homes or other monitoring locations but does not include reimbursement of participants' costs, which are meant to be incorporated in the Retention/Tracking activities.

8.2.4 Retention/Tracking

The Tracking component of this cost area involves maintaining participant addresses and contact information over the course of the study and notifying the appropriate organizations regarding participant moves. The Retention component includes various activities that focus on maintaining participants' involvement in the study – preparation of study newsletters, sending birthday cards, providing incentives for completed data collection visits, etc.

8.2.5 Sample Analysis and Storage (Repository)

This area is focused on the laboratory analysis of biological (blood, urine, hair, etc.) and environmental (dust, paint, soil, etc.) samples collected from study participants. This includes all chain-of-custody and data management work performed by laboratory personnel while conducting the analyses. This is meant to include all sampling handling and preparation costs. Additional costs are included for storing all samples collected over the entire course of the study.

8.2.6 Data Management and Software Development

This work area encompasses 1) development and maintenance of software and database tools required to conduct and manage the study (e.g., a project management system allowing tracking of recruitment and retention), 2) development and maintenance of a study website, 3) data entry and validation of all study data collection forms, and 4) development, maintenance, and updating of study databases containing all data gathered over the course of the study.

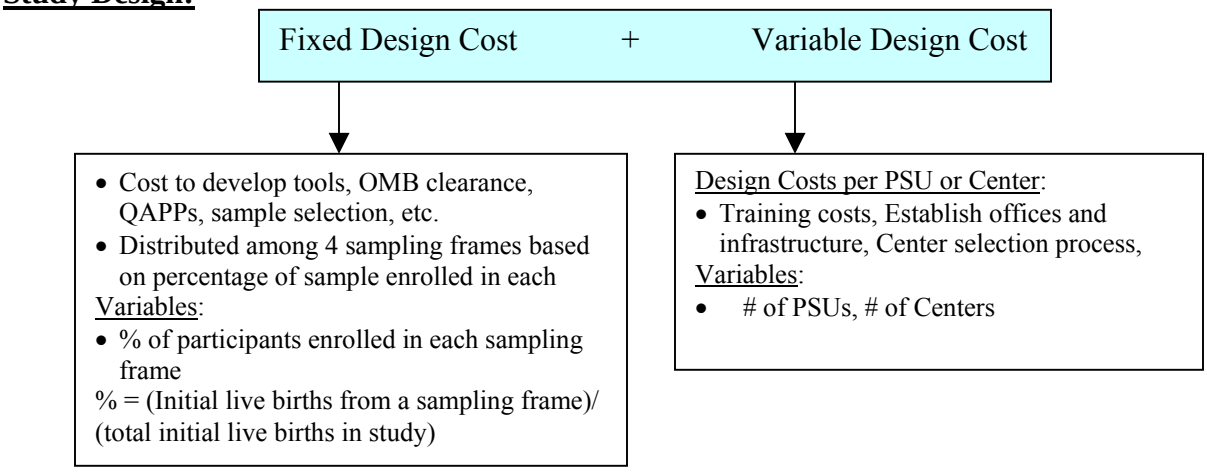
8.2.7 Project Management

This work area includes the Federal government's cost to oversee and manage the study as well as the cost of central coordinating centers or organizations that will assist the government in coordinating all aspects of the study.

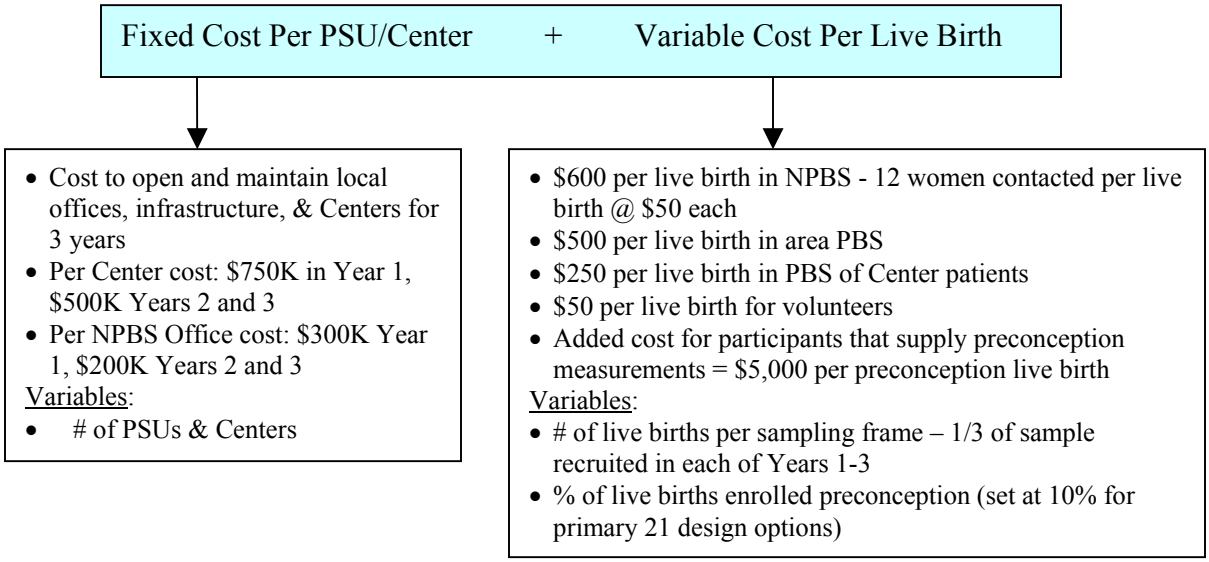
8.3 THE COST ESTIMATING MODEL

For each of the seven work areas, we focused on developing separate estimates of fixed and variable costs associated with conducting the study under each of the four sampling frames. The figures below provide details on the cost assumptions used to develop and model the costs in each of the seven areas. The model was developed using Microsoft Excel. As noted previously, any of the cost inputs or assumptions can easily be changed and all cost estimates automatically revised. This initial cost model uses estimated retention rates for each sampling strategy that were developed separately. Because the retention rates already reflect the different nature of these strategies, costs to retain participants are assumed to be identical across the four strategies. If retention costs were to vary by sampling strategy, then that would have to be reflected in the associated retention rates.

Study Design:

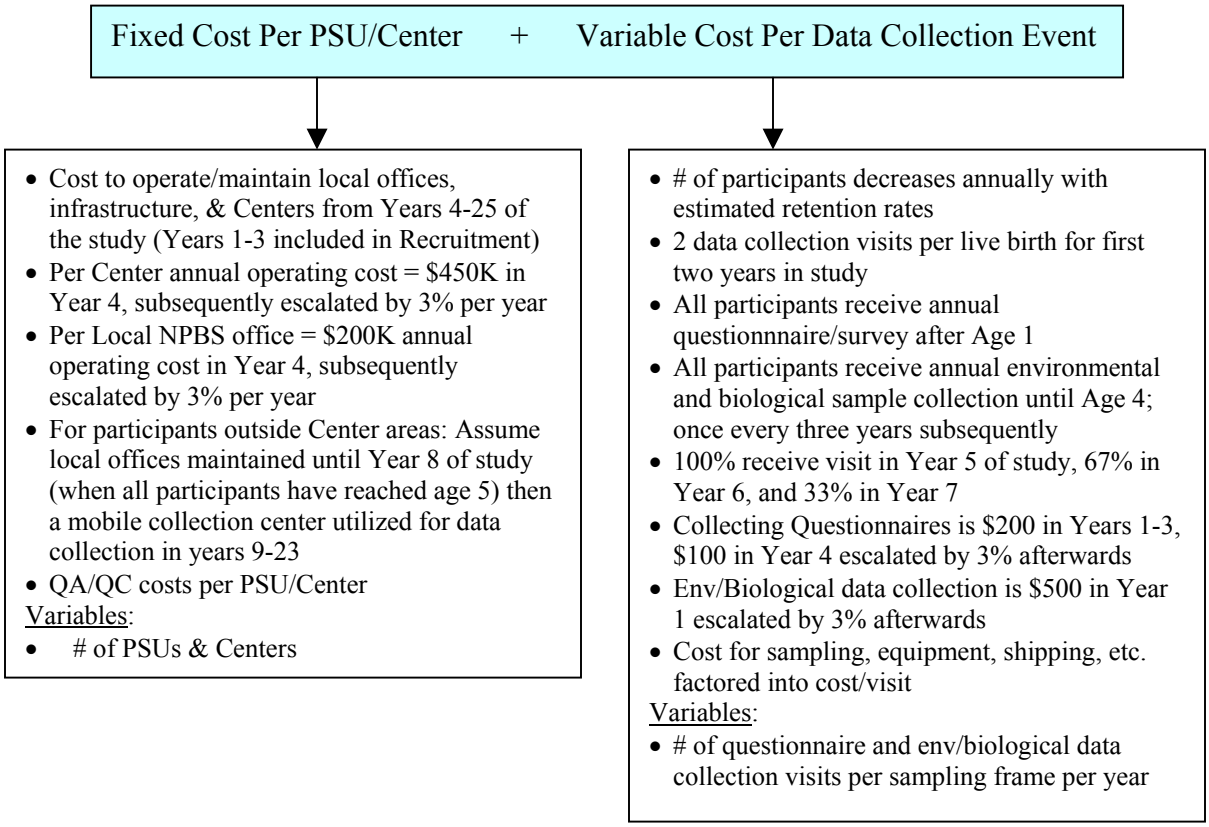


Recruitment:



Note that differing recruitment rates across the sampling frames is reflected in the variable cost per live birth. It is assumed that it will be most difficult to recruit participants into the national PBS as reflected by the $600 cost per live birth. We assume slightly lower costs for the area PBS because of the local presence and recognition of the medical centers. As a starting point, the percent of live births enrolled pre-conception was set at 10 percent for the 23 primary design options; however, two additional options were investigated from a cost perspective with this percentage set at 5 and 25 percent, respectively. Also, we assumed for costing purposes that pre-conception enrollments would only occur in the national and area PBSs, not in the PBS of Center patients or the purposive sample of Center patients.

Data Collection:



Note that in this version of the costing model, data collection costs per visit are assumed to be identical across the different sampling frames. It is possible that some data collection, particularly the biological and medical, may be less expensive in the Center-based sampling frames if costs of equipment, personnel, etc. are being at least partially contributed by the Centers. Future revisions could investigate the cost implications of varying data collection expenses across the sampling frames. Also, data collection costs are meant to represent an average per-person cost per data collection event (with data collection costs being lower for those who undergo the minimal core data collection protocol, and higher costs associated with participants who undergo more detailed data collection efforts). The costs of maintaining medical and environmental monitoring equipment over time are incorporated in the annual fixed cost of operating the study Centers and offices.

Retention/Tracking:

Fixed Cost for Tracking Participants + Variable Cost Per Participant

- Cost to maintain current addresses and contact information and inform data collection organizations of moves/updates
- Cost of producing/mailing semi-annual newsletters to participants
- Assume a single CO handles all tracking responsibilities
- Distribute costs across sampling frames based on % of participants from each frame annually
- Year 1 = $500K, Year 2 = $750K; Year 3 = $1M; Years 4-10 = increase by 10% per year; Years 11-17 = Year 10 cost; Years 18-23 = add 10% per year

Variables:

- % of annual participants in each sampling frame

- Annual cost of incentives
- Assume same incentive structure across all sampling frames since we are using variable retention rates
- $100 annually per participant until Year 8, $125 in Year 9, $150 Years 10-20, $200 Years 21-25
- Structure reflects 3 levels of incentives - $100 annually until Age 7, $150 Ages 8-18, $200 Ages 19-21

Variables:

- # of participants per sampling frame per year

In this initial costing model, the $100/$150/$200 incentive structure can be viewed as an average level of incentives provided to participants. Assuming that some participants are selected for more intensive data collection efforts, we may want to incorporate a higher level incentive structure that would be provided to that subset of participants in any revisions to this cost model.

Sample Analysis & Storage:

Variable Cost Per Data Collection Event

Environmental & Biological Samples

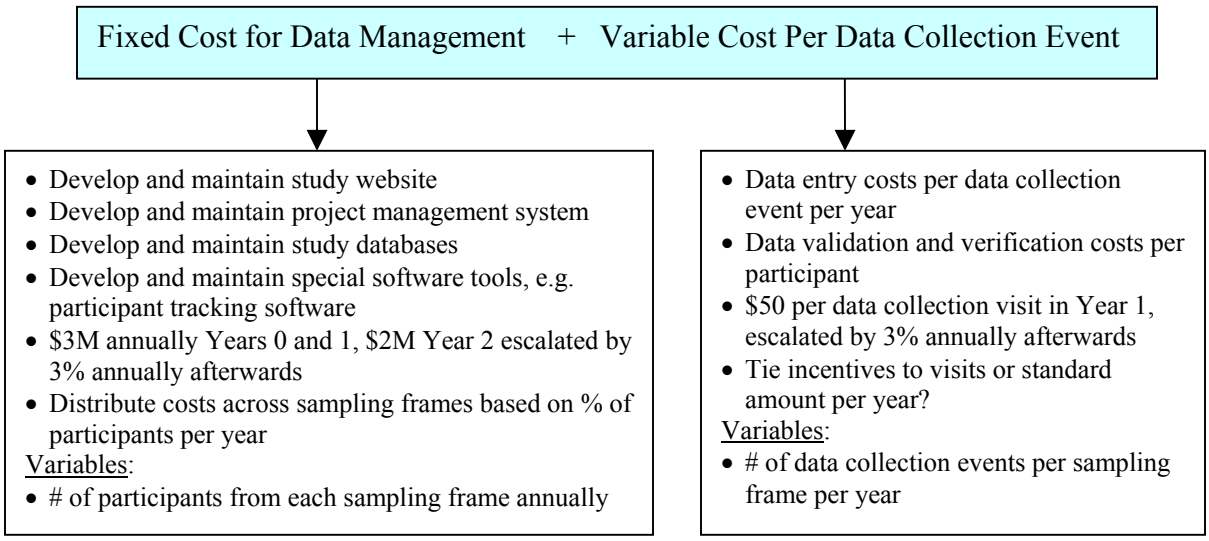
- # of samples per participant per year
- $100 Year 1 cost for analyzing all environmental samples from a data collection visit
- $300 Year 1 cost for analyzing all biological samples from a data collection visit
- Factor in 3% annual inflation rate for sample analysis
- All participants sampled for first 5 years of study, 2/3 of participants in Year 6, 1/3 in remaining years
- Factor in an annual storage cost based on cumulative # of sample collected
- Storage - $.10 per cumulative sample per year ($100K annually per million samples)
- Assumes all samples stored for full length of study (which may not be feasible)

Variables:

- % of data collection events per sampling frame per year

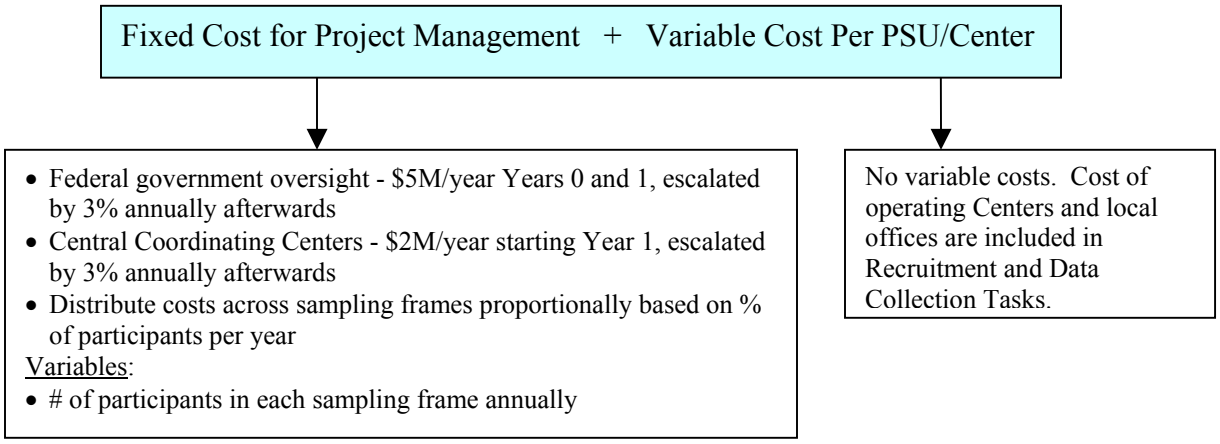
Setting original estimated analysis costs was difficult because of not knowing exactly how many samples will be collected at each data collection point and of what type those samples will be. These original estimates can be replaced with more exact sample analysis cost estimates as necessary to be consistent with the critical measures presented in Chapter 6. The estimated $400 per data collection event for analysis of biological and environmental samples represents an average across all participants. Although all potential analyses that may be conducted on a set of samples from a single data collection event may add up to a much higher cost, archiving of many samples and use of outcome-dependent sampling designs may contribute to lowering the average per participant analysis cost.

Data Management & Software Development:



For this version of the costing model, we assumed equal data management costs across the sampling frames; however, it could be argued that the probability-based samples will generate more data to enter, store, and manage. For example, data collection (screening) forms may need to be completed, entered, and stored for each recruitment attempt. As the national PBS and PBS of MSAs surrounding Centers are expected to have lower recruitment rates, more data would be generated during the recruitment process. In future versions, cost per data collection visit can differ across the sampling frames, if necessary.

Project Management:



The cost model distributed the fixed project management costs proportionally across the sampling frame based on the annual number of participants from each. Similar to the discussion on data management costs above, however, an argument can be made that the inclusion of a national PBS leads to a more complex management structure because of the additional geographic and organizational diversity that will be added. For this reason, we may want to consider altering the model in the future to assign a higher percentage of the management costs to the national PBS frame.

8.4 DESIGN STRATEGIES

For this exercise, we developed cost estimates under two different frameworks. The first framework requires recruitment of 100,000 live births, while the second framework requires a total budget of approximately $2.7 billion. Under each framework we costed 25 design strategies. For the first 21 strategies, the variables are the number of PSUs, the number of Centers, and values of P_1 through P_4 (see Table 8-1 for definitions of these variables):

- Three levels of PSUs are used – 0, 50, and 100 (for 0 PSUs, P_1 also equals 0);
- Four levels of P1 are used – 0.25, 0.5, 0.75, and 0 (for the 0 PSU options)
- Three combinations of P_2-P_3 are used for the remaining sample - 0.25:0.75, 0.5:0.5, and 0.75:0.25 (except for option A2, which sets P_2 to 0),
- P_4 is set at either 0.03, 0.02, or 0.01 (except for Option A1, in which it is set to 0.04, and Option A2, in which it is set to 0.5);
- Each Center is allocated 2,000 study recruits. In the sample size-constrained framework, the three levels of Centers are 38, 25, 13, and 50, which correspond to 75,000, 50,000, 25,000, and 100,000 Center recruits. In the budget-constrained framework that allows a varying number of recruits, the number of Centers varies.

In these first 23 strategies, it is assumed that 10% of the live births in the NPBS and area PBS come from a sample recruited preconception. All live births recruited into the PBS of Center patients and volunteer sample are assumed to be enrolled prenatally. In the last two

strategies (Group I), one design strategy is re-costed with two different levels of preconception recruits in P_1 and P_2 – 5 and 25 percent. Table 8-1 lists all the design strategies for which the cost model was run. Group A represents two options with all participants drawn from a PBS of Center patients and a volunteer sample. Group B represents options with no participants from a NPBS. Groups C through E represent design options with 50 PSUs and participants drawn from each of the four candidate sampling frames, while Groups F through H represent these same options with 100 PSUs. Each individual group represents a single number of Centers and percentage of the sample drawn from a national PBS, with the percentage of live births drawn from each of the Center-based approaches varying across each option.

Note that the variable that indicates the number of participants in each design option is included for the cases where we fix the overall cost of the study, altering the number of participants in order to meet the cost constraint.

Table 8-1. 25 Options for Sample size-Constrained Costing Model

Group	Cost Estimate Run	# of PSUs	# of Centers	National PBS (P_1)[1]	Area PBS (P_2)[2]	PBS of Center Patients (P_3)[3]	Opportunity Sample (P_4)[4]
A	1	0	50	0%	0	97%	3%
	2				0	50%	50%
B	3	0	50	0%	24%	72%	4%
	4				48%	48%	4%
	5				72%	24%	4%
C	6	50	38	25%	18%	54%	3%
	7				36%	36%	3%
	8				54%	18%	3%
D	9		25	50%	12%	36%	2%
	10				24%	24%	2%
	11				36%	12%	2%
E	12		13	75%	6%	18%	1%
	13				12%	12%	1%
	14				18%	6%	1%
F	15	100	38	25%	18%	54%	3%
	16				36%	36%	3%
	17				54%	18%	3%
G	18		25	50%	12%	36%	2%
	19				24%	24%	2%
	20				36%	12%	2%
H	21		13	75%	6%	18%	1%
	22				12%	12%	1%
	23				18%	6%	1%
I	22	50	38	25% (5% preconception 95% pre-birth)	36% 5% preconception 95% pre-birth	36% 100% pre-birth	3% 100% pre-birth
	23			25% 25% preconception 75% pre-birth	36% 25% preconception 75% pre-birth	36% 100% pre-birth	3% 100% pre-birth

* For Groups B through H, live births from national PBS and area PBS are composed of 10% tracked since preconception and 90% tracked since pre-birth. 100% of participants from Center patient lists and opportunity sample are tracked from pre-birth.
[1] The percent of participants recruited for a national PBS
[2] The percent of participants recruited for a PBS in the geographic area covered by a Center
[3] The percent of participants recruited for a PBS of Center patients
[4] The percent of participants recruited for an opportunity sample

8.5 COST MODELING OUTPUT

Tables 8-2 and 8-3 contain summaries of the sample size-constrained and budget-constrained cost estimate frameworks, respectively. Detailed output reporting costs by work area and sampling design option for each design strategy is provided in Appendix H. Much of the difference in cost between design options is driven by the retention rate differences among the sampling frames.

Table 8-2. Summary of Sample Size-Constrained Cost Estimates for 25 Design Strategies

	Constraints							
Strategy	# of PSUs	# of Centers	# Parts. NPBS	# Parts. Area PBS	# Parts. PBS-Patients	# Parts. Volunteer	Cost (millions)	# Year 20 Participants
A1	0	50	0	0	97,000	3,000	$3,473.4	76,956
A2	0	50	0	0	50,000	50,000	$3,449.1	79,400
B3	0	50	0	24,000	72,000	4,000	$3,325.1	66,928
B4	0	50	0	48,000	48,000	4,000	$3,177.4	56,848
B5	0	50	0	72,000	24,000	4,000	$3,029.7	46,768
C6	50	38	25,000	18,000	54,000	3,000	$3,333.1	56,871
C7	50	38	25,000	36,000	36,000	3,000	$3,222.2	49,311
C8	50	38	25,000	54,000	18,000	3,000	$3,111.3	41,751
D9	50	25	50,000	12,000	36,000	2,000	$2,982.8	46,814
D10	50	25	50,000	24,000	24,000	2,000	$2,908.9	41,774
D11	50	25	50,000	36,000	12,000	2,000	$2,835.0	36,734
E12	50	13	75,000	6,000	18,000	1,000	$2,648.7	36,757
E13	50	13	75,000	12,000	12,000	1,000	$2,611.6	34,237
E14	50	13	75,000	18,000	6,000	1,000	$2,574.6	31,717
F15	100	38	25,000	18,000	54,000	3,000	$3,675.2	56,871
F16	100	38	25,000	36,000	36,000	3,000	$3,564.3	49,311
F17	100	38	25,000	54,000	18,000	3,000	$3,453.4	41,751
G18	100	25	50,000	12,000	36,000	2,000	$3,324.9	46,814
G19	100	25	50,000	24,000	24,000	2,000	$3,251.0	41,774
G20	100	25	50,000	36,000	12,000	2,000	$3,177.2	36,734
H21	100	13	75,000	6,000	18,000	1,000	$2,990.8	36,757
H22	100	13	75,000	12,000	12,000	1,000	$2,953.7	34,237
H23	100	13	75,000	18,000	6,000	1,000	$2,916.7	31,717
I24	50	38	25,000*	36,000*	36,000	3,000	$3,207.1	49,311
I25	50	38	25,000**	36,000**	36,000	3,000	$3,267.5	49,311
* 5% of participants recruited pre-conception								
** 25% of participants recruited pre-conception								

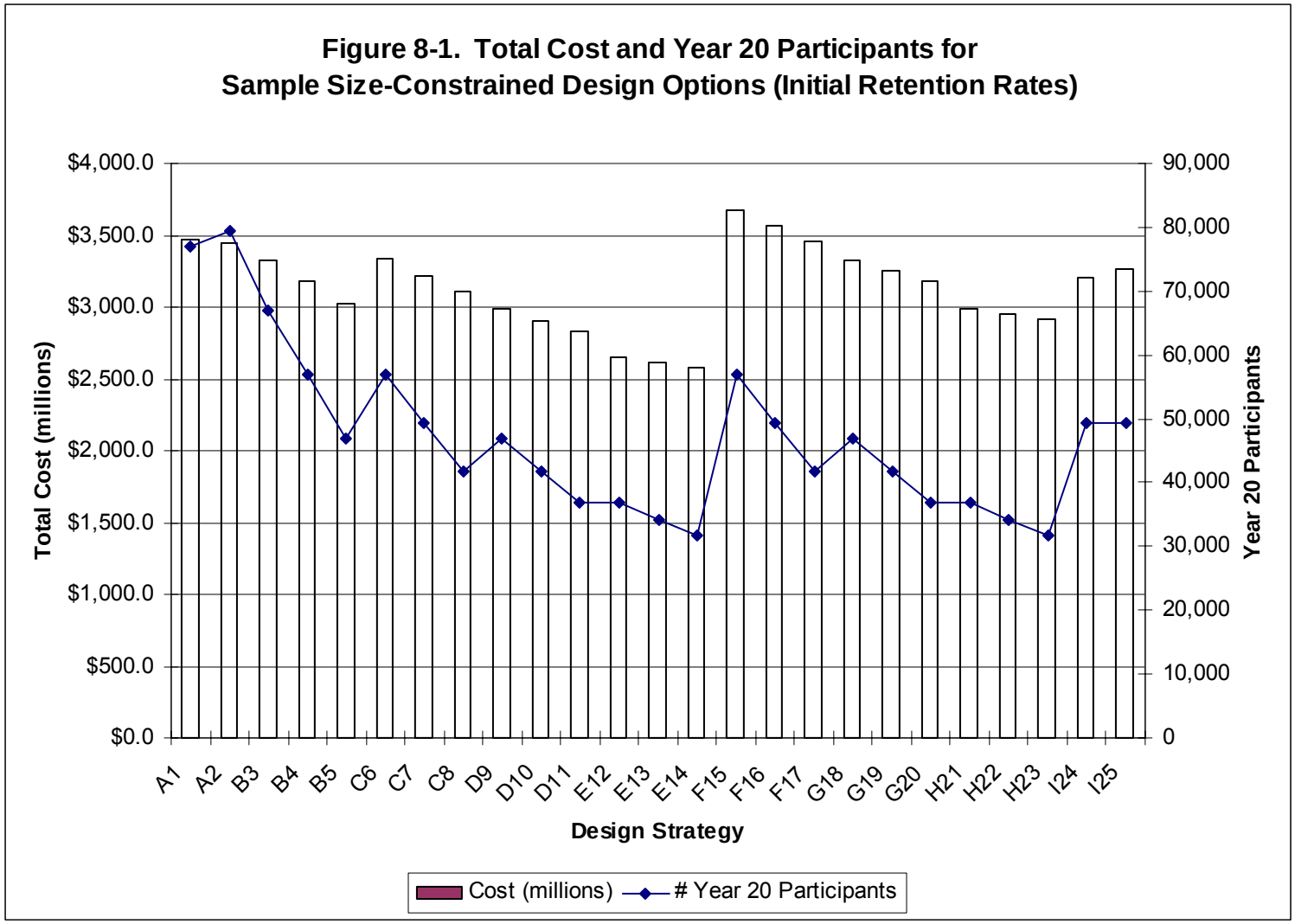


Figure 8-1. Total cost and year 20 participants for sample size-constrained design options.

Table 8-3. Summary of Budget-Constrained Cost Estimates for 25 Design Strategies

	Constraints								
Strategy	# of PSUs	% Parts. NPBS	% Parts. Area PBS	% Parts. PBS-Patients	% Parts. Volunteers	Cost (millions)	# of Centers	Total # Live Births	# Year 20 Participants
A1	0	0	0	0.97	0.03	$2,746.9	38	76,750	59,064
A2	0	0	0	0.50	0.50	$2,705.8	38	75,750	60,146
B3	0	0	0.24	0.72	0.04	$2,735.0	40	80,000	53,542
B4	0	0	0.48	0.48	0.04	$2,728.5	42	84,000	47,752
B5	0	0	0.72	0.24	0.04	$2,710.3	44	88,000	41,156
C6	50	0.25	0.18	0.54	0.03	$2,728.4	29	77,000	43,791
C7	50	0.25	0.36	0.36	0.03	$2,715.7	30	80,000	39,449
C8	50	0.25	0.54	0.18	0.03	$2,705.3	31	83,500	34,862
D9	50	0.5	0.12	0.36	0.02	$2,719.3	22	88,500	41,430
D10	50	0.5	0.24	0.24	0.02	$2,714.5	23	91,000	38,014
D11	50	0.5	0.36	0.12	0.02	$2,714.4	24	94,000	34,530
E12	50	0.75	0.06	0.18	0.01	$2,713.8	13	104,000	38,227
E13	50	0.75	0.12	0.12	0.01	$2,708.9	13	106,000	36,291
E14	50	0.75	0.18	0.06	0.01	$2,694.4	13	107,500	34,096
F15	100	0.25	0.18	0.54	0.03	$2,712.8	24	63,000	35,829
F16	100	0.25	0.36	0.36	0.03	$2,715.5	25	66,000	32,545
F17	100	0.25	0.54	0.18	0.03	$2,711.6	26	69,000	28,808
G18	100	0.5	0.12	0.36	0.02	$2,711.8	18	73,000	34,174
G19	100	0.5	0.24	0.24	0.02	$2,709.6	19	75,000	31,331
G20	100	0.5	0.36	0.12	0.02	$2,697.0	19	77,500	28,469
H21	100	0.75	0.06	0.18	0.01	$2,709.3	11	85,500	31,427
H22	100	0.75	0.12	0.12	0.01	$2,702.8	11	87,000	29,786
H23	100	0.75	0.18	0.06	0.01	$2,703.4	11	89,000	28,228
I24	50	0.25*	0.36*	0.36	0.03	$2,712.9	30	80,500	39,695
I25	50	0.25**	0.36**	0.36	0.03	$2,707.2	29	78,500	38,709
* 5% of participants recruited pre-conception									
** 25% of participants recruited pre-conception									

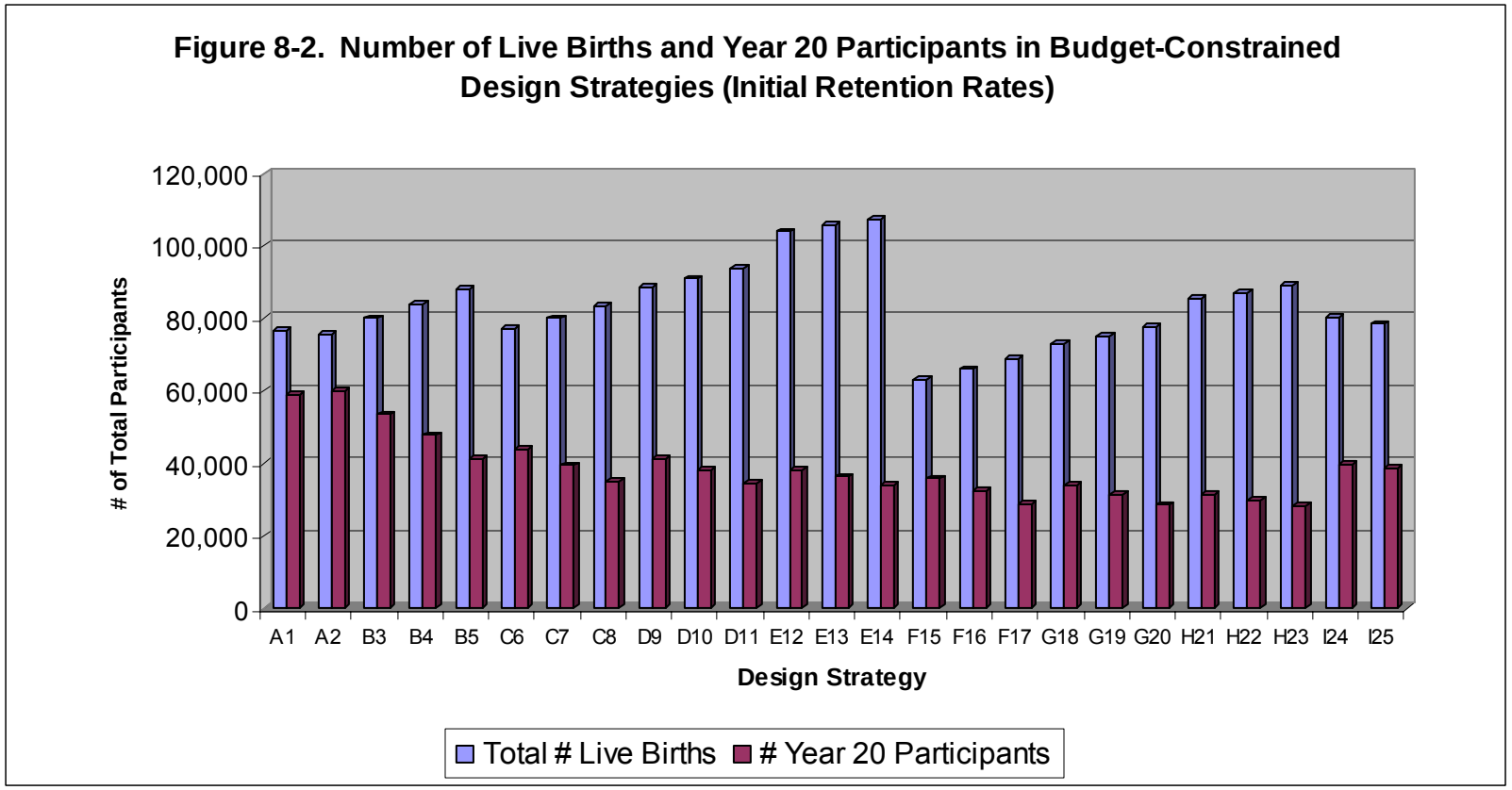


Figure 8-2. Number of live births and year 20 participants in budget-constrained design strategies.

8.6 ANALYSIS AND CONCLUSIONS FROM INITIAL COST MODELING

Review of Tables 8-2 and 8-3 and the detailed tables in Appendix H leads to a few initial conclusions regarding the primary cost drivers.

8.6.1 Effect of Retention Rates

It is extremely important to consider the impact on study quality and results, i.e., the partially noneconomic costs, in conjunction with the estimated costs. This suggests future work to estimate costs associated with maintaining similar retention rates across the sampling frames. There are two possible ways to go about this. One, work can be done to estimate the "cost" of losing participants from the study at different points in time over the course of the study in relation to the number needed to evaluate hypotheses at that life-stage. Among these costs is lost power to make significant conclusions. Two, work can be done to estimate the cost to increase the retention rates in the national and area PBSs, which have the lowest rates. Within the current cost model, setting the national PBS and area PBS retention rates equal to the retention rates of the Center patient PBS and increasing the annual incentive costs (to start at $200 and $300 annually for the doubling and tripling options, respectively) in these two frames resulted in the following changes to the cost estimates and estimated participant retention (summarized in Table 8-4):

- Within sample-size constrained cost run C7 (50 PSUs, 25% enrolled in NPBS):
  - Doubling the annual incentive costs (in conjunction with using identical retention rates for the three types of PBS) increased total estimated study cost by $603 million ($347.1 million in additional data collection, sample analysis, and data management costs and $255.5 million in additional retention costs) to $3.83 billion and raised the number of estimated Year 20 participants from 49,311 to 76,956; and
  - Tripling the annual incentive costs increased total estimated study cost by $783 million ($347.1 million in additional data collection-related costs and $435.8 million in additional retention costs) to $4.01 billion (and raised the number of estimated Year 20 participants by the same amount since all other variables were the same).
- Within sample-size constrained cost run D10 (50 PSUs, 50% enrolled in NPBS):
  - Doubling the annual incentive costs increased total estimated study cost by $758 million ($443.1 million in additional data collection-related costs and $314.6 million in additional retention costs) to $3.67 billion and raised the number of estimated Year 20 participants from 41,774 to 76,904; and
  - Tripling the annual incentive costs increased total estimated study cost by $976 million ($443.1 million in additional data collection-related costs and $533.3 million in additional retention costs to $3.89 billion (and raised the number of estimated Year 20 participants by the same amount since all other variables were the same).

Table 8-4. Estimated Additional Costs (in $Million) from Increasing Retention Rates and Incentives

Design Option	Cost Type	Costs from Initial Sample size-Constrained Model	Additional Cost from Increasing Incentives & Retention Rates	
		Original Incentives ($100/$150/$200)	Double Incentives ($200/$300/$400)	Triple Incentives ($300/$450/$600)
C7 50 PSUs 25% NPBS	Data Collection[1]	$1,613.8	$347.1	$347.1
	Retention	$220.8	$255.5	$435.8
	Total	$3,222.2	$602.6	$782.9
D10 50 PSUs 50% NPBS	Data Collection	$1,523.7	$443.1	$443.1
	Retention	$199.9	$314.6	$533.3
	Total	$2,908.9	$757.7	$976.4

[1] Includes variable data collection costs from data collection, sample analysis and storage, and data management

Note that in Appendix I we have included a full set of cost modeling output based on the revised set of retention rates introduced in Chapter 7.

8.6.2 Cost Comparisons for Options Producing the Same Size Year 20 Cohort

The number of PSUs and the percentage of participants recruited via a national PBS are factors that significantly affect overall costs. For example, comparing design strategies C6 and F10 within the sample size-constrained framework (Table 8-2), we see that doubling the number of PSUs while holding all other design variables constant increases total cost by approximately 10% (from $3.33 billion to $3.68 billion). Additionally, holding all other design variables constant, increasing the percentage of participants recruited via a national PBS leads to significantly lower costs because of lower retention rates. For example, comparing strategies C6 and D10 within the sample size-constrained framework (Table 8-2), we see that costs decrease by approximately 10% by doubling the percentage of participants recruited by a national PBS from 25,000 to 50,000. Tables 8-2 and 8-3 provide a basis for comparing designs from these two perspectives—first by fixing the initial sample size and seeing how costs differ, and second by fixing costs and seeing how the initial and final sample sizes differ.

Another way to look at differences between design options is to focus on the number of participants who remain in the cohort over the full lifespan of the study and compare design options that produce the same size Year 20 cohort. The information in Tables 8-5 and 8-6 was compiled to provide this comparison of selected design options when the year 20 sample size is the same. Table 8-5 employs the original retention rates while Table 8.6 employs the revised retention rates. Table 8-7 provides information on the characteristics of the design options that are compared in this cost analysis.

Within each line of Tables 8-5 and 8-6, information for a reference option and a comparison option is presented. For example, the first line of Table 8-5 compares Option F17 (the reference option) with Option B3. The number of live births for Option B3 has been modified to achieve the same Year 20 cohort size as Option F17. Two cost figures are reported

for the comparison option. The first is the total cost for the selected option. The second cost figure labeled "Information Acquisition Costs" is the sum of the per visit data collection costs, all sample analysis and storage costs, and the per visit data management costs for the comparison option. This figure is meant to be proportional to the volume of information collected for the cohort. Savings versus the reference option are presented in terms of both millions of dollars and as a percentage of the Information Acquisition Costs. One interpretation of the percentage figure is that it represents the percentage of additional information that could be collected for the cohort if the savings were rolled back into the Option without changing the number of live births. Comparisons include:

- Options F17 and B3 are compared to contrast an option having 100 PSUs and 79% of the cohort selected via a national or area PBS (25% national PBS) with an option having no PSUs and 25% of the cohort selected via an area PBS (no national PBS).
- Option C8 and B3 offer a similar comparison but with Option C8 having 50 rather than 100 PSUs.
- Options G19 and B5 are compared to contrast two options having similar percentages of the cohort selected via PBS of an unrestricted population, but different fractions of the cohort recruited via a national or area PBS. Option G19 has 100 PSUs and 50% of the cohort selected via a national PBS while Option B5 has no PSUs and selects none of the cohort members via a national PBS.
- Option D10 and B5 offer a similar comparison but with Option D10 having 50 rather than 100 PSUs.

It should be emphasized that, while the reference and comparison options presented in Tables 8-5 and 8-6 are equivalent in terms of Year 20 cohort size, they are decidedly not equivalent at earlier time points. This can be seen by comparing the number of live births for the reference and comparison option. In all cases, the reference scenario has a larger number of live births, which would result in higher power to address hypotheses that focus on health outcomes that can be observed in earlier stages of life. The focus in these tables is on the Year 20 cohort size, emphasizing the advantages associated with participants who have remained in the study for 20 years..

Returning the focus to the first line of Table 8-5, employing Option B3 rather than Option F17 to obtain a 20 Year cohort of size 28,808 results in a cost savings of $1,073.4 million. These cost savings are primarily due to 1) eliminating the NPBS component and the associated operations in 100 PSUs and 2) the ability to begin with a smaller number of live births because of higher retention rates for Option B3. The cost savings are equivalent to 137% of the information collection costs associated with Option B3. If the savings were rolled back into Option B3 for additional information collection, the information collection budget could be increased by 137%. Smaller savings are realized for the other comparisons in Table 8.5 because the options being compared are not as dissimilar as Options F17 and B3.

A comparison of the last columns of Tables 8-5 and 8-6 illustrates the sensitivity of the analysis to assumptions about retention rates. The revised retention rates employed in Table 8-6 assume more favorable retention behavior for cohort members selected via a national or area PBS. This results in a significant decrease in the savings achieved by shifts away from national and area probability-based sampling.

The data reported in Tables 8-5 and 8-6 show, under certain retention rate assumptions and with a singular focus on the Year 20 cohort size, that (1) significant cost savings can be achieved by shifts away from national and area probability-based sampling and (2) these cost savings could be used to significantly increase the information collection budget for the cohort. These tables also show that the magnitude of achievable savings is very sensitive to the retention rate assumptions employed in the analysis.

Table 8.5. Cost Comparisons for Options Producing the Same Size Year 20 Cohort Using Original Retention Rates

Reference Option				Comparison Option						
Option	Number of Live Births Enrolled	Year 20 Cohort Size	Total Costs	Option	Number of Live Births Enrolled	Year 20 Cohort Size	Information Acquisition Costs	Total Costs	Savings vs. Reference Option	Savings as Percentage of Information Acquisition Costs
F17	69,000	28,808	$2,711.6	B3*	43,043	28,808	$782.4	$1,661.6	$1,073.4	137%
C8	83,500	34,862	$2,705.3	B3*	52,089	34,862	$946.8	$1,917.7	$817.3	86%
G19	75,000	31,331	$2,709.6	B5*	66,992	31,331	$1,065.4	$2,149.7	$560.6	53%
D10	91,000	38,014	$2,714.5	B5*	81,283	38,014	$1,292.7	$2,539.2	$171.1	13%

*Number of live births modified to achieve same Year 20 cohort size as reference option

Table 8.6. Cost Comparisons for Options Producing the Same Size Year 20 Cohort Using Revised Retention Rates

Reference Option				Comparison Option						
Option	Number of Live Births Enrolled	Year 20 Cohort Size	Total Costs	Option	Number of Live Births Enrolled	Year 20 Cohort Size	Information Acquisition Costs	Total Costs	Savings vs. Reference Option	Savings as Percentage of Information Acquisition Costs
F17	59,000	41,584	$2,706.1	B3*	55,189	41,584	$1,056.2	$2,080.3	$665.1	63%
C8	71,500	50,395	$2,714.1	B3*	66,883	50,395	$1,280.0	$2,422.3	$323.1	25%
G19	64,000	44,619	$2,714.2	B5*	61,914	44,619	$1,163.2	$2,241.9	$479.1	41%
D10	77,500	54,031	$2,711.1	B5*	74,974	54,031	$1,408.6	$2,623.1	$97.9	7%

*Number of live births modified to achieve same Year 20 cohort size as reference option

Table 8.7. Overview of Specific Designs Explored in Cost Comparisons for Options Producing the Same Size Year 20 Cohort

Design	P_1	P_2	P_3	Fraction of the National Children's Study Recruited From				# of PSUs in National PBS	# of Purposively Selected Centers
				National PBS	PBS of Center MSAs	PBS of Center Patients	Purposive/ Convenience Sample		
B3	0.00	0.24	0.72	0.00	0.24	0.72	0.04	0	50
B5	0.00	0.72	0.24	0.00	0.72	0.24	0.04	0	50
C8	0.25	0.72	0.24	0.25	0.54	0.18	0.03	50	38
D10	0.50	0.48	0.48	0.50	0.24	0.24	0.02	50	25
F17	0.25	0.72	0.24	0.25	0.54	0.18	0.03	100	38
G19	0.50	0.48	0.48	0.50	0.24	0.24	0.02	100	25

8.6.3 Key Assumptions

All of the key assumptions we made can easily be varied to explore how different scenarios impact overall cost. For example, we made the assumption that annual Center operating costs would be higher than operating costs for the national PBS in single PSU - $500,000 vs. $200,000 in Year 1. This assumed higher overhead costs because of affiliations with a university or hospital. If this assumption is questioned, the annual operating costs can be revised and all cost runs automatically updated to reflect the new per-Center and per-PSU costs. As noted throughout the section, there are other areas that may need to be revised or improved to reflect more accurate cost estimates or more appropriate distribution of costs across the sampling frames (e.g., with a higher percentage of fixed costs being assigned to the national PBS in areas such as Project Management).

9 POWER CALCULATIONS

One of the primary uses of the NCS data will be to evaluate the study hypotheses. These hypotheses are generally concerned with assessing whether there is a significant relationship between an exposure of interest (recall that exposures are broadly defined as physical, chemical, biological, and psychosocial) and some adverse health outcome. For example, one of the hypotheses (hypothesis 1.1 discussed in Chapter 6) poses the question of whether impaired glucose metabolism during pregnancy, among women without diabetes prior to pregnancy, is related to a variety of birth defects such as major congenital malformations of the heart or central nervous system. In this Chapter we present a set of examples that illustrate the ability of a sample to *detect relationships of interest* identified in the NCS hypotheses. These examples are meant to demonstrate some of the tradeoffs associated with different design approaches, and to provide a starting point for evaluating different designs in terms of their ability to assess a set of selected hypotheses (as well as other hypotheses identified during the course of the study, associated with outcomes or exposures with characteristics similar to these selected hypotheses).

When determining differences in study designs relative to their ability to address the hypotheses, it is helpful to conduct the assessment in terms of the probability of a given design to minimize both Type 1 statistical error (the probability of concluding there is a relationship when in fact there is not) and Type 2 error (the probability of concluding there is no relationship when in fact there is). From a statistical standpoint, Type 1 errors are controlled by the choice of significance level (or alpha level) for statistical tests conducted as part of the data analysis (i.e., the chance of a Type 1 error is typically fixed at some small value, such as α=0.05). Related to Type 1 errors are errors resulting from concluding there is an effect when the effect may be due to confounding, or when measurement of the effect has been biased due to model misspecification, biases in measurement, or biases (both known and unknown) in the sample. On the other hand, Type 2 errors are usually characterized by examining the power of the study to detect a specified effect level (e.g., an odds ratio of 1.5) for a key study hypothesis. Power assesses the probability of the complement of a Type 2 error, or the probability of correctly concluding that there is an effect/relationship when an effect/relationship of specified size is present.

For a study like the NCS, with multiple hypotheses and multiple inferences of interest, there are many ways to assess power, and the results can be quite dissimilar. For example, power is dependent upon the statistical model chosen to characterize the relationship (e.g., regression models, logistic regression models, survival analysis models, longitudinal models where multiple observations for each individual are available, etc.). In particular, power may be greatly increased when the data and underlying biological model support estimation of a continuous dose-response relationship between an exposure and outcome (as opposed to estimation of a categorical effect such as an odds ratio or relative risk). Additionally, power may be heavily dependent on the inference goals associated with an analysis (i.e., whether inferences are being drawn for only the population of individuals participating in the study or for a broader population that could expand up to the entire sampling frame from which the study population was drawn). Conceptually, this can be thought of as doing separate power analyses associated with a weighted or unweighted analysis of the data (i.e., using or not using sampling weights, if they are calculable). Thus, power should be assessed for both inferences that apply only to the

study population, as well as inferences that can be generalized to the wider sampling frame population. Inferences beyond the specified sampling frame (e.g., children born during the recruitment period of the NCS but not specifically enumerated by the selected sampling frame) of the study cannot be based on statistical analysis alone.

Since evaluating the relationships of interest within the study population (i.e., assessing internal validity) is the logical first step to analyzing the data obtained in the NCS, power for detecting these relationships within the cohort, through an unweighted analysis, is the initial focus in the results presented here. We assume relationships observed within the NCS cohort using an unweighted analysis will be used to assess internal validity. In some cases, combining other scientific assessments with these internal NCS results may offer the possibility of generalizing the results to other (larger) populations. However, the ability to estimate the statistical significance of a hypothesized relationship for a broader population within the sampling frame is also extremely valuable – providing a basis for inferring significant relationships based on the statistical analysis alone. Thus, we also discuss the power results for a weighted analysis of the data, which allows inference to the wider sampling frame population.

In addition to the statistical model and the inference goals associated with a particular analysis, power will depend on a variety of other factors, such as:

- The Type 1 error rate (or significance level) for determining that there is a significant relationship between the exposure of interest, *X*, and the adverse health outcome, *Y*.
- The strength (or severity) of the relationship between *X* and *Y*.
- The degree of unequal weighting and clustering in a design. By clustering, we refer to the possible correlation of responses (and exposures) among individuals selected in the same cluster. Note that there could be clustering in both *X* and *Y*, and both types of clustering can have an impact on power. One of the factors that affects clustering is the number of PSUs (or number of clusters) in a sample. In the examples in this section, counties are generally considered the PSUs, with individuals living in the same county assumed to be more similar than individuals living in different counties. It should also be noted that the number of PSUs may potentially have significant impact on the cost of a given design (see Chapter 8).
- The incidence or prevalence for binary health outcomes, *Y*. Depending on the hypothesis, the prevalence (or rate of occurrence) of the disease may vary considerably. For example, autism is known to have a very low rate of occurrence (e.g., on the order of 2 or 3 cases per 1000 children or 0.2% of children) in the population whereas the rate of occurrence of asthma is on the order of 100 cases per 1000 children (or 10% of the population). For continuous outcomes, the distribution of the outcome will be important.
- The distribution of *X* (e.g., the frequency of the exposure for a categorical exposure measure). Similar to the incidence or prevalence of *Y*, the distribution of *X* will vary depending on the hypothesis of interest, which will also affect the power to detect relationships between *X* and *Y*.
- Availability of other covariates that assist in explaining *Y* and/or that modify the effect of *X* on *Y*. This could be highly related to implementation costs for the

study, in that a study design which has higher cost in other areas, such as recruitment and retention of study participants, may not have the ability to collect as much explanatory/covariate information.

- The size of the sample, which is affected by several factors:
  - Cost
  - The period of assessment (i.e., relevant life stages for testing the hypothesis) for the selected hypothesis and the attrition rate associated with that length of time. For example, if 100,000 children are originally enrolled, and the hypothesis of interest cannot be assessed until age 6 or 7, then presumably some percentage of the original 100,000 cohort will have dropped out of the study, thereby reducing the number of subjects available for analysis.
  - The availability of the necessary data. For example, missed measurements and/or designed missingness (e.g., sub-sampling for more detailed exposure measurements) for factors that are either difficult or expensive to collect will impact the data that is available for analysis (and the amount of information contained in that data).

The large number of factors affecting the calculation of power, each of them different for different hypotheses, makes power studies for the NCS as a whole relatively difficult. For this reason, in the examples below we generally focus on specific scenarios (and specific settings for all of the above factors) to demonstrate the effect that these factors have on the power of a selected design. For each scenario considered, we highlight the assumptions (e.g., retention assumptions, distributional assumptions, etc.) that were made so that the results can be interpreted in light of these assumptions. It should be noted that some of these assumptions, while seemingly reasonable, are based on limited information and could be better refined if pilot studies were used to evaluate their validity.

To narrow the focus to a manageable number of scenarios, we begin by limiting this investigation to the evaluation of power for diseases and exposures that are binary (i.e., categorical X and Y variables). Since many of the hypotheses of the study involve evaluation of the relationship between a categorical or binary risk factor (e.g., presence or absence of the risk factor) and a binary health outcome variable (e.g., presence or absence of the health outcome), this simplification seems a reasonable starting point for evaluating power and indicating some of the main differences in various design approaches. Additionally, we envision that similar results will hold for related, more complex analyses. Thus, we evaluate the power of a design to detect relationships between a categorical (two-level) outcome Y and a categorical risk factor X. In the following sections, power is computed for both an un-weighted analysis (i.e., an analysis that provides inference for the NCS cohort), and a weighted analysis that takes into account the complex design of the study and is appropriate for inferences to the sampling frame population. (Note that the unweighted analysis corresponds to a model-based approach to inference.)

Another reduction strategy that we employ is to begin by ignoring the longitudinal aspect of the data, and focus on outcomes, and exposures, that are simply assessed for their presence/absence by a specified age (e.g., did the child have asthma by age 10). In other words, the data will consist of a single observation for each individual in the available cohort, and we do

not consider the case of multiple observations for an individual. Certainly, we recognize that the NCS will involve measures that are assessed repeatedly over time and that are correlated within an individual and between individuals within the same cluster. However, given the current hypothesis statements and considering that longitudinal analysis may provide increased power for detecting relationships of interest (e.g., since additional information would be available if multiple observations are obtained for each individual), we begin by ignoring the longitudinal aspect of the NCS design and consider power calculations for the setting where subjects are clustered within groups defined by the clustered design. This strategy is a sensible one in terms of providing a conservative estimate of power; however it also implies that the true exposure can be characterized by a single measure (which may be very naïve).

Finally, it should be noted that the power studies that are presented in this report are based on sample sizes that include all study participants that are available at the time that the health outcome is observed. However, many of the study hypotheses will likely be addressed by more efficient outcome dependent sub-studies of the NCS cohort, especially those hypotheses that address rare health outcomes such as birth defects, autism and schizophrenia. It is easy to show that a well designed nested case control study from within the NCS study population that includes all observed cases of disease and a reasonable number of control children (e.g., three times as many controls as cases) will have largely the same power as a study that includes the entire cohort. This result holds both for unweighted analyses of the data, as well as analyses that incorporate the sampling weights for extrapolation to the sampling frame population. Although this result may help study planners build additional cost efficiencies into the study design when feasible (e.g., rely upon archived biological and environmental samples when possible, and only chemically analyze the substantially smaller subset of those samples from study participants that are selected for the appropriate nested case control study), it is critical that the study maintain high rates of retention so that the maximum number of rare cases can be observed – as the nested case control study derives its power from the number of cases observed (provided there are suitable controls available for each case).

The remainder of this Chapter is organized in the following manner. We begin by summarizing the designs that will be considered, the methods for implementing these designs, and the methods used in combining individuals sampled using the different sampling frames in the design. Section 9.1 provides this description and highlights some of the issues involved in selecting a national probability-based design and in combining individuals selected from different sampling frames. Section 9.2 describes the methods that will be used in calculating the power associated with a design beginning with the assumption of a simple random sample (i.e., an unweighted analysis that assumes no clustering), for which power can be calculated analytically, and moving to designs that involve clustering and unequal weighting of the observations (i.e., a weighted analysis that accounts for the clustering in a given design), which entails calculation of power via simulation and the use of generalized estimating equations (GEE) (see Liang and Zeger, 1986) analysis methods. Section 9.3 briefly describes the hypotheses that are investigated in the power analyses, and Section 9.4 displays and discusses the results of the power analyses for both a simple random sample and a design involving both clustering and unequal weighting. Finally, Section 9.5 discusses the implications and possible interpretations of these results, their limitations, and areas for future research that could lead to more refined pictures of the power of various NCS design scenarios under different settings.

Note that those readers interested in general results and conclusions should simple read the summaries provided in Section 9.5 and Chapter 10, rather than the detailed results of Section 9.4.

9.1 DESIGNS CONSIDERED

Since the degree of unequal weighting and clustering in a design can have an effect on power, it will be necessary to specify a limited number of designs that will be considered in these power calculations. Chapter 3 of this report outlines the family of designs that are proposed as candidate NCS designs, and Chapter 5 outlines a set of 23 designs that are considered in these power analyses. The specific designs selected from the family of designs are meant to provide a range of possible designs so that an indication of the effect of changing the various design parameters can be obtained. In other words, the 23 designs are selected in an attempt to span the range of possible designs outlined in Chapter 3. Table 9-1 displays these 23 designs based on the number of PSUs in the NPBS (50 or 100), the percent of the cohort selected in the NPBS (P_1), the percent of the cohort selected in the Centers MSA ($(1-P_1)*P_2$), and the percent of the cohort selected from the Center patient lists. (Note that designs A1 and A2 and designs B3 through B5 correspond to smaller sampling frame populations then the sampling frame population corresponding to designs C6 through H23. This sampling frame difference should be considered when comparing these designs in terms of their statistical power.)

Recall that the family of designs outlined in Chapter 3 allows a portion of the cohort to be selected as a volunteer sample; however, for most of the designs (all except design A2) displayed in Table 9-1 we concentrate more directly on the probability-based sampling aspects of the family of designs by assuming that the entire cohort will be selected in some probabilistic manner. As described in Chapter 5, this is not to say that we consider volunteerism an unimportant aspect for the NCS, since there are a number of possible advantages in dealing with volunteer subjects (e.g., higher retention rates, increased motivation to fully participate in the study, etc.). Thus, it may be the case that NCS planners choose to include some portion of the cohort as a volunteer sample. In this case, power for an unweighted analysis will likely be very similar to the unweighted analysis powers presented for these designs (actually, the power would likely be greater since the increased retention rates assumed for volunteer subjects would result in increased numbers of individuals available for analysis). For a weighted analysis, on the other hand, the powers would presumably be adversely affected as the volunteer portion of the cohort is increased. To demonstrate this impact, one of the designs in Table 9-1 calls for selection of a significant portion of the cohort as a volunteer sample. Specifically, design A2 specifies selection of 50 percent of the cohort as volunteers. Due to the potential biases resulting from a volunteer sample and in an effort to provide a conservative estimate of the weighted analysis power for this design, we do not include the information collected from volunteer subjects when conducting the weighted analyses associated with this design (i.e., volunteer subjects are not included in a weighted analysis).

In addition to outlining the 23 design specifications, Table 9-1 also provides the number of clusters associated with each design. For designs A1 and A2, the number of clusters in the design is defined as two times the number of Centers. We include two clusters per center since for these designs we nominally assume that each Center is required to sample 20 percent of their subjects from a rural area, denoting one cluster of subjects from the Center, and the remaining 80

percent from an urban area, denoting the other cluster of subjects from the Center. Under the assumption that each Center can recruit and follow 2000 participants (i.e., 50 Centers are necessary to recruit and follow 100,000 participants), this results in a total of 100 clusters for designs A1 and A2. For designs B3 through H23, the number of clusters in the design will also depend on the number of Centers and the number of PSUs. On the NPBS side of sampling, counties are considered the clusters, and on the Centers side of sampling the MSA and the patient lists are each considered as separate clusters (i.e., two clusters per Center or a clustered sample with stratification within Centers). In other words, for designs B3 through H23 the number of clusters in a design is the number of PSUs in the NPBS plus two times the number of Centers. [It should be noted that these designs could also be analyzed (and simulated) as nested cluster designs with the MSA as the first level of clustering and the different modes of recruitment of individuals as a second level of clustering. Due to the increased complexity of this nested correlation structure, a single level correlation structure is utilized for purposes of these power analyses.]

Finally, for each of the 23 designs we also consider two approaches to cost (see Chapter 8). The first approach is the fixed sample size approach in which each design recruits and follows 100,000 children, regardless of cost. The second approach is the fixed cost approach where a total fixed cost of approximately 2.7 billion dollars is assumed, and each design recruits and follows as many individuals as possible given the cost constraints. (Chapter 8 provides a more detailed description of the methods for evaluating costs and identifying appropriate sample sizes for the fixed cost designs.) The last two columns of Table 9-1 display the initial sample sizes for both the fixed sample size designs (all designs have initial sample size of 100,000), and for the fixed cost designs.

Table 9-1. Set of 23 designs considered in the power analysis.

Design	Number of PSUs in NPBS	% Cohort in NPBS	% Cohort in Centers Area	% Cohort in Center-Patients	# of Clusters for Fixed Sample Size	N for Fixed Sample Size Designs	N for Fixed Cost Designs
A1*	0	0	0	1.00	100	100000	76750
A2*	0	0	0	0.50	100	100000	75750
B3+	0	0	0.25	0.75	100	100000	80000
B4+	0	0	0.50	0.50	100	100000	84000
B5+	0	0	0.75	0.25	100	100000	88000
C6	50	0.25	0.19	0.56	126	100000	77000
C7	50	0.25	0.38	0.38	126	100000	80000
C8	50	0.25	0.56	0.19	126	100000	83500
D9	50	0.5	0.13	0.38	100	100000	88500
D10	50	0.5	0.25	0.25	100	100000	91000
D11	50	0.5	0.38	0.13	100	100000	94000
E12	50	0.75	0.06	0.19	76	100000	104000
E13	50	0.75	0.13	0.13	76	100000	106000
E14	50	0.75	0.19	0.06	76	100000	107500
F15	100	0.25	0.19	0.56	176	100000	63000
F16	100	0.25	0.38	0.38	176	100000	66000
F17	100	0.25	0.56	0.19	176	100000	69000
G18	100	0.5	0.13	0.38	150	100000	73000
G19	100	0.5	0.25	0.25	150	100000	75000
G20	100	0.5	0.38	0.13	150	100000	77500
H21	100	0.75	0.06	0.19	126	100000	85500
H22	100	0.75	0.13	0.13	126	100000	87000
H23	100	0.75	0.19	0.06	126	100000	89000

* Note that the sampling frame population for these designs is limited to patients of the set of purposively selected Centers
+ Note that the sampling frame population for these designs is limited to individuals living in the geographic area and/or individuals that are patients of the set of purposively selected Centers

9.2 METHODS

In many statistical analyses involving a binary outcome, logistic regression models (Hosmer and Lemeshow, 2000) are used to assess the statistical significance of the relationship between the outcome and any risk factors of interest. In the simple univariate case (i.e., a single risk factor and a single outcome) the model is as follows:

$$ln(p/1\text{-}p) = \beta_0 + \beta_1 \cdot X,$$

where p is the conditional probability of having the outcome of interest (i.e., the probability of disease) given X, the risk factor of interest (e.g., for a categorical risk factor X=1 if the risk factor is present and X=0 otherwise), and β_0 and β_1 are the parameters of the model. The significance

of the relationship between X and Y is assessed by evaluating the significance of the parameter β_1, which is the natural log of the odds ratio. The odds ratio is a popular measure of association for binary outcomes that is defined as the ratio of the odds of the outcome for a unit increase in X. More particularly, for a binary risk factor indicating presence or absence of exposure, it is the ratio of the odds of disease for exposed individuals to the odds of disease for unexposed individuals. For example, if the odds ratio is 2.0, then the odds of disease for exposed individuals is twice as high as the odds of disease for unexposed individuals. Odds ratios that are greater than 1 imply that increases in X are related to increases in the odds (or probability) of disease, and odds ratios less than 1 imply that decreases in X are related to increases in the odds (or probability) of disease. For this investigation, we are generally interested in detecting odds ratios that are greater than 1, which indicate a positive relationship between the risk factor of interest and the probability of disease. This would suggest that the power for one-sided hypothesis tests is appropriate; however, to provide more conservative power estimates that may be relevant to other hypotheses for which two-sided tests are necessary (e.g., for situations where no prior assumption on the direction of the relationship between the outcome and the risk factor is made), all power calculations are done assuming a two-sided test will be utilized.

In more formulaic terms, letting p_1 be the probability of disease for individuals with X=1, and p_0 be the probability of disease for individuals with X=0, the formula for the odds ratio (OR) is as follows:

$$OR = \frac{p_1/(1-p_1)}{p_0/(1-p_0)}.$$

Note that for the case of $(1-p_0)/(1-p_1) \approx 1$, the odds ratio is approximately equal to the relative risk, p_1/p_0. Table 9-2 and Figure 9-1 provide examples of the relationship between p_0 and p_1 as the odds ratio varies. Note that an odds ratio of 1.00 indicates no relationship between disease and exposure (i.e., p_0 and p_1 are identical), and as the odds ratio increases the difference between p_0 and p_1 increases.

Table 9-2. Probability of disease for exposed individuals (p_1) as a function of the odds ratio and the probability of disease for unexposed individuals (p_0).

Prob. of Disease for Unexposed Individuals (p_0)	Prob. of disease for exposed individuals (p_1)								
	Odds Ratios								
	1.00	1.01	1.05	1.10	1.20	1.35	1.50	1.75	2.00
0.0025	0.0025	0.0025	0.0026	0.0027	0.0030	0.0034	0.0037	0.0044	0.0050
0.005	0.0050	0.0050	0.0052	0.0055	0.0060	0.0067	0.0075	0.0087	0.0100
0.01	0.0100	0.0101	0.0105	0.0110	0.0120	0.0135	0.0149	0.0174	0.0198
0.05	0.0500	0.0505	0.0524	0.0547	0.0594	0.0663	0.0732	0.0843	0.0952
0.1	0.1000	0.1009	0.1045	0.1089	0.1176	0.1304	0.1429	0.1628	0.1818
0.2	0.2000	0.2016	0.2079	0.2157	0.2308	0.2523	0.2727	0.3043	0.3333

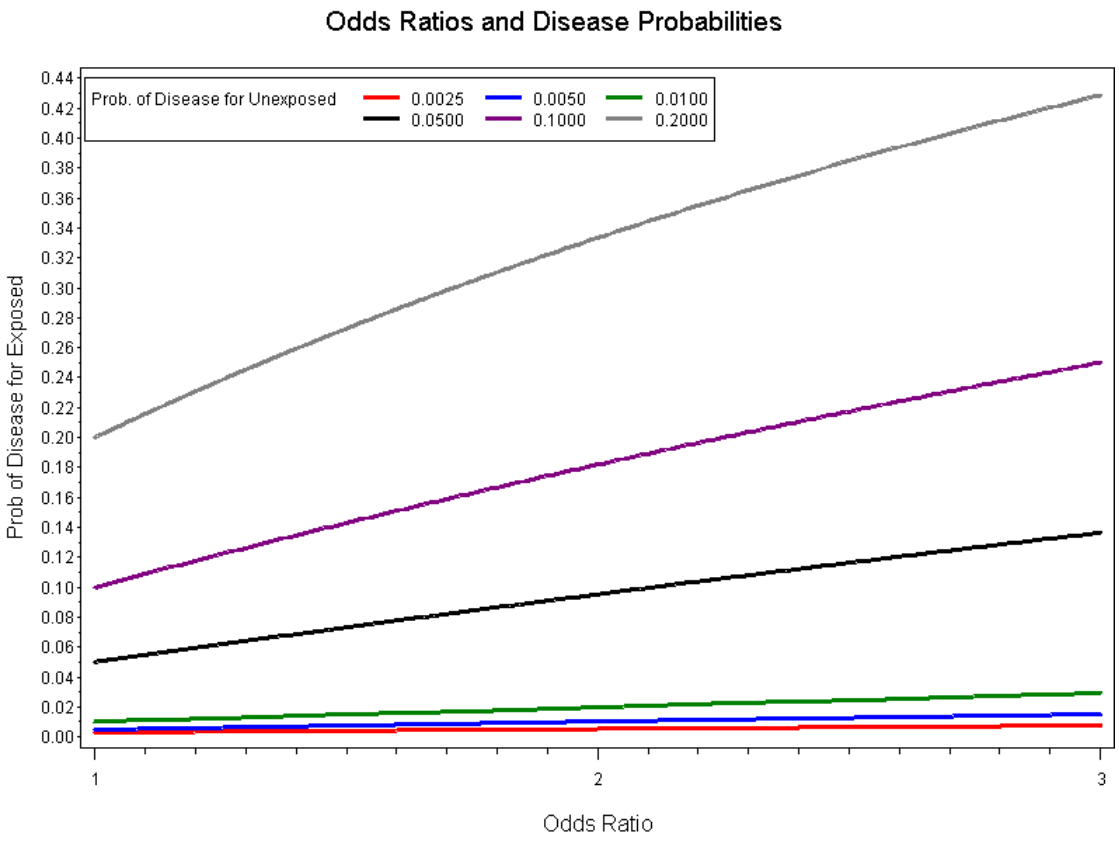


Figure 9-1. Probability of disease for exposed individuals (p_1) as a function of the odds ratio and the probability of disease for unexposed individuals (p_0).

Of course, there are other measures of association between a binary exposure factor X and a binary health outcome Y (e.g., the relative risk), and there are other plausible tests (other than using a logistic regression model) for evaluating the significance of the relationship between X and Y. For example, comparing the proportion of individuals with the disease in the unexposed group to that of the exposed group is likely the simplest means of evaluating the significance of this relationship. In general, the logistic regression paradigm allows a high degree of flexibility and generalization that will certainly be important when analyzing data obtained in the NCS. For example, logistic regression models allow incorporation of further levels of complexity such as including other important covariates (e.g., continuous exposure measures or scores) and possible confounders in the model, and/or including correlated data structures, through the use of GEE models (Liang and Zeger, 1986). Thus, since it is easily extended to more complex cases, logistic regression seems a natural starting point for evaluating power in the simple univariate case (i.e., through testing the significance of the log-odds ratio, β_1). We refer the reader to Hosmer and Lemeshow (2000) for a more detailed description of the logistic regression paradigm, and to Liang and Zeger (1986) or Zeger and Liang (1986) for a description of GEE methods.

In the following subsections we describe methods for calculating the power to detect relationships between a binary exposure factor X and a binary health outcome Y. As suggested previously (see above), these calculations will depend on a large number of factors (e.g., strength

of relationship, degree of clustering, prevalence of *X* and *Y*, etc.). Thus, we begin with the simple case of assuming the data are a simple random sample (i.e., all individuals are selected with equal probability and there is no clustering in the design). For this situation, analytical formulas for statistical power exist, making consideration of a large number of scenarios possible and offering an important starting point for these power analyses. However, this simple approach does not account for unequal probabilities of selection (i.e., unequal weights in a weighted analysis) and it does not account for the clustering that would likely be apparent in any plausible NCS design. In other words, the simple random sample results need to be interpreted as a sort of "best-case" scenario for the NCS. The second approach that we describe addresses both the unequal weighting and the clustering issues by conducting a series of simulations for a small set of selected scenarios. The simulation involves generating candidate designs (according to the family of designs described in Chapters 3 and/or 5), simulating the data realized under that design, analyzing the simulated data, and repeating this process a large number of times to obtain an estimate of the power associated with each scenario. Since this simulation approach is computationally intensive, a smaller number of scenarios, arising from a subset of the hypotheses of the study, are evaluated, but the results can be considered more realistic and offer further refinement and interpretation of the differences associated with candidate designs.

9.2.1 Power for a Simple Random Sample

For simple inferences that treat the cohort as a simple random sample, analytical formulas can be derived to compute the power of detecting a specified effect. Assuming that the estimate of the log-odds ratio is approximately normally distributed (asymptotically the maximum likelihood estimate of the log-odds ratio is normally distributed), an analytical formula for the two-sided power to detect an odds ratio of a specified size is given by:

$$\text{Power} = 1 - \Phi\left(z_{1-\alpha/2} - \frac{\ln(OR)}{\sqrt{V\left(\ln(\hat{OR})\right)}} \right) + \Phi\left(-z_{1-\alpha/2} - \frac{\ln(OR)}{\sqrt{V\left(\ln(\hat{OR})\right)}} \right), \qquad (9\text{-}1)$$

where Φ is the CDF of a standard normal distribution, $z_{1-\alpha/2}$ is the upper $\alpha/2$ percentile of a normal distribution, OR is the assumed value of the true odds ratio, and $V\left(\ln\left(\hat{OR}\right)\right)$ is the variance of the estimate of the log-odds ratio. (Note that this variance will depend on the true odds ratio, the sample size, the prevalence of the disease, and the prevalence of the risk factor.) Section D-6 of Appendix D provides further details of the derivation of this formula, and Figure 9-2 displays an example of the results that can be obtained using an analytical formula. The figure displays the power of a two-sided α=0.05 level test to detect odds ratios of a specified size for a simple random sample with a disease prevalence of 5% (i.e., on average 5 out of 100 individuals get the disease), an exposure prevalence of 20%, and sample sizes of n=5000 (right-most curve), 10000, 20000, 35000, 50000, 75000, and 100000 (left-most curve).

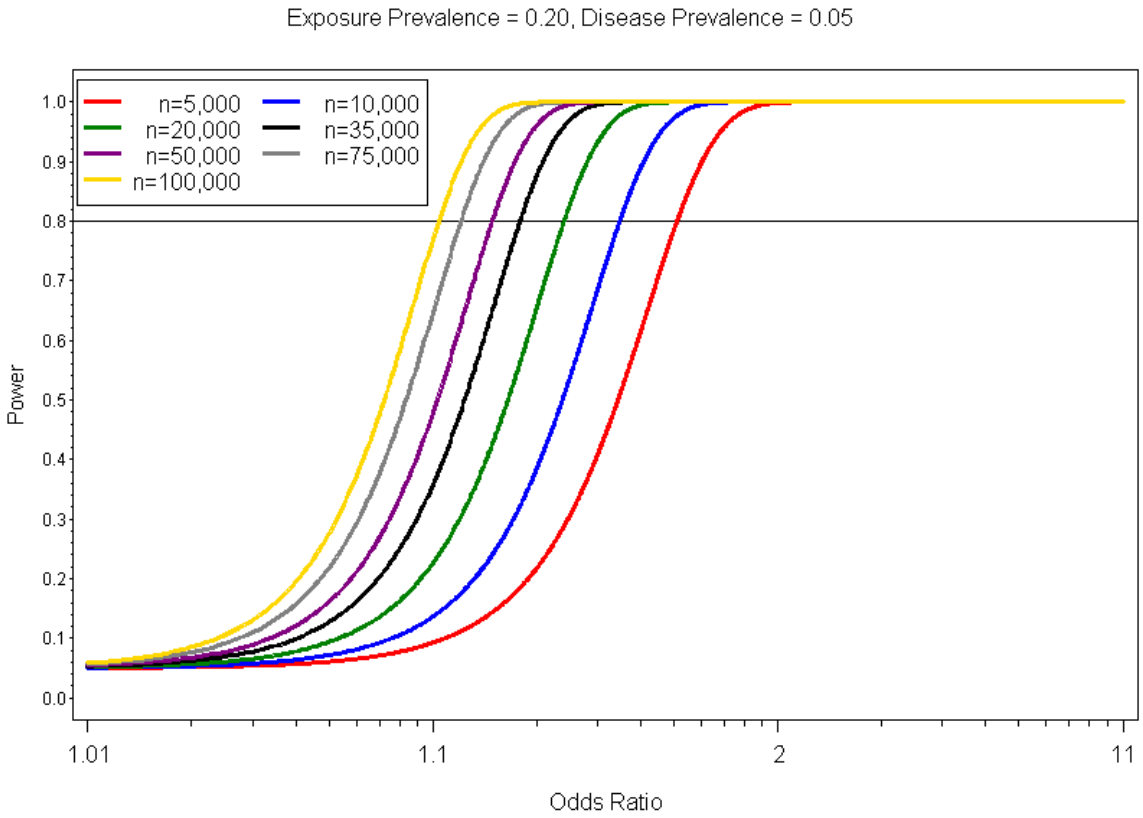


Figure 9-2. SRS power for detecting a significant relationship between a health outcome with a prevalence of 5% and a binary risk factor with a prevalence of 20%.

Thus, due to its ease of computation, the above analytical formula provides a means of assessing power for a large number of scenarios, and can be used to evaluate the effect of sample size, disease and exposure occurrence rates, and the strength of the relationship between disease and exposure. This may be especially useful in assessing the sensitivity of the power calculations to the assumptions that are used in conducting them. For example, the effect of assuming different occurrence rates for the risk factor and the effect of assuming different sample sizes, resulting from assuming different retention rates associated with the length of follow-up for the hypothesis of interest or different costing scenarios, can be evaluated. The hypothesis-specific results presented in Section 9.4 will highlight the usefulness of these kinds of investigations (e.g., in evaluating different hypotheses and in assessing the sensitivity of the calculations to the assumptions of the scenario).

It is important to note, however, that power values resulting from these calculations, and any conclusions based on the results, should be interpreted in light of the *simple random sample* assumption. More particularly, we note that the formulas do not account for the effect of clustering and unequal weighting that will likely be elements of any feasible NCS design. (e.g., recall that the design effects discussed in Chapter 5 are affected by clustering and unequal weighting). Considered in another light, these calculations could approximately correspond to an unweighted analysis (i.e., an analysis where all observations have equal weight) in which there is no clustering (or no effect of clustering) of the design. Thus, while they offer an important

starting point for evaluating the power associated with different hypotheses and different designs, they must be interpreted with these limitations in mind.

9.2.2 Calculating Power via Simulation

To calculate power under a design that involves both unequal weighting and clustering, a simulation approach to calculating power is utilized, in which the following process is repeated:

1. Obtain a realization of the proposed design (i.e., sample 100,000 individuals and compute their probability of selection according to the specified design scheme).
2. Simulate the binary exposure and binary disease variables according to the specified scenario. This will depend on the prevalence of the exposure, the prevalence of the disease, the amount of within-cluster correlation in the *X*s and the *Y*s, the assumed odds ratio, and the assumed retention rates for the design at the life stage when *Y* is observed.
3. Fit a logistic regression GEE model that accounts for the possible clustering of the observations and assumes equal weights for all the observations (i.e., conduct an unweighted analysis). Each fit will provide an estimate of the log-odds ratio, and its corresponding standard error and statistical significance, for an unweighted analysis under the selected design.
4. Fit a logistic regression GEE model that accounts for both clustering and unequal weighting of the observations (i.e., conduct a weighted analysis). Again, each fit will provide an estimate of the log-odds ratio, and its corresponding standard error and statistical significance, for a weighted analysis under the selected design.

Of course, it should be noted that each of the above steps involves a number of complexities that are explained in other sections or appendices of this report. In particular, Section 9.1 provides a brief description of the set of proposed designs that we investigate here, and the process for obtaining a realization of the proposed designs. More detailed descriptions of the class of designs (or family of designs) that are considered can be found in Chapters 3 and 5 of this report, and a detailed description of the methods for obtaining the designs can be found in Chapter 5. The methods for simulating binary exposure and disease variables are briefly described below, and are more fully specified in Section D-7 of Appendix D. Finally, the GEE models (using an independence working correlation (Heagerty and Zeger, 2000)), and the corresponding estimates of the log-odds ratio, its standard error, and statistical significance, were implemented using the SUDAAN software package.

There are generally two approaches to simulating correlated binary data, the random effects logistic approach and the beta-binomial approach (see Section D-7 of Appendix D). For these power calculations we simulated correlated binary data using a beta-binomial approach in which cluster-specific baseline occurrence probabilities are selected from a beta distribution with moments defined by the desired within-cluster correlation and the assumed marginal probabilities (or overall prevalence rates), and within-cluster subject specific binary outcomes are simulated from a bernoulli trial using the cluster-specific occurrence probabilities. The assumed marginal odds ratio between unexposed and exposed subjects, along with the prevalence of the exposure and the disease, will determine the cluster-specific constant odds

ratios (or conditional odds ratios). Section D-7 of Appendix D provides a more detailed discussion of the possible approaches to simulating correlated binary data, and a more detailed description of the beta-binomial method utilized here. Additionally, Section 9.5 provides a discussion of the possible implications and interpretations when assuming a constant odds ratio across clusters.

In the ideal situation, power would be calculated by repeating the above process a large number of times (e.g., 1000 times), and calculating the proportion of those times that the estimated log-odds ratio is statistically significant. Due to computational limitations and the large number of scenarios under consideration for this work, a more computationally efficient, although perhaps less accurate, approach was used in which the above process was repeated 50 times and the median of the 50 log-odds ratio standard errors was used as an estimate of the standard error that would be realized under the selected design. Like Equation (9-1), the power is calculated as:

$$\text{Power} = 1 - \Phi\left(z_{1-\alpha/2} - \frac{\ln(OR)}{SE(\ln(\hat{OR}))} \right) + \Phi\left(-z_{1-\alpha/2} - \frac{\ln(OR)}{SE(\ln(\hat{OR}))} \right), \qquad (9\text{-}2)$$

where Φ is the CDF of a standard normal distribution, $z_{1-\alpha/2}$ is the upper $\alpha/2$ percentile of a normal distribution, OR is the assumed value of the true odds ratio, and $SE\left(\ln\left(\hat{OR}\right)\right)$ *SE(ln(OR))* is the estimate of the standard error of the log-odds ratio estimate under the selected design (i.e., the median of the estimated standard errors in the 50 simulated data sets).

Thus, using the above procedure we will evaluate power for both an unweighted (model-based) analysis that accounts for clustering in a design and is appropriate for testing the significance of estimates for the NCS cohort, and we will evaluate power for a weighted analysis that accounts for both clustering and unequal weighting in a design and is appropriate for generalization of results to the sampling frame population (recall that the volunteer subjects of design A2 are not included in the weighted analysis). Section 9.2.3 provides a discussion of how these results will be presented and discussed for each of the scenarios considered in Section 9.4.

9.2.3 Displaying the Results of the Power Calculations

Figure 9-3 displays an example of the types of results that are available for each of the hypotheses considered (see Section 9.3). The figure displays the power of a model-based (unweighted) analysis to detect the relationship of interest as a function of the odds ratio and the design for a health outcome with an occurrence rate of 0.6% individuals and an exposure risk factor with an occurrence rate of 5%. The different panels of the plot correspond to different proportions of the cohort selected from the NPBS with 100 PSUs (see Chapter 5 and Table 9-1), and the different lines in each panel correspond to different proportions of the Centers sample selected from the Center's MSA (i.e., P_2). For example, the red line in the upper left panel of the figure corresponds to a design for which 25% of the cohort is selected in the NPBS with 100

PSUs, and 25% of the within-Centers (i.e., 25% of the 75% selected from a set of purposively selected Centers) cohort is selected from the MSA corresponding to the Center. Similar figures are available for weighted analyses and for the other designs considered.

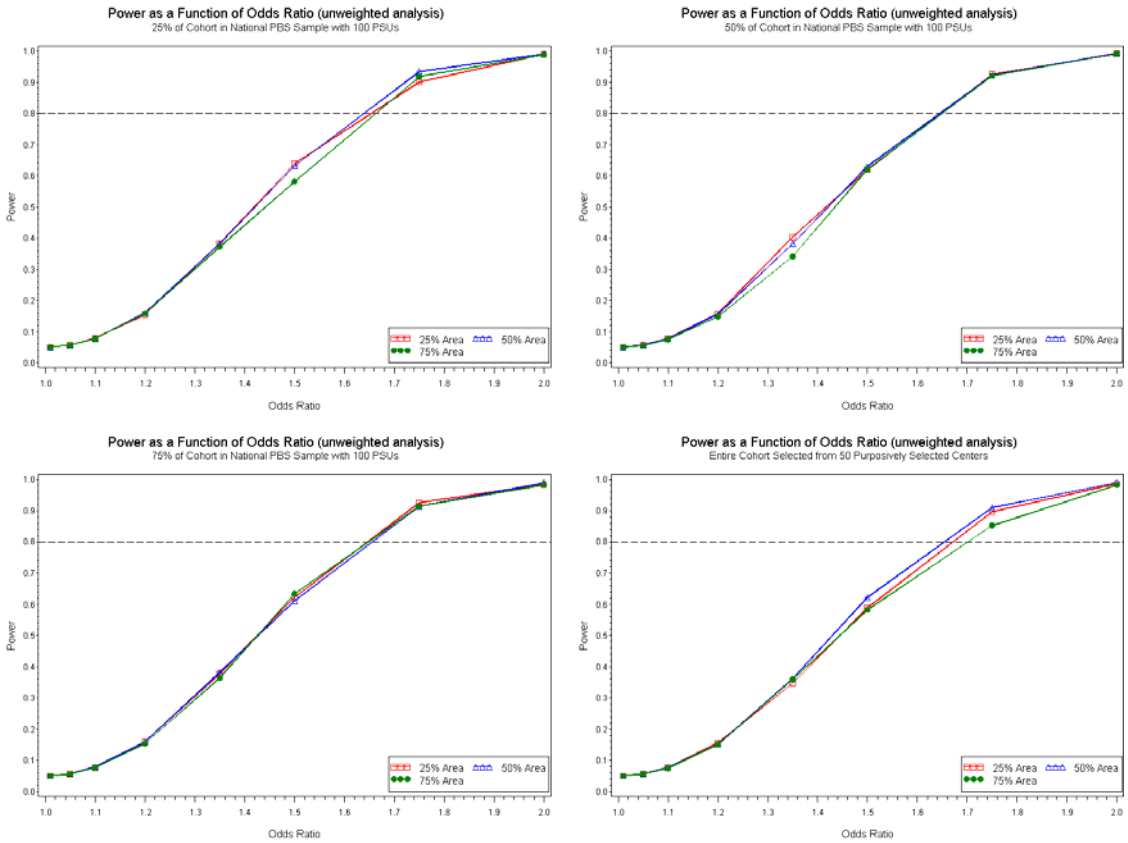


Figure 9-3. Example of the unweighted analysis power for detecting a significant relationship between a health outcome with a prevalence of 0.6% and a binary risk factor with a prevalence of 5%.

In looking across all the results for both unweighted and weighted analyses, there are generally small differences in power between the 100 and 50 PSU designs with the same cohort selection proportions. Additionally, there are relatively small differences between the designs that select different proportions of the cohort in the Center-area sample assuming a fixed proportion of the cohort is selected in the NPBS (i.e., the different lines in each panel of Figure 9-3), indicating that perhaps the more important differences are related to the percent of the cohort in the NPBS. For this reason, Figure 9-4 focuses on the power results for designs A1, A2, B4, F16, G19, and H22 (see Table 9-1). The different panels of the figure correspond to unweighted (left-hand panels) and weighted (right-hand panels) analyses for the fixed sample size (top panels) and fixed cost designs (bottom panels). The different lines in each panel correspond to the six different designs (A1, A2, B4, F16, G19, and H22). This figure may provide a more succinct summary of the differences between the designs for both unweighted and weighted analyses, and for the fixed sample size and fixed cost scenarios. Thus, to present

the results of the power calculations, figures such as the one displayed in Figure 9-4 will be provided along with a discussion of the relevance of the results.

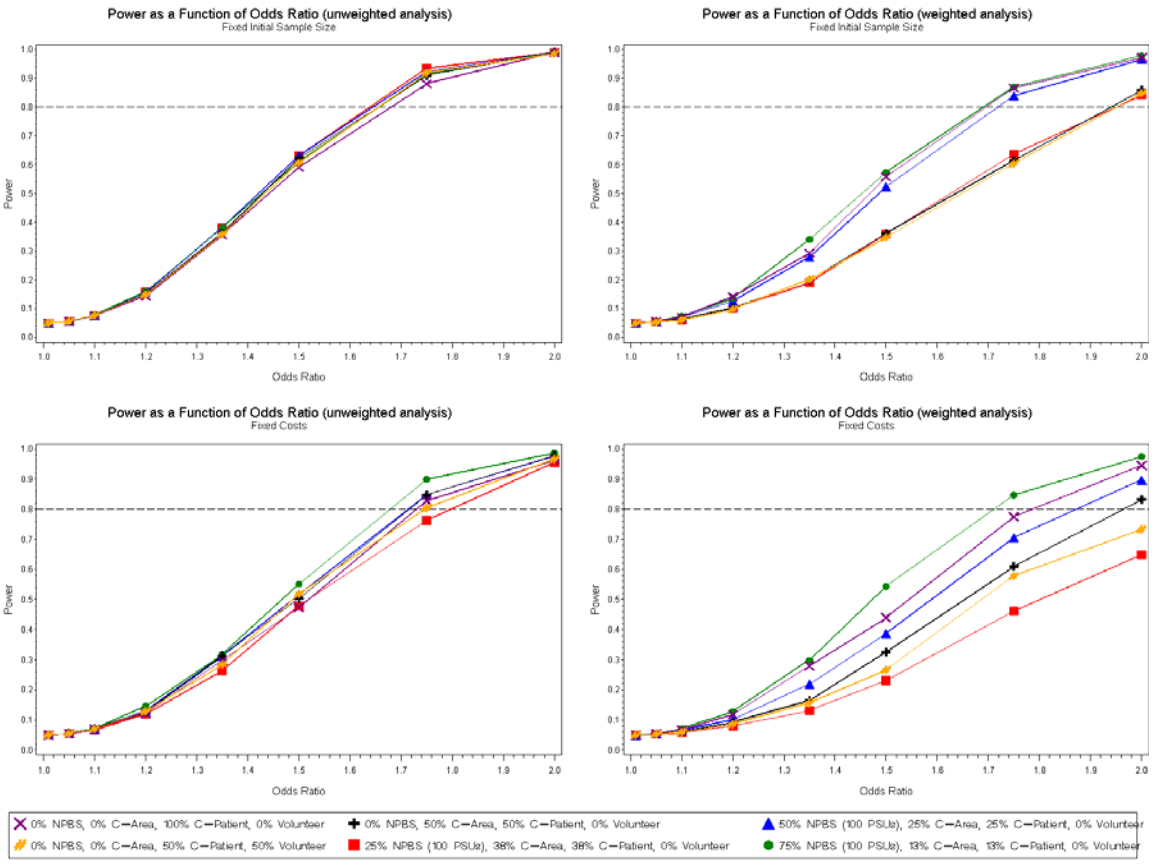


Figure 9-4. Example figure displaying the power to detect a significant odds ratio for unweighted and weighted analyses, and for fixed sample size and fixed cost samples corresponding to designs A1, A2, B4, F16, G19, and H22 (assuming a health outcome with a prevalence of 0.6% and a binary risk factor with a prevalence of 5%).

In addition to figures like Figure 9-4, tables of the power to detect a specific odds ratio (determined by evaluating the figures and determining what odds ratio provides around 80% power for a weighted analysis) as a function of the design, the selected scenario, and whether a fixed sample size or fixed cost is assumed, are also provided to assist in interpreting the results. As an example, Table 9-3 displays the power to detect an odds ratio of 1.75 for a fixed initial sample size of 100,000 individuals and as a function of the design parameters (i.e., the percent of the cohort selected from the different sampling frames), the type of analysis (unweighted, weighted, or SRS), and the number of PSUs for the NPBS selected individuals. Note that the sample size available for analysis displayed in these tables will depend on the initial sample size, the hypothesis of interest through its required years of follow-up, and the retention rates for the desired follow-up period and the method of selection. Since the retention rates differ for the different sampling frames, with individuals selected from Center patient lists having generally

the highest retention rates and individuals selected in the NPBS having the lowest retention rates, the proportion of the individuals available for analysis from the different sampling frames will change over time. Thus, the first three columns display both the percent of the original cohort selected in each frame, and the percent of the available or remaining cohort selected in each frame. Finally, for each column, a star denotes the maximum power for the designs that include selection of some portion of the cohort as a NPBS.

Table 9-3. Example table for the power to detect an odds ratio of 1.75 as a function of the design (assuming a disease prevalence of 0.6%, an exposure prevalence of 5%, and a fixed initial sample size).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]	Power Unweighted		Power Weighted		Power SRS
				50 PSUs	100 PSUs	50 PSUs	100 PSUs	
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	99000	0.881		0.866		0.964
0.00 (0.00)	0.00 (0.00)	0.50 (0.50)	99000[b]	0.919		0.604		0.964
0.00 (0.00)	0.25 (0.24)	0.75 (0.76)	97750	0.896		0.436		0.963
0.00 (0.00)	0.50 (0.49)	0.50 (0.51)	96500	0.910		0.615		0.961
0.00 (0.00)	0.75 (0.74)	0.25 (0.26)	95250	0.853		0.699		0.958
0.25 (0.24)	0.19 (0.18)	0.56 (0.58)	96560	0.917	0.901	0.606	0.622	0.961*
0.25 (0.24)	0.38 (0.37)	0.38 (0.39)	95630	0.928*	0.935*	0.647	0.637	0.959
0.25 (0.25)	0.56 (0.56)	0.19 (0.20)	94690	0.908	0.917	0.642	0.649	0.957
0.50 (0.49)	0.13 (0.12)	0.38 (0.39)	95380	0.922	0.926	0.780	0.771	0.959
0.50 (0.49)	0.25 (0.25)	0.25 (0.26)	94750	0.902	0.923	0.783	0.839	0.958
0.50 (0.49)	0.38 (0.37)	0.13 (0.13)	94130	0.901	0.921	0.832	0.828	0.956
0.75 (0.74)	0.06 (0.06)	0.19 (0.20)	94190	0.883	0.926	0.755	0.859	0.957
0.75 (0.74)	0.13 (0.13)	0.13 (0.13)	93880	0.882	0.914	0.848	0.870*	0.956
0.75 (0.75)	0.19 (0.19)	0.06 (0.07)	93560	0.853	0.915	0.876*	0.851	0.955

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.
[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

9.3 ASSUMPTIONS AND SELECTION OF HYPOTHESES TO STUDY

As mentioned previously, in addition to the sample design (Section 9.1) and the analysis method utilized (Section 9.2), power calculations for a binary health outcome, *Y*, and binary risk factor, *X*, will also depend on the hypothesis of interest through the sample size, the strength of the relationship (i.e., the odds ratio), the degree of clustering in *X* and *Y*, and the prevalence of *X* and *Y*. Thus, to focus the results on reasonable settings of these factors (i.e., reasonable scenarios), power analysis results are provided for a subset of the hypotheses of interest. Chapter 6 of this report provides a set of descriptions of the hypotheses of the NCS, including estimates of the prevalence of the health outcome, the years of assessment for the hypothesis, and, for some cases, estimates for the prevalence of *X*. Additionally, Chapter 7 provides a set of assumptions for retention and recruitment rates, as a function of the years of follow-up, associated with different methods of selecting the NCS cohort. Using the data provided in these Chapters, the power analysis results are focused on eight scenarios (with at least one scenario selected from each major outcome area). Table 9-4 summarizes the hypotheses that will be studied and outlines the corresponding assumptions necessary in the power calculations.

Table 9-4. Table of hypotheses studied and corresponding assumptions.

Priority Outcome Area	Health Outcome (prevalence)	Risk Factor (prevalence)	W/in Cluster Correlation in Y	W/in Cluster Correlation in X	Years of Follow-Up
Birth Defects and Preterm Birth	Central Nervous System (0.60%)	Impaired Glucose Metabolism During Pregnancy (5%)	0.01	0.02	1
	Congenital Malformations of the Heart (0.60%)	Impaired Glucose Metabolism During Pregnancy (5%)	0.01	0.02	19
Altered Neurobehavioral Development	Cerebral Palsy and Autism (0.25%)	Prenatal Infection and Mediators of Inflammation (20%)	0.005	0.02	7
	Schizophrenia (1%)	Pre/Perinatal Infection and Mediators of Inflammation (20%)	0.01	0.02	21
Injury	Increased Risk of Injury (10%)	Exposure to Neurotoxins or Behavioral Attributes of Childcare (5%)	0.02	0.02	10
	Increased Risk of Injury (10%)	Exposure to Neurotoxins or Behavioral Attributes of Childcare (10%)	0.02	0.02	10
Asthma	Development of Asthma (6%)	Respiratory Viral Infection or Maternal Stress During Pregnancy or Exposure to Air Pollution (1%)	0.02	0.02	10
	Development of Asthma (6%)	Respiratory Viral Infection or Maternal Stress During Pregnancy or Exposure to Air Pollution (5%)	0.02	0.02	10
Obesity and Altered Physical Development	Obesity (15%)	Impaired Glucose Metabolism During Pregnancy (5%)	0.02	0.02	10

It should be noted that while some of the assumptions outlined above, such as the disease prevalence, represent relatively accurate estimates of a "real" scenario, other assumptions are based on limited data. In particular, there is little information on realistic values of the within-cluster correlation for many of the risk factors of interest and the health outcomes of interest (see Chapter 5). Since intraclass correlations are generally very small when dealing with relatively rare outcomes, we will typically assume that there is little within-cluster correlation for both the health outcome and exposure variables. Section D-7 of Appendix D provides further discussion of the implementation and possible interpretation of the within-cluster correlations.

9.4 RESULTS

In the following sections the results of the power calculations are presented in graphical and tabular form. Section 9.4.1 begins by providing an example of the simple random sample (SRS) power results for a selected scenario, and indicating how these results can be utilized to compare different designs, evaluate other hypotheses that may be of interest, evaluate the effect of increasing and/or decreasing the expected occurrence rates of the disease and the exposure

risk factor, and evaluate the effect of increasing or decreasing the expected odds ratio (note that more extensive SRS results are provided in Section D-6 of Appendix D. Section 9.4.2 moves into the more realistic simulation-based results, for the specific scenarios and hypotheses that were selected for investigation.

9.4.1 Simple Random Sample Power Results

As described in Section 9.2, analytical formulas for the power of detecting a specified relationship are available when the data are selected as a simple random sample. For this reason, power can be calculated for a large number of scenarios; however, the power values and resulting conclusions must be interpreted in light of this *simple random sample* assumption (see discussion in Section 9.2). More particularly, the power values do not account for the effect of clustering and unequal weighting that will likely be elements of any feasible NCS design; however, their ease of computation allows investigation of a large number of scenarios and provides insight into the:

- Effect of sample size (which is influenced by retention rates associated with a selected hypothesis and the costs associated with different designs),
- Effect of differing levels of disease and exposure occurrence rates, and
- Odds ratios that can be detected for the different scenarios.

Recall from Section 9.2 that the factors affecting the simple random sample power of detecting a significant relationship between a categorical outcome Y, and a binary risk factor X are: sample size, strength of the relationship, rate of occurrence of X and Y, and the desired significance level of the hypothesis tests. Figure 9-5 displays an example figure for the power of detecting the relationship between a rare health outcome (disease prevalence of 0.25%), such as autism and cerebral palsy, and a binary risk factor. The upper left panel of the figure displays the power as a function of sample size (the different lines) and the odds ratio (the horizontal axis) for a binary exposure risk factor with a prevalence of 1% (the horizontal line is drawn at a power of 0.80). As expected, power increases as a function of sample size and as a function of the strength of the relationship between the outcome and the risk factor. The graph demonstrates that even for a sample size of n=100,000, an odds ratio of close to 3 is required in order to detect a significant relationship with 80% power for this scenario. The other panels of the figure display the same type of information but for different values for the prevalence of the risk factor (5%, 10%, and 20%, respectively). Note that the power picture becomes somewhat more promising as the prevalence of the exposure (or exposure occurrence rate) increases, with odds ratios on the order of 1.5 being detectable with 80% power when the exposure prevalence is 20% (bottom right panel of the figure). From an overall design perspective this generally implies that for diseases with very low prevalence, weak relationships (e.g., odds ratios of 1.1) will be difficult to detect even with a simple random sample of size 100,000 individuals.

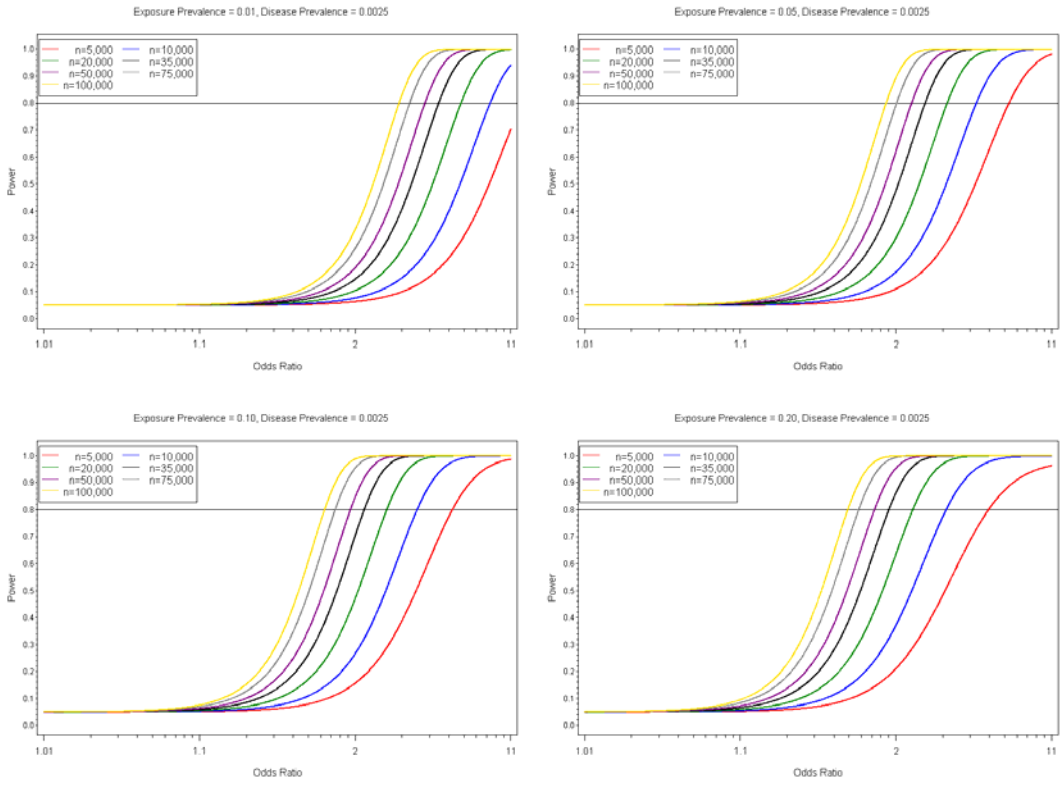


Figure 9-5. SRS power for detecting a significant relationship between a rare health outcome with a prevalence of 0.25% and a binary risk factor.

In Section D-6 of Appendix D we provide these types of figures for a number of scenarios in which the sample size, odds ratio, rate of occurrence of *X*, and rate of occurrence of *Y* are varied over ranges of plausible values. As mentioned previously, these results can be used to evaluate other hypotheses of interest and to assess the sensitivity of the power calculations to the assumptions that are used in conducting them (e.g., sample size assumptions, disease and exposure occurrence rate assumptions, etc.). For example, the effect of assuming different occurrence rates for the risk factor and the effect of assuming different sample sizes, resulting from assuming different retention rates associated with the length of follow-up for the hypothesis of interest or different costing scenarios, can be evaluated.

9.4.2 Simulation-Based Results

The results of the power analyses for each of the hypotheses outlined in Table 9-4 are provided below. Recall that focus is first given to the power for an unweighted analysis since evaluating the relationships of interest within the study population (i.e., assessing internal validity) is the logical first step to analyzing the data obtained in the NCS. In other words, the logical progression for these analyses of the NCS data begins by establishing the presence (or absence) of the relationship within the NCS cohort, and then attempting to generalize that relationship to some larger population (i.e., the NCS sampling frame population). For this

reason, the power results for the unweighted analysis are presented first, followed by the results for the weighted analysis.

Undesirable Outcomes of Pregnancy: Early Diagnosis

For the birth defects priority outcome area, we provide power results for investigating whether impaired glucose metabolism, in women without diabetes before pregnancy, is associated with an increase in the risk of major congenital malformations of the heart, central nervous system, musculoskeletal system, and all birth defects combined. More particularly, as identified in Section 9.3, we focus on congenital heart defects and central nervous system defects, which both occur with a rate of approximately 0.60%, or for around 6 in 1000 individuals (see Chapter 6). The prevalence of the risk factor, impaired glucose metabolism, is assumed to occur in 5% of individuals (see Chapter 6, Table 6-2).

In terms of the year of diagnosis for these health outcomes, typically central nervous system defects are measurable at a very early stage of life, and so for this outcome it is assumed a one-year follow-up period will be necessary for evaluation of the hypothesis. In other words, the one-year retention rates will be used to compute each design's realized sample size for this hypothesis, and, since Chapter 7 suggests that these retention rates are all relatively high (greater than 90% for each mode of sample selection), the assumed sample size for this hypothesis will be close to the initial sample sizes assumed for each design. In contrast, malformations of the heart are often not discovered until later stages of life; thus, in the next section we present results for the hypothesis relating malformations of the heart to impaired glucose metabolism during pregnancy. Note that the only thing changing between the two scenarios is the period of follow-up, and so a comparison of the results will indicate the effect that retention rates can have on hypotheses that require long-term follow-up.

Figure 9-6 displays the power to detect a significant odds ratio for unweighted and weighted analyses, and for fixed sample size and fixed cost cohorts selected using designs A1. A2, B4, F16, G19, and H22 (see Table 9-1). For fixed sample size designs (the top panels of the figure) there appear to be no differences between the designs for an unweighted analysis (top left panel), indicating that all the designs appear to provide the same power for detecting a relationship. This is likely due to their similar sample sizes for fixed sample size designs and hypotheses that are evaluated early in the study. For a weighted analysis (top right panel), as expected, designs with a larger proportion of the cohort selected in the NPBS result in generally higher power to detect the relationship of interest. Additionally, note that the design that selects all subjects from the patient lists of the set of Centers has a weighted analysis power that is similar to the 75% NBPS design; however, recall that this design corresponds to a smaller sampling frame population. Interpreting the graphs in terms of the minimum odds ratios that are detectable with 80% power, an unweighted analysis appears to have sufficient power to detect odds ratios around 1.6 for all of the designs. For a weighted analysis, odds ratios around 1.7 are detectable with 80% power for designs A1, G19, and H22, whereas odds ratios around 1.9 are detectable with 80% power for designs A2, B4, and F16. Considered in this light, there does not appear to be much difference between these six designs if the initial sample size for all the designs is assumed to be 100,000 individuals.

Since an unweighted analysis indicates little differences between the designs, the differences in the weighted analyses are likely attributable to their corresponding differences in the variability of the weights. As described in Chapter 5, the NPBS sampling attempted to obtain a self-weighting sample (i.e., all individuals have approximately the same probability of selection). Thus, designs that involve a larger degree of NPBS sampling will likely have less variability in the weights, and designs that involve higher degrees of Centers sampling (where attempts were not made to obtain a self-weighting sample) will likely have more variability in the weights. This highlights one of the difficulties, or limitations, of a design that attempts to combine individuals sampled from different sampling frames. Namely, the difficulty of obtaining a self-weighting sample when combining individuals sampled from different sampling frames. Chapter 5 describes the methods used in combining the samples for the current power study; however, it should be noted that there may be more optimal ways of combining different sampling frames to obtain a sample that is "closer" to a self-weighting sample. Additionally, it should be noted that there may be more optimal methods for obtaining a self-weighting sample when all of the individuals are selected from a set of 50 purposively selected Centers (i.e., the design that does not include any individuals selected in the NPBS). This may be an important area for further research in the design of the NCS, and is further discussed in Section 9.5.

For the fixed cost designs (bottom panels of Figure 9-6), the picture is slightly different due to the differing initial sample sizes (and resulting differences in the sample size available for analysis) associated with the designs. Note that in this case there are some differences for the unweighted analyses (lower left panel) with designs that result in smaller sample sizes having lower power. In general, for the fixed cost analyses, the designs that select a larger percentage of the cohort in the NPBS appear to have higher power for both the weighted and the unweighted analyses. This is likely due to the larger one-year sample sizes available for these designs. Of course, due to the lower long-term retention rates associated with individuals selected in the NPBS, hypotheses requiring longer periods of follow-up may alter this comparison (in fact, this is one of the factors that allows a larger number of people to be included in the initial sample size for these designs). Evaluating the approximate odds ratios that are detectable with 80% power in a weighted analysis, design H22 (i.e., 75% NPBS) has sufficient power to detect an odds ratio around 1.7, while the other designs are capable of detecting only larger odds ratios.

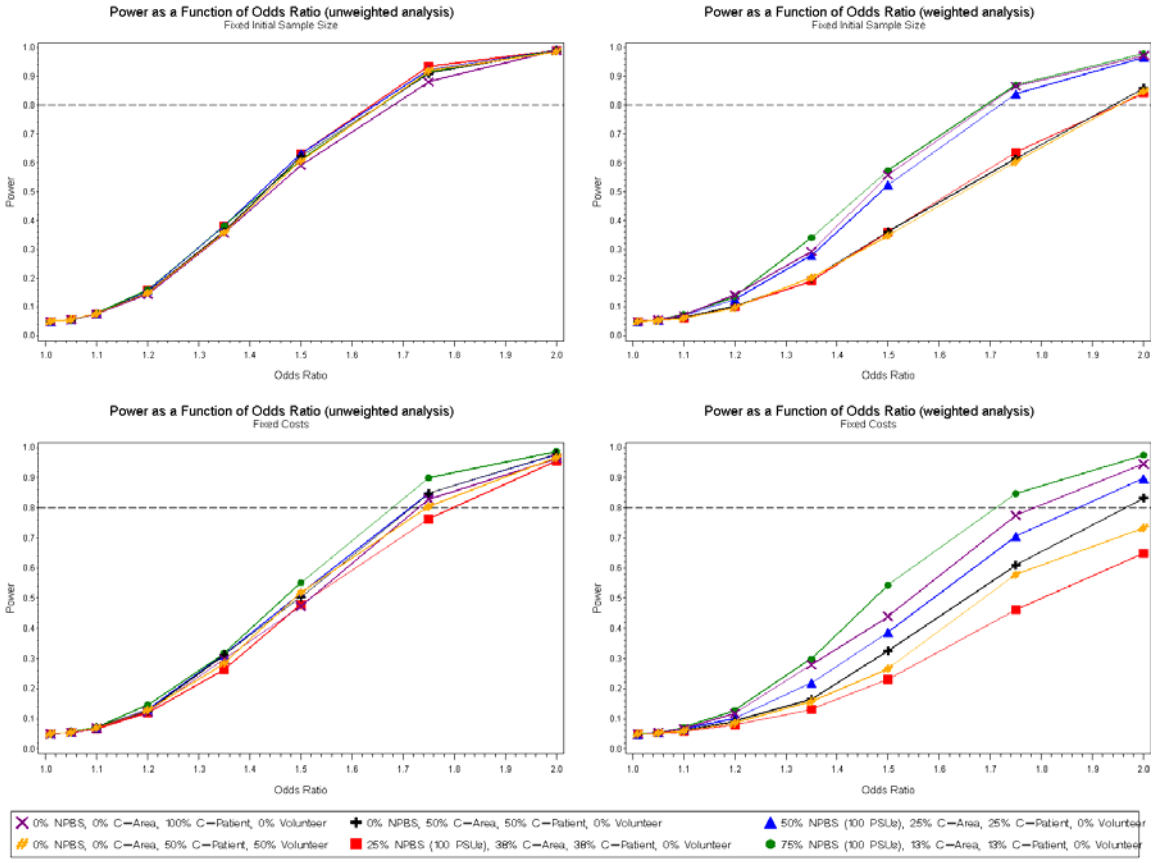


Figure 9-6. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 0.6%, an exposure prevalence of 5%, and 100 PSUs).

As indicated in Figure 9-6, odds ratios around 1.75 are detectable with 80% power for at least some of the designs, indicating that only relatively strong relationships between the outcome and the exposure are detectable for a rare health outcome. Thus, Table 9-5 displays the power to detect an odds ratio of 1.75 for the 23 fixed initial sample size design scenarios under an unweighted analysis, a weighted analysis, and a simple random sample. The realized sample size (N) changes as a function of the design due to the different retention rates associated with the multiple sampling arms of the family of designs. Since a fixed initial sample size for all the designs in this table is assumed and since generally lower retention rates are assumed for individuals selected from the NPBS, designs that have a larger portion of the cohort selected in the NPBS will have smaller realized sample sizes. Note that there are small differences between the power for an unweighted analysis and the SRS power values (designs with higher sample sizes have slightly higher power), and note that there are small differences between the 50 PSU and 100 PSU powers for both an unweighted and a weighted analysis. However, as above, for some of the designs (i.e., those involving larger variability in the weights) there appears to be a significant impact of an unequally weighted design. The implication of these results is that there is generally little effect of clustering when evaluating relationships (see discussion in Section 9.5 for the limitations and how they relate to the simulation assumptions), but, depending on the

amount of unequal weighting in the design, there can be a significant loss of power for a weighted analysis due to unequal probabilities of selection.

Table 9-5. Power to detect an odds ratio of 1.75 as a function of the design (assuming a disease prevalence of 0.6%, an exposure prevalence of 5%, and a fixed initial sample size).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]	Power Unweighted		Power Weighted		Power SRS
				50 PSUs	100 PSUs	50 PSUs	100 PSUs	
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	99000	0.881		0.866		0.964
0.00 (0.00)	0.00 (0.00)	0.50 (0.50)	99000[b]	0.919		0.604		0.964
0.00 (0.00)	0.25 (0.24)	0.75 (0.76)	97750	0.896		0.436		0.963
0.00 (0.00)	0.50 (0.49)	0.50 (0.51)	96500	0.910		0.615		0.961
0.00 (0.00)	0.75 (0.74)	0.25 (0.26)	95250	0.853		0.699		0.958
0.25 (0.24)	0.19 (0.18)	0.56 (0.58)	96560	0.917	0.901	0.606	0.622	0.961*
0.25 (0.24)	0.38 (0.37)	0.38 (0.39)	95630	0.928*	0.935*	0.647	0.637	0.959
0.25 (0.25)	0.56 (0.56)	0.19 (0.20)	94690	0.908	0.917	0.642	0.649	0.957
0.50 (0.49)	0.13 (0.12)	0.38 (0.39)	95380	0.922	0.926	0.780	0.771	0.959
0.50 (0.49)	0.25 (0.25)	0.25 (0.26)	94750	0.902	0.923	0.783	0.839	0.958
0.50 (0.49)	0.38 (0.37)	0.13 (0.13)	94130	0.901	0.921	0.832	0.828	0.956
0.75 (0.74)	0.06 (0.06)	0.19 (0.20)	94190	0.883	0.926	0.755	0.859	0.957
0.75 (0.74)	0.13 (0.13)	0.13 (0.13)	93880	0.882	0.914	0.848	0.870*	0.956
0.75 (0.75)	0.19 (0.19)	0.06 (0.07)	93560	0.853	0.915	0.876*	0.851	0.955

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.

[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

For the 23 fixed cost design scenarios, Table 9-6 displays the power to detect an odds ratio of 1.75 under an unweighted analysis, a weighted analysis, and a simple random sample. Since a fixed cost (and varying initial sample sizes depending on the costs) is assumed for all the designs in this table, the available sample sizes change depending on the cost of implementing the selected design and its corresponding retention rates (note that the 50 PSU and 100 PSU designs result in different sample sizes due to their differing costs). As in Table 9-5 above, the realized sample size (N) in Table 9-6 changes as a function of the design due to the different retention rates associated with the multiple sampling arms of the family of designs. In Table 9-6 we begin to see the effect of cost on the power for a given design. For example, comparing the 50 PSU and 100 PSU power values, we see that the powers for the 50 PSU designs are generally higher as a result of their lower cost (and corresponding larger sample sizes) when compared to the 100 PSU design. As described above, designs that select a larger percentage of the cohort in the NPBS have generally higher power for the weighted analysis, presumably due to both unequal weighting of the designs and the larger resulting sample sizes for these designs. For the unweighted analysis, again designs that select a larger percentage of the cohort in the NPBS have generally higher power, probably due to their larger available samples sizes for this hypothesis (although note that there are small differences between the designs for the unweighted analysis).

Table 9-6. Power to detect an odds ratio of 1.75 as a function of the design (assuming a disease prevalence of 0.6%, an exposure prevalence of 5%, and a fixed design cost).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]		Power Unweighted		Power Weighted		Power SRS	
			50 PSUs	100 PSUs	50 PSUs	100 PSUs	50 PSUs	100 PSUs	50 PSUs	100 PSUs
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	75980		0.828		0.774		0.910	
0.00 (0.00)	0.00 (0.00)	0.50 (0.50)	74990[b]		0.804		0.578		0.906	
0.00 (0.00)	0.25 (0.24)	0.75 (0.76)	78200		0.828		0.364		0.917	
0.00 (0.00)	0.50 (0.49)	0.50 (0.51)	81060		0.848		0.610		0.926	
0.00 (0.00)	0.75 (0.74)	0.25 (0.26)	83820		0.843		0.669		0.934	
0.25 (0.24)	0.19 (0.18)	0.56 (0.58)	74350	60830	0.842	0.756	0.465	0.398	0.904	0.839
0.25 (0.24)	0.38 (0.37)	0.38 (0.39)	76500	63110	0.851	0.762	0.524	0.461	0.911	0.852
0.25 (0.25)	0.56 (0.56)	0.19 (0.20)	79060	65330	0.863	0.801	0.560	0.455	0.920	0.864
0.50 (0.49)	0.13 (0.12)	0.38 (0.39)	84410	69620	0.870	0.852	0.695	0.664	0.935	0.884
0.50 (0.49)	0.25 (0.25)	0.25 (0.26)	86220	71060	0.866	0.848	0.716	0.705	0.940	0.891
0.50 (0.49)	0.38 (0.37)	0.13 (0.13)	88480	72950	0.893	0.841	0.771	0.666	0.945	0.898
0.75 (0.74)	0.06 (0.06)	0.19 (0.20)	97960	80530	0.889	0.864	0.819	0.798	0.963	0.924
0.75 (0.74)	0.13 (0.13)	0.13 (0.13)	99510	81670	0.876	0.899*	0.863*	0.847*	0.965	0.928
0.75 (0.75)	0.19 (0.19)	0.06 (0.07)	100580	83270	0.912*	0.871	0.856	0.830	0.967*	0.932*

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.
[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

Undesirable Outcomes of Pregnancy: Late Diagnosis

To contrast the results for diseases diagnosed early in life to those diagnosed later in life, we re-evaluate the above scenario assuming that a long-term follow-up period is necessary. Above, the hypothesis relating impaired glucose metabolism to malformations of the central nervous system (typically diagnosed at an early age) were studied assuming that a one-year follow-up period was necessary. Here, we evaluate the hypothesis relating impaired glucose metabolism to major congenital malformations of the heart, which may not be diagnosed until later in life (e.g., adolescence). Thus, we assume a 19-year follow-up period so that the outcome of interest is diagnosis of a heart malformation by the age of 18 (e.g., upon graduation from high school). Since the only thing that changes between these two scenarios is the period of follow-up (i.e., both diseases occur with the same prevalence, and the risk factor of interest is the same), a comparison of the results will indicate the effect that retention rates can have on hypotheses requiring long-term follow-up.

Figure 9-7 displays the power to detect a significant odds ratio for unweighted and weighted analyses, and for fixed sample size and fixed cost cohorts selected using designs A1, A2, B4, F16, G19, and H22 (see Table 9-1). Comparing the power curves in this figure to those displayed in Figure 9-6, we immediately see the loss in power that is a result of a longer follow-up period and the corresponding decrease in sample size. Additionally, note that for this case there are differences in both the fixed sample size and fixed cost designs in terms of their power

for an unweighted analysis, with designs resulting in larger sample sizes (i.e., the designs that select none of the cohort in the NPBS) generally having higher power.

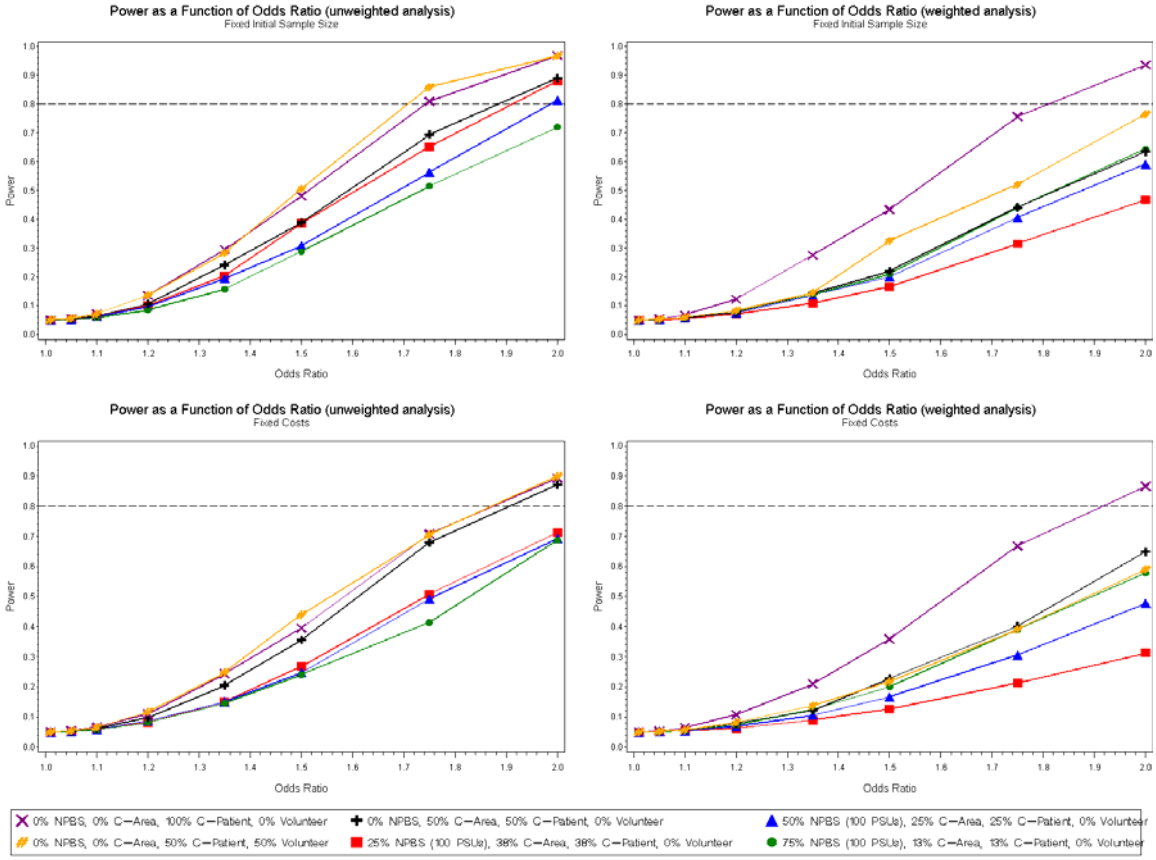


Figure 9-7. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 0.6%, an exposure prevalence of 5%, and 100 PSUs).

Figure 9-7 indicates again that only relatively strong relationships between the outcome and the exposure are detectable for rare health outcomes, especially if that outcome requires an extended period of follow-up. Thus, Table 9-7 displays the power to detect an odds ratio of 2.0 for the 23 fixed initial sample size design scenarios under an unweighted analysis, a weighted analysis, and a simple random sample. Again, note that there is little difference between the 50 PSU and 100 PSU power for both an unweighted and a weighted analysis; however, in contrast to Table 9-5, there are differences between the designs in terms of their power to detect an odds ratio of 2.0 in an unweighted analysis. In particular, designs corresponding to higher retention rates, and thereby larger sample sizes, have higher power for an unweighted analysis. For a weighted analysis, even though they result in smaller sample sizes, the designs that involve 75% of the cohort in the NPBS tend to have the largest power (at least for those designs involving some amount of NPBS sampling).

Table 9-7. Power to detect an odds ratio of 2.0 as a function of the design (assuming a disease prevalence of 0.6%, an exposure prevalence of 5%, and a fixed initial sample size).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]	Power Unweighted		Power Weighted		Power SRS
				50 PSUs	100 PSUs	50 PSUs	100 PSUs	
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	74000	0.969		0.936		0.989
0.00 (0.00)	0.00 (0.00)	0.50 (0.48)	77000[b]	0.967		0.765		0.991
0.00 (0.00)	0.25 (0.12)	0.75 (0.88)	63250	0.929		0.490		0.976
0.00 (0.00)	0.50 (0.30)	0.50 (0.70)	52500	0.890		0.635		0.948
0.00 (0.00)	0.75 (0.56)	0.25 (0.44)	41750	0.844		0.684		0.892
0.25 (0.11)	0.19 (0.11)	0.56 (0.78)	53190	0.860*	0.905*	0.433	0.468	0.950*
0.25 (0.13)	0.38 (0.26)	0.38 (0.61)	45130	0.853	0.879	0.521	0.468	0.913
0.25 (0.16)	0.56 (0.47)	0.19 (0.37)	37060	0.819	0.806	0.515	0.474	0.853
0.50 (0.27)	0.13 (0.09)	0.38 (0.64)	43130	0.847	0.847	0.567	0.604	0.901
0.50 (0.30)	0.25 (0.21)	0.25 (0.49)	37750	0.789	0.813	0.636*	0.591	0.860
0.50 (0.36)	0.38 (0.36)	0.13 (0.29)	32380	0.765	0.763	0.617	0.642	0.803
0.75 (0.52)	0.06 (0.06)	0.19 (0.42)	33060	0.755	0.786	0.625	0.627	0.812
0.75 (0.57)	0.13 (0.13)	0.13 (0.30)	30380	0.748	0.720	0.609	0.643*	0.778
0.75 (0.62)	0.19 (0.21)	0.06 (0.17)	27690	0.695	0.688	0.618	0.608	0.740

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.
[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

For the fixed design costs, Table 9-8 displays the power to detect an odds ratio of 2.0 for the 23 design scenarios under an unweighted analysis, a weighted analysis, and a simple random sample. Conclusions are similar to those discussed above.

Table 9-8. Power to detect an odds ratio of 2.0 as a function of the design (assuming a disease prevalence of 0.6%, an exposure prevalence of 5%, and a fixed design cost).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]		Power Unweighted		Power Weighted		Power SRS	
			50 PSUs	100 PSUs	50 PSUs	100 PSUs	50 PSUs	100 PSUs	50 PSUs	100 PSUs
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	56800		0.894		0.866		0.961	
0.00 (0.00)	0.00 (0.00)	0.50 (0.48)	58330[b]		0.900		0.590		0.965	
0.00 (0.00)	0.25 (0.12)	0.75 (0.88)	50600		0.887		0.379		0.940	
0.00 (0.00)	0.50 (0.30)	0.50 (0.70)	44100		0.872		0.649		0.907	
0.00 (0.00)	0.75 (0.56)	0.25 (0.44)	36740		0.826		0.595		0.850	
0.25 (0.11)	0.19 (0.11)	0.56 (0.78)	40950	33510	0.850*	0.713	0.358	0.274	0.886*	0.817*
0.25 (0.13)	0.38 (0.26)	0.38 (0.61)	36100	29780	0.771	0.714	0.407	0.313	0.844	0.770
0.25 (0.16)	0.56 (0.47)	0.19 (0.37)	30950	25570	0.732	0.638	0.417	0.332	0.786	0.706
0.50 (0.27)	0.13 (0.09)	0.38 (0.64)	38170	31480	0.842	0.751*	0.473	0.463	0.863	0.792
0.50 (0.30)	0.25 (0.21)	0.25 (0.49)	34350	28310	0.771	0.692	0.613	0.477	0.826	0.749
0.50 (0.36)	0.38 (0.36)	0.13 (0.29)	30430	25090	0.714	0.607	0.561	0.433	0.779	0.697
0.75 (0.52)	0.06 (0.06)	0.19 (0.42)	34390	28270	0.777	0.707	0.621	0.555	0.826	0.748
0.75 (0.57)	0.13 (0.13)	0.13 (0.30)	32200	26430	0.753	0.687	0.624	0.578	0.801	0.720
0.75 (0.62)	0.19 (0.21)	0.06 (0.17)	29760	24640	0.708	0.650	0.665*	0.591*	0.770	0.690

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.

[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

Altered Neurobehavioral Development: Cerebral Palsy

For rare neurobehavioral health outcomes, power calculations were conducted to investigate whether prenatal infection or mediators of inflammation (with an assumed prevalence of 20%) are associated with an increase in the risk of cerebral palsy (or autism since these have similar occurrence rates). As indicated in Chapter 6, the prevalence of cerebral palsy is on the order of 0.20% and the prevalence of autism is on the order of 0.30%, thus, for this scenario we assume a health outcome prevalence of 0.25%. Finally, we use the year seven retention rates provided in Chapter 6.

Figure 9-8 displays the power to detect a significant odds ratio for unweighted and weighted analyses, and for fixed sample size and fixed cost cohorts selected using designs A1, A2, B4, F16, G19, and H22 (see Table 9-1). For the fixed sample size designs (top panels of the figure), we begin to see the effect of the varying retention rates (producing different sample sizes) associated with the different modes of selection. For example, since larger retention rates are realized when a smaller portion of the cohort is selected in the NPBS, the unweighted analysis results suggest that a design that includes no NPBS selected individuals has the highest

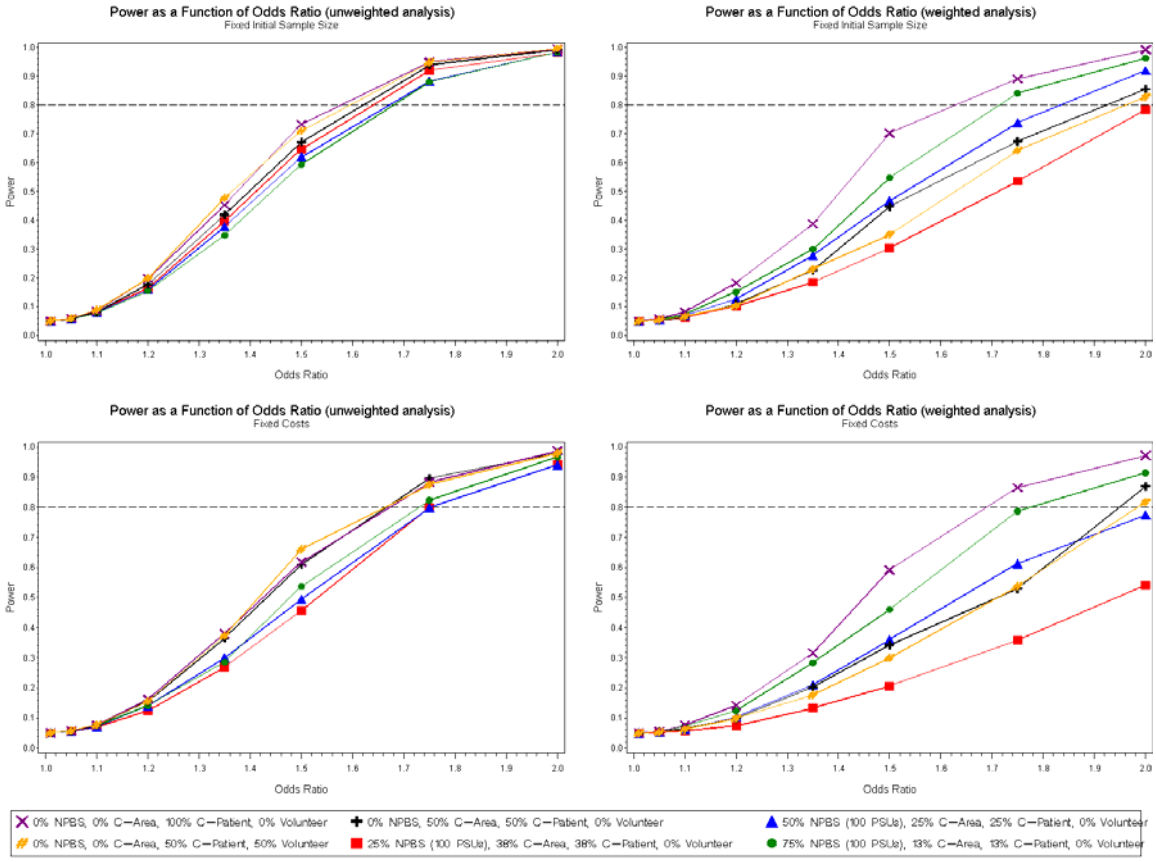


Figure 9-8. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 0.25%, an exposure prevalence of 20%, and 100 PSUs).

power. For weighted analyses, design A1 (i.e., 100% Center patients) has the highest power followed by design H22 (i.e., 75% NBPS design); however, recall that design A1 corresponds to a smaller sampling frame population. Interpretations are relatively similar for the fixed cost designs with the exception that the powers tend to be slightly lower due to the smaller initial sample sizes for these designs.

For the unweighted analyses, the odds ratios that are detectable with 80% power are on the order of 1.6 to 1.7 for all of the designs; however, for the weighted analyses odds ratios between 1.7 and 2.0 are detectable within 80% power (except for the fixed cost design that selects 25% of the cohort in the NPBS, which does not provide adequate power to detect even an odds ratio of 2.0).

As indicated in Figure 9-8, odds ratios around 1.75 are detectable with 80% power for at least some of the designs (i.e., only relatively strong relationships between the outcome and the exposure are detectable for these rare outcomes). Thus, Table 9-9 displays the power to detect an odds ratio of 1.75 for the 23 fixed initial sample size design scenarios under an unweighted analysis, a weighted analysis, and a simple random sample. Again, note that for an unweighted analysis the designs that result in larger sample sizes (e.g., that have smaller portions of the cohort selected in the NPBS) have higher power; whereas for those designs involving some NBPS subjects the designs that have larger portions of the cohort selected in the NPBS have higher power for a weighted analysis even though they correspond to smaller sample sizes. As above, since there are small differences between the unweighted and SRS powers and since there are small differences between the 50 PSU and 100 PSU powers, there again appears to be little effect of clustering on the power to detect relationships of interest (see discussion in Section 9.5).

Table 9-9. Power to detect an odds ratio of 1.75 as a function of the design (assuming a disease prevalence of 0.25%, an exposure prevalence of 20%, and a fixed initial sample size).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]	Power Unweighted		Power Weighted		Power SRS
				50 PSUs	100 PSUs	50 PSUs	100 PSUs	
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	90000	0.949		0.890		0.971
0.00 (0.00)	0.00 (0.00)	0.50 (0.49)	91000[b]	0.948		0.643		0.972
0.00 (0.00)	0.25 (0.19)	0.75 (0.81)	83750	0.956		0.437		0.961
0.00 (0.00)	0.50 (0.42)	0.50 (0.58)	77500	0.938		0.675		0.947
0.00 (0.00)	0.75 (0.68)	0.25 (0.32)	71250	0.881		0.668		0.929
0.25 (0.19)	0.19 (0.16)	0.56 (0.65)	77310	0.916	0.939*	0.582	0.546	0.947*
0.25 (0.20)	0.38 (0.34)	0.38 (0.46)	72630	0.917*	0.921	0.623	0.536	0.934
0.25 (0.21)	0.56 (0.54)	0.19 (0.25)	67940	0.904	0.902	0.599	0.561	0.918
0.50 (0.41)	0.13 (0.11)	0.38 (0.48)	70880	0.903	0.901	0.711	0.730	0.928
0.50 (0.43)	0.25 (0.24)	0.25 (0.33)	67750	0.903	0.883	0.779	0.738	0.917
0.50 (0.45)	0.38 (0.38)	0.13 (0.17)	64630	0.894	0.880	0.770	0.730	0.904
0.75 (0.68)	0.06 (0.06)	0.19 (0.26)	64440	0.912	0.879	0.753	0.759	0.904
0.75 (0.69)	0.13 (0.13)	0.13 (0.18)	62880	0.874	0.880	0.790*	0.842*	0.897
0.75 (0.71)	0.19 (0.20)	0.06 (0.09)	61310	0.880	0.856	0.789	0.836	0.889

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.
[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

Table 9-10 displays the power to detect an odds ratio of 1.75 for the 23 fixed cost design scenarios under an unweighted analysis, a weighted analysis and a simple random sample. For an unweighted analysis, since most of the designs result in similar available sample sizes, similar powers are realized. Comparing the weighted analysis powers for designs that involve some subjects selected in the NPBS, the designs that have larger portions of the cohort selected in the NPBS have higher power, partly due to the larger available sample sizes and partly due to the smaller variability of the weights. In fact, the differences in power between a design with 25% of the original cohort in the NPBS and a design with 75% of the original cohort in the NPBS is on the order of 0.30 to 0.40, indicating a large increase in power as the percentage of the cohort selected in the NPBS gets larger. Finally, note that since a 50 PSU design has lower costs, the sample size available for these designs are much larger than the corresponding 100 PSU design, which results in higher power for the 50 PSU designs.

Table 9-10. Power to detect an odds ratio of 1.75 as a function of the design (assuming a disease prevalence of 0.25%, an exposure prevalence of 20%, and a fixed cost design).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]		Power Unweighted		Power Weighted		Power SRS	
			50 PSUs	100 PSUs	50 PSUs	100 PSUs	50 PSUs	100 PSUs	50 PSUs	100 PSUs
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	69080		0.882		0.865		0.922	
0.00 (0.00)	0.00 (0.00)	0.50 (0.49)	68930[b]		0.875		0.538		0.921	
0.00 (0.00)	0.25 (0.19)	0.75 (0.81)	67000		0.888		0.380		0.914	
0.00 (0.00)	0.50 (0.42)	0.50 (0.58)	65100		0.895		0.530		0.906	
0.00 (0.00)	0.75 (0.68)	0.25 (0.32)	62700		0.900		0.685		0.896	
0.25 (0.19)	0.19 (0.16)	0.56 (0.65)	59530	48710	0.855	0.773	0.458	0.320	0.880	0.809
0.25 (0.20)	0.38 (0.34)	0.38 (0.46)	58100	47930	0.864	0.800	0.474	0.359	0.872	0.803
0.25 (0.21)	0.56 (0.54)	0.19 (0.25)	56730	46880	0.857	0.746	0.519	0.382	0.864	0.794
0.50 (0.41)	0.13 (0.11)	0.38 (0.48)	62720	51740	0.868	0.815	0.687	0.571	0.896	0.832
0.50 (0.43)	0.25 (0.24)	0.25 (0.33)	61650	50810	0.893*	0.798	0.761	0.613	0.891	0.825
0.50 (0.45)	0.38 (0.38)	0.13 (0.17)	60750	50080	0.842	0.789	0.731	0.606	0.886	0.820
0.75 (0.68)	0.06 (0.06)	0.19 (0.26)	67020	55090	0.885	0.796	0.758	0.777	0.914*	0.855*
0.75 (0.69)	0.13 (0.13)	0.13 (0.18)	66650	54700	0.889	0.823	0.848*	0.787*	0.913	0.852
0.75 (0.71)	0.19 (0.20)	0.06 (0.09)	65910	54570	0.870	0.844*	0.832	0.765	0.910	0.851

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.
[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

Altered Neurobehavioral Development: Schizophrenia

To investigate power for an outcome that is assessed later in life (i.e., requires participation in the study for the full 20 years), power calculations were performed investigating whether pre/perinatal infection or mediators of inflammation (with an assumed prevalence of 20%) is associated with an increase in the risk of schizophrenia. As indicated in Chapter 6, the prevalence of schizophrenia is on the order of 1%, and since it can be diagnosed anywhere from infancy through adulthood we use the year 21 retention rates provided in Chapter 7.

As in the previous scenarios, Figure 9-9 displays the power to detect a significant odds ratio for unweighted and weighted analyses, and for fixed sample size and fixed cost cohorts selected using designs A1, A2, B4, F16, G19, and H22 (see Table 9-1). For all of these scenarios we again see the significant impact that retention rates can have on the power to detect relationships between an exposure and a disease that are evaluated later in life (i.e., require longer periods of

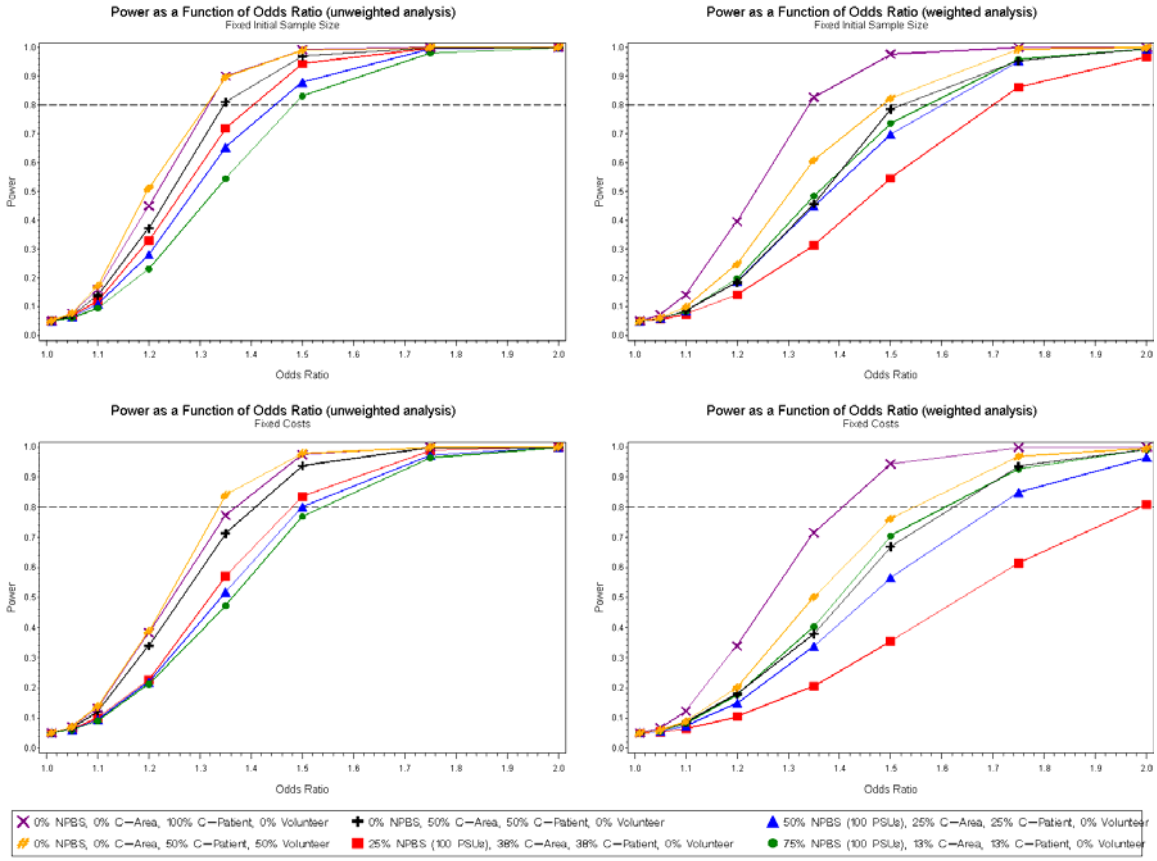


Figure 9-9. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 1%, an exposure prevalence of 20%, and 100 PSUs).

follow-up). For example, all of the panels indicate that the designs with generally highest power are those designs that select none of the individuals in the NPBS (i.e., designs A1, A2, and B4). One reason for this is the significantly higher long-term retention rates for individuals selected in the purposively selected Centers and the resulting gains in the sample size available at the end of the study. Another cause for this result is the fact that designs that select none of the cohort in the NPBS correspond to designs that limit the sampling frame population to a smaller number of individuals. Thus, the acceptability of this reduction in the sampling frame population must be considered when interpreting these results. If we include at least some NPBS sampling, then designs with the largest portion of the cohort selected in the NPBS have the highest power in the weighted analyses (as expected).

For the unweighted analyses, the odds ratios that are detectable with 80% power are on the order of 1.3 to 1.5 for both the fixed cost and fixed sample size approaches. For a weighted analysis with fixed sample size, the detectable odds ratios for most of the designs are on the order of 1.5 to 1.7 (except for design A1, where an odds ratio around 1.35 is detectable with 80% power). Finally, for the weighted analyses under fixed cost designs, the largest differences are apparent, with design A1 able to detect odds ratios around 1.4 with sufficient power, and designs including only 25% of the cohort selected in the NPBS able to detect odds ratios around 2.0.

As indicated in Figure 9-9, odds ratios around 1.5 are detectable with 80% power for a weighted analysis in at least some of the designs. Thus, Table 9-11 displays the power to detect an odds ratio of 1.5 for the 23 fixed initial sample size design scenarios under an unweighted analysis, a weighted analysis, and a simple random sample. For the unweighted analyses, designs with the largest sample size have the highest power (i.e., there appears to be little effect of clustering), and for the weighted analyses, if we exclude the designs that do not include some portion of the cohort selected in the NPBS, the designs with highest power correspond to those designs that select the largest portion of the cohort in the NPBS. However, again note that a design that limits the sampling frame population may offer the largest power to detect odds ratios of 1.5 for this scenario.

Table 9-11. Power to detect an odds ratio of 1.5 as a function of the design (assuming a disease prevalence of 1%, an exposure prevalence of 20%, and a fixed sample size design).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]	Power Unweighted		Power Weighted		Power SRS
				50 PSUs	100 PSUs	50 PSUs	100 PSUs	
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	73000	0.992		0.977		0.998
0.00 (0.00)	0.00 (0.00)	0.50 (0.48)	76000[b]	0.991		0.823		0.999
0.00 (0.00)	0.25 (0.11)	0.75 (0.89)	61750	0.979		0.546		0.994
0.00 (0.00)	0.50 (0.28)	0.50 (0.72)	50500	0.970		0.785		0.981
0.00 (0.00)	0.75 (0.54)	0.25 (0.46)	39250	0.924		0.801		0.944
0.25 (0.10)	0.19 (0.10)	0.56 (0.80)	51310	0.961*	0.974*	0.560	0.520	0.982*
0.25 (0.12)	0.38 (0.24)	0.38 (0.64)	42880	0.927	0.944	0.591	0.546	0.960
0.25 (0.15)	0.56 (0.46)	0.19 (0.40)	34440	0.895	0.899	0.552	0.529	0.914
0.50 (0.24)	0.13 (0.09)	0.38 (0.67)	40880	0.937	0.907	0.666	0.701	0.952
0.50 (0.28)	0.25 (0.20)	0.25 (0.52)	35250	0.901	0.879	0.719	0.698	0.920
0.50 (0.34)	0.38 (0.35)	0.13 (0.31)	29630	0.837	0.850	0.720	0.693	0.870
0.75 (0.49)	0.06 (0.06)	0.19 (0.45)	30440	0.833	0.858	0.669	0.690	0.879
0.75 (0.54)	0.13 (0.13)	0.13 (0.33)	27630	0.794	0.832	0.689	0.737*	0.846
0.75 (0.60)	0.19 (0.21)	0.06 (0.18)	24810	0.789	0.801	0.726*	0.700	0.806

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.
[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

Table 9-12 displays the corresponding results for a fixed cost design. Here we again see the effect of cost in producing designs with different sample sizes (e.g., comparing the 50 PSU and 100 PSU sample sizes and corresponding powers), and we see the effect of unequal weighting in the analysis. As in the previous scenarios, if we exclude those designs that do not include some portion of the cohort selected in the NPBS, the designs that have the highest degree of unequal weighting tend to have the lowest power for a weighted analysis, even if they correspond to larger sample sizes.

Table 9-12. Power to detect an odds ratio of 1.5 as a function of the design (assuming a disease prevalence of 1%, an exposure prevalence of 20%, and a fixed cost design).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]		Power Unweighted		Power Weighted		Power SRS	
			50 PSUs	100 PSUs	50 PSUs	100 PSUs	50 PSUs	100 PSUs	50 PSUs	100 PSUs
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	56030		0.974		0.944		0.989	
0.00 (0.00)	0.00 (0.00)	0.50 (0.48)	57570[b]		0.978		0.762		0.990	
0.00 (0.00)	0.25 (0.11)	0.75 (0.89)	49400		0.956		0.453		0.979	
0.00 (0.00)	0.50 (0.28)	0.50 (0.72)	42420		0.937		0.669		0.958	
0.00 (0.00)	0.75 (0.54)	0.25 (0.46)	34540		0.886		0.721		0.915	
0.25 (0.10)	0.19 (0.10)	0.56 (0.80)	39510	32330	0.919*	0.872*	0.421	0.329	0.946*	0.897*
0.25 (0.12)	0.38 (0.24)	0.38 (0.64)	34300	28300	0.896	0.836	0.465	0.355	0.913	0.854
0.25 (0.15)	0.56 (0.46)	0.19 (0.40)	28760	23760	0.847	0.774	0.459	0.357	0.860	0.789
0.50 (0.24)	0.13 (0.09)	0.38 (0.67)	36170	29840	0.885	0.820	0.592	0.524	0.926	0.872
0.50 (0.28)	0.25 (0.20)	0.25 (0.52)	32080	26440	0.871	0.801	0.680	0.566	0.894	0.830
0.50 (0.34)	0.38 (0.35)	0.13 (0.31)	27850	22960	0.803	0.738	0.686	0.578	0.849	0.775
0.75 (0.49)	0.06 (0.06)	0.19 (0.45)	31660	26020	0.838	0.780	0.715	0.643	0.891	0.824
0.75 (0.54)	0.13 (0.13)	0.13 (0.33)	29280	24030	0.875	0.769	0.724	0.705*	0.866	0.794
0.75 (0.60)	0.19 (0.21)	0.06 (0.18)	26670	22080	0.812	0.707	0.782*	0.660	0.833	0.759

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.
[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

Injury

To investigate power for the injury outcomes, power calculations for the association between injury (with an assumed prevalence of 10%) and two "exposure" variables, one with a prevalence of 5% and one with a prevalence of 10%, were investigated. In this case, we assume a 10-year follow-up period (i.e., injury and exposure evaluated at age 9) and use the corresponding retention rates supplied in Chapter 7.

Figures 9-10 and 9-11 display the power to detect a significant odds ratio for unweighted and weighted analyses, and for fixed sample size and fixed cost cohorts selected using designs A1, A2, B4, F16, G19, and H22 (see Table 9-1). In particular, Figure 9-10 corresponds to the case of an exposure variable with a prevalence of 5%, and Figure 9-11 corresponds to the case of an exposure variable with a prevalence of 10%. As indicated in the figures, for diseases with higher prevalence, smaller odds ratios are detectable (e.g., comparing these figures to the figures for the previous scenarios). In addition, in comparing Figure 9-10 to Figure 9-11 we see the effect of increasing the exposure prevalence. In particular, note that the power displayed in Figure 9-11 is greater than the corresponding power displayed in Figure 9-10, as expected.

Comparing the designs displayed in Figure 9-10, there are essentially no differences in the designs for an unweighted analysis, but for a weighted analysis there appear to be some slight differences. In particular, the fixed sample size designs (top panels) have slightly higher power then the fixed cost designs (bottom panels), due to their larger available sample sizes. Additionally, note that the design that selects all participants from the set of Center patients

generally has the highest power for both weighted and unweighted analyses. Finally, excluding the designs that select no NPBS participants, for both the fixed cost and fixed sample size designs, the designs that include 75% of the cohort selected in the NPBS have the highest power and can detect odds ratios on the order of 1.25 with 80% power, whereas, the designs that include 50% of the cohort selected in the NPBS have slightly lower power and can detect odds ratios on the order of 1.3 with 80% power.

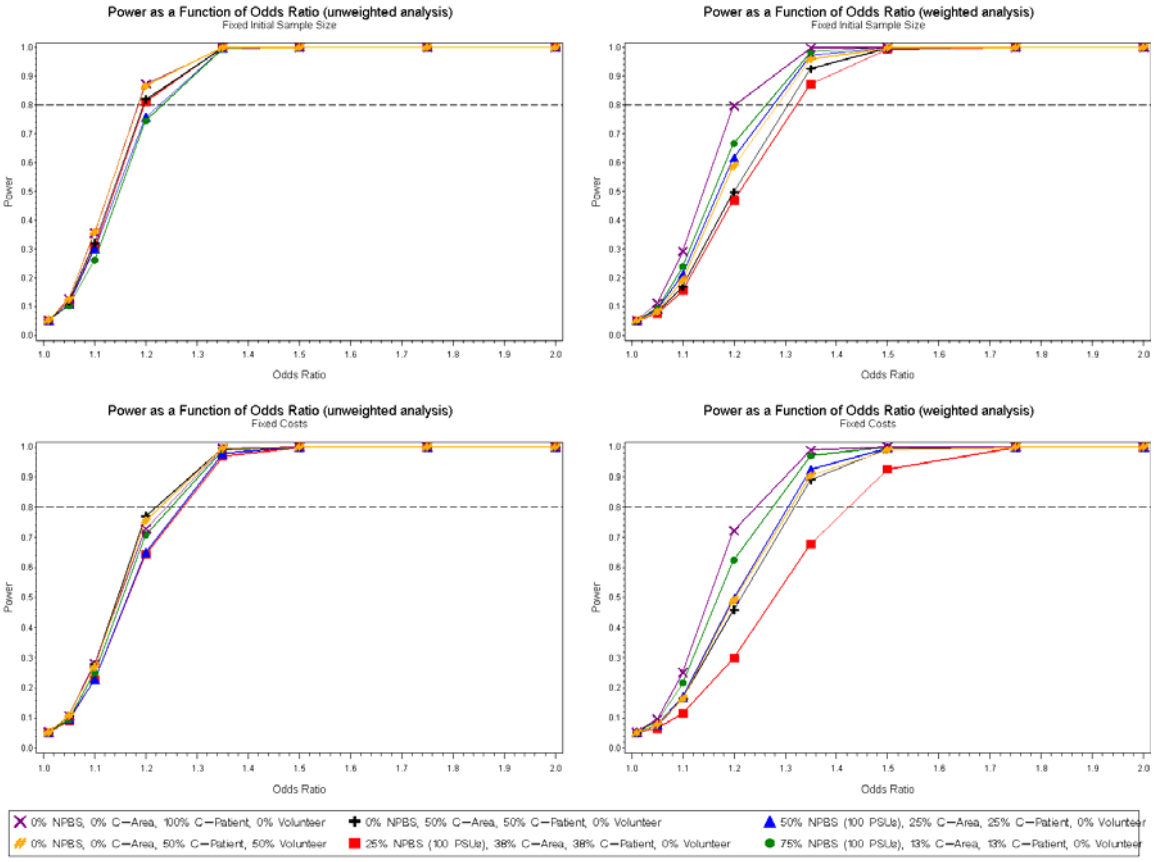


Figure 9-10. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 10%, an exposure prevalence of 5%, and 100 PSUs).

Comparing the designs displayed in Figure 9-11 (for which a higher exposure prevalence was assumed) slightly higher powers are exhibited but similar characteristics are apparent. The range of odds ratios detectable with 80% power is approximately 1.15 to 1.25 for the fixed sample size designs, and is approximately 1.15 to 1.3 for the fixed cost designs.

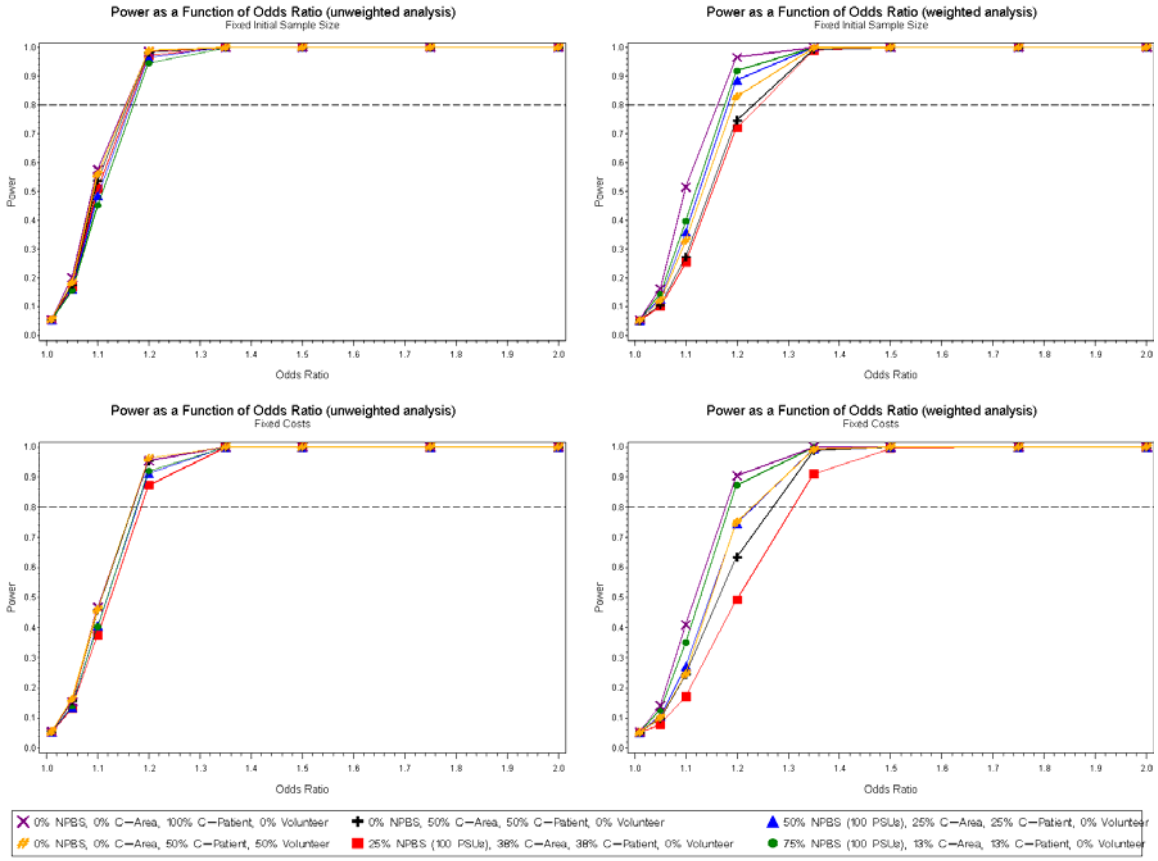


Figure 9-11. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 10%, an exposure prevalence of 10%, and 100 PSUs).

Focusing on the results corresponding to an exposure prevalence of 5%, Tables 9-13 and 9-14 display the power to detect an odds ratio of 1.2 for the fixed sample size designs and the fixed cost designs, respectively. As in many of the previous analyses, for the fixed sample size designs (Table 9-13), the differing sample sizes result in differing powers for the unweighted analyses (i.e., larger sample sizes have higher powers), and for the weighted analyses the designs with a larger portion of the cohort selected in the NPBS tend to have higher powers when comparing designs that include some participants selected in the NPBS.

Table 9-13. Power to detect an odds ratio of 1.2 as a function of the design (assuming a disease prevalence of 10%, an exposure prevalence of 5%, and a fixed sample size design).

% of Original Cohort in NPBS (% of N)	% of Original Cohort in C-Area (% of N)	% of Original Cohort in C-Patient (% of N)	N[a]	Power Unweighted		Power Weighted		Power SRS
				50 PSUs	100 PSUs	50 PSUs	100 PSUs	
0.00 (0.00)	0.00 (0.00)	1.00 (1.00)	85000	0.873		0.797		0.959
0.00 (0.00)	0.00 (0.00)	0.50 (0.49)	87000[b]	0.866		0.588		0.963
0.00 (0.00)	0.25 (0.17)	0.75 (0.83)	77000	0.822		0.354		0.941
0.00 (0.00)	0.50 (0.38)	0.50 (0.62)	69000	0.820		0.498		0.916
0.00 (0.00)	0.75 (0.65)	0.25 (0.35)	61000	0.801		0.519		0.881
0.25 (0.16)	0.19 (0.14)	0.56 (0.69)	69000	0.815	0.825*	0.458	0.455	0.916*
0.25 (0.18)	0.38 (0.32)	0.38 (0.51)	63000	0.821*	0.811	0.452	0.469	0.890
0.25 (0.20)	0.56 (0.52)	0.19 (0.28)	57000	0.771	0.765	0.484	0.445	0.858
0.50 (0.37)	0.13 (0.11)	0.38 (0.52)	61000	0.752	0.812	0.539	0.608	0.881
0.50 (0.39)	0.25 (0.23)	0.25 (0.37)	57000	0.779	0.758	0.596	0.617	0.858
0.50 (0.42)	0.38 (0.38)	0.13 (0.20)	53000	0.749	0.764	0.630	0.611	0.833
0.75 (0.64)	0.06 (0.06)	0.19 (0.30)	53000	0.728	0.749	0.594	0.654	0.833
0.75 (0.66)	0.13 (0.13)	0.13 (0.21)	51000	0.725	0.747	0.644*	0.667*	0.818
0.75 (0.69)	0.19 (0.20)	0.06 (0.11)	49000	0.710	0.753	0.615	0.658	0.803

[a] N is the sample size available for analysis, which depends on the retention rates, the original sample size, and the period of follow-up for the hypothesis.
[b] Note that the volunteer subjects are excluded when conducting a weighted analysis.

The fixed cost designs of Table 9-14 again display the effects of cost on power if we compare the 50 PSU design to the 100 PSU design under the same selection scenario. Due to its higher cost, and resulting smaller sample size, the 100 PSU design has lower power than the 50 PSU design. Again, many of the conclusions mentioned for previous examples are relevant to the results displayed in this table (e.g., higher power for the 75% NPBS design when comparing designs F16, G19, and H22).