Table 9-14. 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 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)	65240		0.729		0.722		0.901	
0.00 (0.00)	0.00 (0.00)	0.50 (0.49)	65900[b]		0.756		0.491		0.904	
0.00 (0.00)	0.25 (0.17)	0.75 (0.83)	61600		0.757		0.252		0.884	
0.00 (0.00)	0.50 (0.38)	0.50 (0.62)	57960		0.770		0.459		0.864	
0.00 (0.00)	0.75 (0.65)	0.25 (0.35)	53680		0.744		0.534		0.837	
0.25 (0.16)	0.19 (0.14)	0.56 (0.69)	53130	43470	0.685	0.656	0.368	0.291	0.834	0.755
0.25 (0.18)	0.38 (0.32)	0.38 (0.51)	50400	41580	0.743	0.644	0.386	0.299	0.814	0.736
0.25 (0.20)	0.56 (0.52)	0.19 (0.28)	47600	39330	0.695	0.648	0.401	0.312	0.792	0.712
0.50 (0.37)	0.13 (0.11)	0.38 (0.52)	53990	44530	0.744	0.649	0.535	0.450	0.839	0.765
0.50 (0.39)	0.25 (0.23)	0.25 (0.37)	51870	42750	0.749*	0.650	0.560	0.497	0.825	0.748
0.50 (0.42)	0.38 (0.38)	0.13 (0.20)	49820	41080	0.732	0.655	0.582	0.515	0.809	0.731
0.75 (0.64)	0.06 (0.06)	0.19 (0.30)	55120	45320	0.733	0.678	0.619	0.595	0.847*	0.772*
0.75 (0.66)	0.13 (0.13)	0.13 (0.21)	54060	44370	0.734	0.709*	0.655	0.624	0.840	0.763
0.75 (0.69)	0.19 (0.20)	0.06 (0.11)	52680	43610	0.714	0.691	0.702*	0.630*	0.831	0.756

[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.

Asthma

To investigate power for the asthma outcomes, power calculations for the association between asthma (with an assumed prevalence of 6%) and two "exposure" variables (e.g., exposure to air pollution, maternal stress during pregnancy, or respiratory viral infection), one with a prevalence of 1% and one with a prevalence of 5%, were investigated. We provide results for two values of exposure prevalence since there are several hypotheses related to the asthma outcome and its association with a variety of exposures. The hope is that these two levels of exposure prevalence are representative of several of the risk factors of interest for this hypothesis. A 10-year follow-up period (i.e., asthma by age 9) is assumed and the corresponding retention rates are utilized.

Figures 9-12 and 9-13 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 9-12 corresponds to the case of an exposure variable with a prevalence of 1%, and Figure 9-13 corresponds to the case of an exposure variable with a prevalence of 5%. Again, the differences in the exposure prevalence for these two figures result in relatively large differences in the power to detect associations of interest. For example, assuming an exposure prevalence of 1%, odds ratios on the order of 1.75 or 2 are detectable with 80% power in a weighted analysis. On the other hand, assuming an exposure prevalence of 5%, odds ratios on the order of 1.3 are detectable with 80% power in a weighted analysis.

As in each of the above examples, designs that provide larger sample sizes generally provide higher power for the unweighted analyses (this is true for both the fixed sample size and fixed cost designs). For the fixed sample size and fixed cost weighted analyses displayed (right panels), the design that selects the entire cohort from Center patients has consistently higher power than the other designs (again note that this design corresponds to a fundamentally smaller sampling frame population).

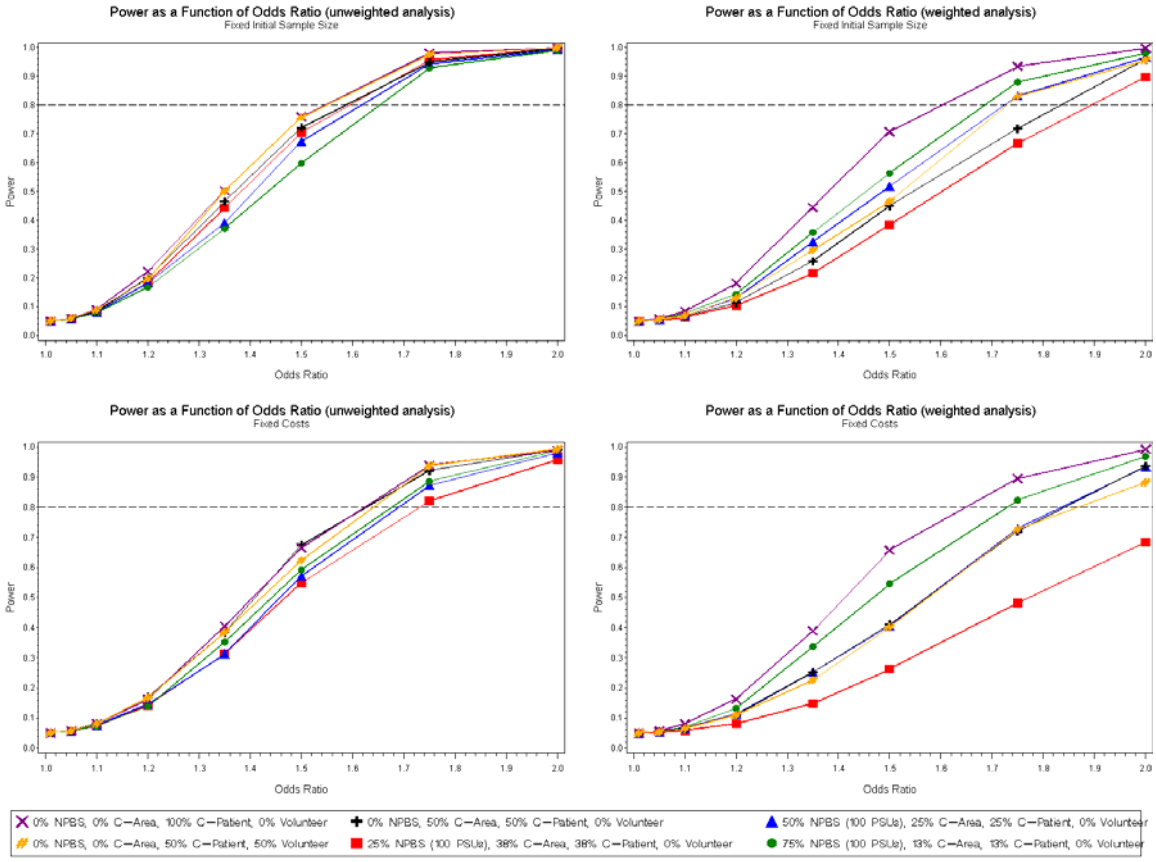

Figure 9-12. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 6%, an exposure prevalence of 1%, and 100 PSUs).

Comparing the designs displayed in Figure 9-13 (where a higher exposure prevalence was assumed), slightly higher powers are exhibited but similar characteristics are apparent. For a weighted analysis, the range of odds ratios detectable with 80% power is approximately 1.3 to 1.4 in the fixed sample size designs, and is approximately 1.3 to 1.5 in the fixed cost designs.

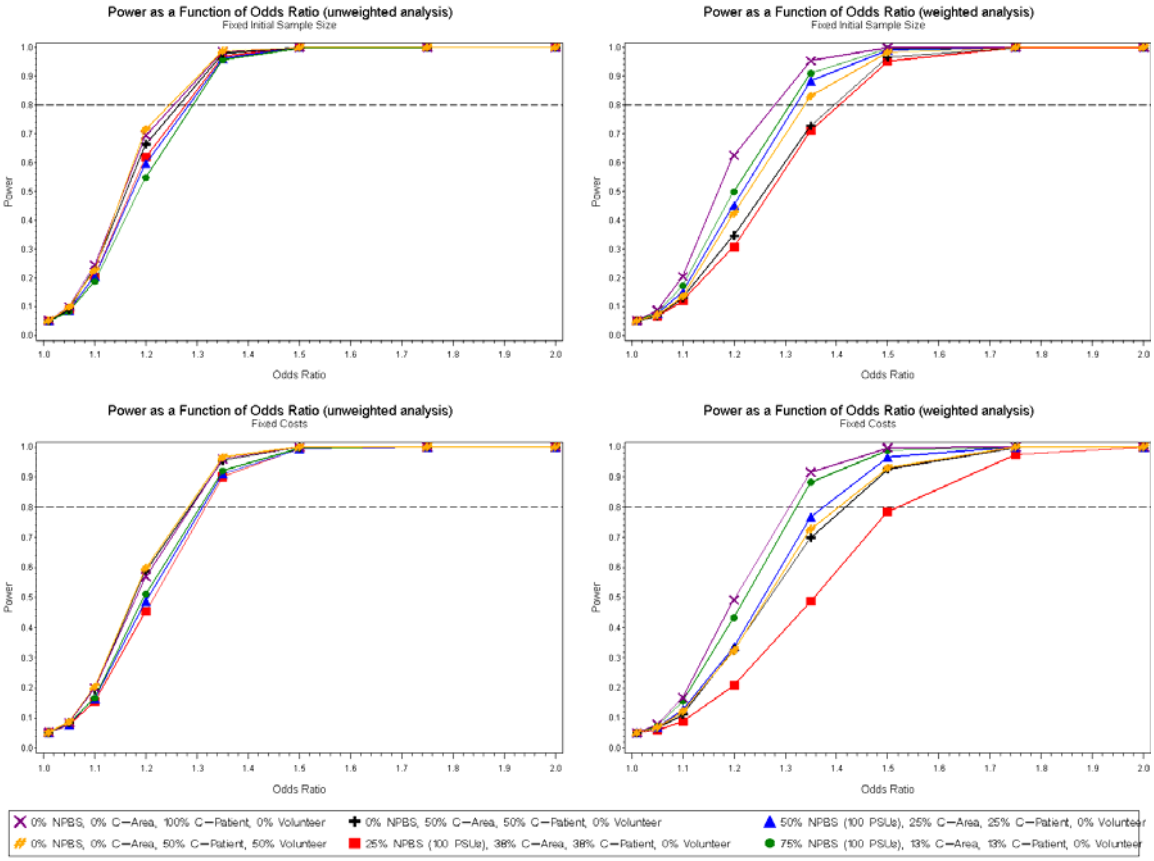

Figure 9-13. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 6%, an exposure prevalence of 5%, and 100 PSUs).

Focusing on the results corresponding to an exposure prevalence of 5%, Tables 9-15 and 9-16 display the power to detect an odds ratio of 1.35 for the fixed sample size designs and the fixed cost designs, respectively. Note that since the disease prevalence tends to be higher, these smaller odds ratios are detectable with sufficient power. This may be a very important characteristic for diseases such as asthma, where there are likely a large number of exposures, each having a weak relationship to the development of asthma. From a public health perspective, being able to detect these weak relationships for diseases that are more common is likely a desirable characteristic for the NCS, since improving occurrence rates (i.e., decreasing them by educating people on the factors that aid in development of the disease) for diseases that are common would result in improving the health of a larger number of individuals.

Table 9-15. Power to detect an odds ratio of 1.35 as a function of the design (assuming a disease prevalence of 6%, 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	0.986		0.953		0.999
0.00 (0.00)	0.00 (0.00)	0.50 (0.49)	87000[b]	0.988		0.831		0.999
0.00 (0.00)	0.25 (0.17)	0.75 (0.83)	77000	0.979		0.547		0.998
0.00 (0.00)	0.50 (0.38)	0.50 (0.62)	69000	0.979		0.728		0.996
0.00 (0.00)	0.75 (0.65)	0.25 (0.35)	61000	0.970		0.805		0.990
0.25 (0.16)	0.19 (0.14)	0.56 (0.69)	69000	0.976*	0.976*	0.691	0.727	0.996*
0.25 (0.18)	0.38 (0.32)	0.38 (0.51)	63000	0.972	0.969	0.741	0.711	0.992
0.25 (0.20)	0.56 (0.52)	0.19 (0.28)	57000	0.969	0.963	0.734	0.724	0.986
0.50 (0.37)	0.13 (0.11)	0.38 (0.52)	61000	0.955	0.961	0.800	0.853	0.990
0.50 (0.39)	0.25 (0.23)	0.25 (0.37)	57000	0.963	0.962	0.874	0.884	0.986
0.50 (0.42)	0.38 (0.38)	0.13 (0.20)	53000	0.943	0.961	0.866	0.868	0.980
0.75 (0.64)	0.06 (0.06)	0.19 (0.30)	53000	0.919	0.951	0.849	0.907	0.980
0.75 (0.66)	0.13 (0.13)	0.13 (0.21)	51000	0.931	0.957	0.880	0.910	0.976
0.75 (0.69)	0.19 (0.20)	0.06 (0.11)	49000	0.932	0.938	0.908*	0.915*	0.971

[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 9-16. Power to detect an odds ratio of 1.35 as a function of the design (assuming a disease prevalence of 6%, an exposure prevalence of 5%, 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)	65240		0.960		0.916		0.994	
0.00 (0.00)	0.00 (0.00)	0.50 (0.49)	65900[b]		0.966		0.728		0.994	
0.00 (0.00)	0.25 (0.17)	0.75 (0.83)	61600		0.950		0.541		0.991	
0.00 (0.00)	0.50 (0.38)	0.50 (0.62)	57960		0.955		0.699		0.987	
0.00 (0.00)	0.75 (0.65)	0.25 (0.35)	53680		0.942		0.799		0.981	
0.25 (0.16)	0.19 (0.14)	0.56 (0.69)	53130	43470	0.928	0.893	0.583	0.479	0.980	0.953
0.25 (0.18)	0.38 (0.32)	0.38 (0.51)	50400	41580	0.935	0.901	0.619	0.488	0.974	0.944
0.25 (0.20)	0.56 (0.52)	0.19 (0.28)	47600	39330	0.920	0.896	0.671	0.510	0.967	0.933
0.50 (0.37)	0.13 (0.11)	0.38 (0.52)	53990	44530	0.942	0.889	0.792	0.735	0.982	0.957
0.50 (0.39)	0.25 (0.23)	0.25 (0.37)	51870	42750	0.943	0.910	0.822	0.768	0.978	0.950
0.50 (0.42)	0.38 (0.38)	0.13 (0.20)	49820	41080	0.933	0.900	0.841	0.756	0.973	0.942
0.75 (0.64)	0.06 (0.06)	0.19 (0.30)	55120	45320	0.945*	0.909	0.848	0.859	0.983*	0.960*
0.75 (0.66)	0.13 (0.13)	0.13 (0.21)	54060	44370	0.937	0.921*	0.896*	0.883*	0.982	0.956
0.75 (0.69)	0.19 (0.20)	0.06 (0.11)	52680	43610	0.927	0.909	0.876	0.871	0.979	0.953

[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.

Obesity and Altered Physical Development

Finally, to investigate power for the obesity outcomes, power calculations were performed for the association between the risk of obesity (with an assumed prevalence of 15%) and maternal impaired glucose metabolism during pregnancy (with an assumed prevalence of 5%). Again, a 10-year follow-up period (i.e., obesity assessed at age 9) is assumed and the corresponding retention rates are utilized.

Figure 9-14 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). Again, note that the higher assumed disease prevalence results in sufficient power to detect smaller odds ratios. In particular, for an unweighted analysis, odds ratios on the order of 1.1 to 1.2 are detectable for all of the designs, and for a weighted analysis, odds ratios on the order of 1.2 to 1.3 are detectable for all of the designs. In addition, for the unweighted analyses, there are again relatively small differences in the designs resulting from the different realized sample sizes associated with each design. On the other hand, for the weighted analyses corresponding to designs that select some portion of the cohort in the NPBS, the effect of unequal weighting of the cohort is apparent with the design that selects 75% of the cohort in the NPBS having the largest power, followed by the 50% NPBS design, and the 25% NPBS design.

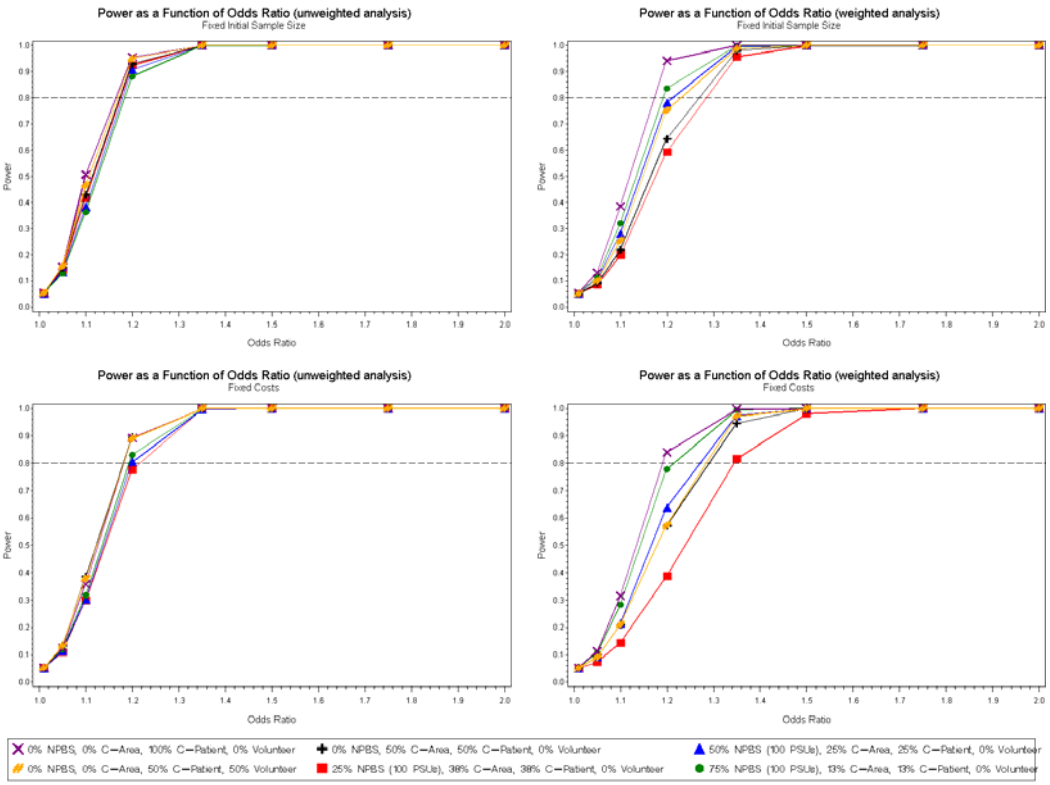

Figure 9-14. Power to detect a significant odds ratio for unweighted (left side panels) and weighted (right side panels) analyses and for fixed sample size (top panels) and fixed cost samples (bottom panels) (assuming a disease prevalence of 15%, an exposure prevalence of 5%, and 100 PSUs).

As indicated in Figure 9-14, odds ratios around 1.20 are detectable with 80% power for a weighted analysis in at least some of the designs. Thus, Table 9-17 displays the power to detect an odds ratio of 1.20 for the 23 fixed initial sample size design scenarios under an unweighted analysis, a weighted analysis, and a simple random sample. For the unweighted analysis, designs with the largest sample size have the highest power (i.e., there appears to be little effect of clustering), and for the weighted analysis, 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 larger portions of the cohort in the NPBS.

Table 9-17. Power to detect an odds ratio of 1.20 as a function of the design (assuming a disease prevalence of 15%, 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	0.956		0.942		0.992
0.00 (0.00)	0.00 (0.00)	0.50 (0.49)	87000[b]	0.950		0.752		0.993
0.00 (0.00)	0.25 (0.17)	0.75 (0.83)	77000	0.944		0.407		0.986
0.00 (0.00)	0.50 (0.38)	0.50 (0.62)	69000	0.930		0.643		0.976
0.00 (0.00)	0.75 (0.65)	0.25 (0.35)	61000	0.913		0.729		0.959
0.25 (0.16)	0.19 (0.14)	0.56 (0.69)	69000	0.935*	0.929*	0.579	0.559	0.976*
0.25 (0.18)	0.38 (0.32)	0.38 (0.51)	63000	0.915	0.925	0.620	0.593	0.964
0.25 (0.20)	0.56 (0.52)	0.19 (0.28)	57000	0.895	0.891	0.605	0.575	0.947
0.50 (0.37)	0.13 (0.11)	0.38 (0.52)	61000	0.894	0.910	0.734	0.720	0.959
0.50 (0.39)	0.25 (0.23)	0.25 (0.37)	57000	0.878	0.908	0.734	0.781	0.947
0.50 (0.42)	0.38 (0.38)	0.13 (0.20)	53000	0.870	0.890	0.800*	0.763	0.932
0.75 (0.64)	0.06 (0.06)	0.19 (0.30)	53000	0.840	0.881	0.726	0.770	0.932
0.75 (0.66)	0.13 (0.13)	0.13 (0.21)	51000	0.875	0.882	0.783	0.836*	0.923
0.75 (0.69)	0.19 (0.20)	0.06 (0.11)	49000	0.839	0.864	0.798	0.826	0.913

[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.

Finally, Table 9-18 displays the power to detect an odds ratio of 1.20 for the 23 fixed cost design scenarios under an unweighted analysis, a weighted analysis and a simple random sample. For the unweighted analysis, the designs corresponding to smaller sampling frame populations have the highest power, due to their larger available sample size. For the weighted analyses, the 75% NPBS designs generally have higher power when comparing designs that include some portion of the cohort selected in the NPBS. Also, again note that the 50 PSU designs have higher power than the corresponding 100 PSU design due to their lower costs and corresponding larger sample sizes.

Table 9-18. Power to detect an odds ratio of 1.20 as a function of the design (assuming a disease prevalence of 15%, an exposure prevalence of 5%, 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)	65240		0.892		0.840		0.969	
0.00 (0.00)	0.00 (0.00)	0.50 (0.49)	65900[b]		0.890		0.574		0.971	
0.00 (0.00)	0.25 (0.17)	0.75 (0.83)	61600		0.872		0.391		0.961	
0.00 (0.00)	0.50 (0.38)	0.50 (0.62)	57960		0.894		0.571		0.950	
0.00 (0.00)	0.75 (0.65)	0.25 (0.35)	53680		0.851		0.659		0.935	
0.25 (0.16)	0.19 (0.14)	0.56 (0.69)	53130	43470	0.854	0.771	0.465	0.371	0.933	0.878
0.25 (0.18)	0.38 (0.32)	0.38 (0.51)	50400	41580	0.851	0.776	0.506	0.388	0.920	0.864
0.25 (0.20)	0.56 (0.52)	0.19 (0.28)	47600	39330	0.843	0.758	0.525	0.420	0.905	0.845
0.50 (0.37)	0.13 (0.11)	0.38 (0.52)	53990	44530	0.840	0.768	0.646	0.625	0.936	0.886
0.50 (0.39)	0.25 (0.23)	0.25 (0.37)	51870	42750	0.878*	0.805	0.674	0.638	0.927	0.873
0.50 (0.42)	0.38 (0.38)	0.13 (0.20)	49820	41080	0.877	0.795	0.737	0.636	0.917	0.860
0.75 (0.64)	0.06 (0.06)	0.19 (0.30)	55120	45320	0.865	0.835*	0.752	0.744	0.941*	0.891*
0.75 (0.66)	0.13 (0.13)	0.13 (0.21)	54060	44370	0.874	0.831	0.799*	0.778*	0.937	0.885
0.75 (0.69)	0.19 (0.20)	0.06 (0.11)	52680	43610	0.860	0.823	0.798	0.743	0.931	0.879

[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.

9.5 CONCLUSIONS AND LIMITATIONS

As suggested previously, 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. The results presented above exhibit some of these dissimilarities for the different hypotheses and different analysis methods (i.e., weighted versus unweighted analyses). This makes it very difficult to identify a single "optimal" design strategy for the NCS; however, some general conclusions based on the results presented above may be appropriate:

- For unweighted analyses (i.e., analyses that account for design clustering but not unequal weighting), the design that provides the largest available sample size generally corresponds to the design with the highest power.
- For the fixed sample size designs, there is relatively little difference in power when using a 50 PSU design versus a 100 PSU design (assuming the same proportion of the cohort is selected in the NPBS), suggesting that there is only a small affect of clustering for PSU sizes on the order of 100. This result is consistent for both the unweighted and weighted analyses. On the other hand, for the fixed cost designs, the 50 PSU design provides greater power than the 100 PSU design (again, assuming the same proportion of the cohort is selected in the NPBS) due to its lower costs and resulting ability to follow a larger cohort of children.

- Comparing the simple random sample power results to the corresponding results for an unweighted analysis (i.e., an analysis that accounts for the possible clustering in the design), there appear to be small differences in the powers, indicating little effect of clustering on the ability to detect relationships of interest in these scenarios.
- Comparing the power for a weighted analysis to that for an unweighted analysis, many of the designs indicate a larger effect of unequal weighting (at least larger than the effect of clustering) in the cohort with designs that have a larger portion of the cohort selected in the NPBS having generally higher power when comparing designs that select some portion of the cohort in NPBS. This remains the case even if these designs result in smaller sample sizes, and indicates that alternative statistical approaches to combining individuals selected from different sampling frames in order to obtain a "more" self-weighting sample may be a promising avenue for further research
- Unless it is acceptable to limit the sampling frame population (i.e., select 0% of the cohort in the NPBS) to maximize power for a weighted analysis, the results suggest that as much of the cohort as possible should be selected in the NPBS.
- Alternatively, to maximize power for an unweighted analysis, the results suggest that the cohort should be selected in such a way as to obtain the highest possible retention rates.
- If it is acceptable to limit the sampling frame population to a set of cities (or MSAs) that correspond to a group of purposively selected Centers, then there may be cost and retention gains associated with these designs. In addition, we envision that there may be more optimal ways of conducting the sampling in the purposively selected Centers to obtain a more self-weighting sample (resulting in higher powers for a weighted analysis). Due to time constraints, the sample selection for the Centers was intended simply to obtain random samples, *with the same size from every Center,* from the appropriate sampling frames. Allowing the size of the sample to vary as a function of the "size" of the Center could provide a more optimal approach to the design within a set of purposively selected Centers; however, this may detract from the benefits of conducting a study of substantial size in smaller Centers. Alternatively, perhaps efforts could be made to equalize the Center populations (e.g., by including geographic areas of different sizes depending on the population density associated with a Center), or to select Centers in some probabilistic manner.
- When comparing the power study results from the weighted analyses of designs that involve some NPBS sampling to the designs that select all of the individuals from a set of purposively selected Centers it should be noted that the sampling frame populations are fundamentally different.
- Interpreting the designs in terms of their power to detect a fixed odds ratio, there can be somewhat large differences between the powers (e.g., with the less optimal designs having 30 to 40% less power than the most optimal designs).
- On the other hand, comparing the designs in terms of the odds ratio that can be detected with 80% power, there often appear to be "small" differences between the designs. Of course, the definition of a "small" difference in the odds ratio is a subjective characterization; however, comparing a design that has 80% power to

detect an odds ratio of 1.3 to a design that has 80% power to detect an odds ratio of 1.5 does not intuitively seem to suggest that one design is **far** superior to the other (especially if other considerations, such as cost, are taken into account).

- As expected, increases in the prevalence of the disease, and/or prevalence of the health outcome, result in increased power (or decreases in the odds ratio detectable with 80% power). In general, for rare diseases, such as autism, cerebral palsy, and rare birth outcomes, only stronger relationships (e.g., odds ratios greater than 1.5) are detectable with sufficient power. However, for the more common outcomes, such as asthma and injury, weaker relationships (e.g., odds ratios around 1.2 or 1.3) are detectable. This may be an important characteristic for diseases such as asthma, where there are likely a large number of potential risk factors, each having a weak relationship to the development of asthma. In addition, from a public health perspective, being able to detect these weak relationships for diseases that are more common is likely a desirable characteristic for the NCS, since improving public awareness of the potential risk factors for common diseases has the potential to result in improving the health of a larger number of individuals. On the other hand, for rare diseases (for which there may be limited information on potential risk factors) it seems most appropriate to begin by identifying those risk factors that are most related to the development of the disease.

These general conclusions offer very suitable considerations for the design of the NCS and to the specific scenarios investigated in Section 9.4. However, it is important to note that they are general conclusions and can offer reasonable results only under the assumptions used in generating them. In other words, it is likely the case that some of these conclusions rely heavily on the assumptions that were used in obtaining the design, and in simulating the data according to a selected hypothesis. For example, other prevalence estimates and other intraclass correlations could provide different power results, as could alternative simulation methods. Additionally, other types of analyses (e.g., survival analyses or relationships between binary outcomes and continuous risk factors) could provide different power results. Thus, consideration of the validity of the assumptions that were utilized, and consideration of the limited number of scenarios investigated, must be an element of interpreting these results.

One key assumption in the power studies is that of a constant odds ratio across clusters (the conditional odds ratio described in Section D-7 of Appendix D). This is likely limiting the effect of clustering in detecting the relationships of interest, and has the interpretation that the true relationship between disease and exposure is the same within every cluster. In reality, a univariate model that expresses the probability of disease as a function of a single measure of exposure may not be valid, since there are likely factors that vary between (and within) clusters, such as race or socioeconomic status, that will act to modify the effect of the exposure of interest. In some cases, the univariate model is of interest (e.g., for setting national policy), and the use of a weighted analysis to generalize to a broader population may be important in guarding against potential biases from the effects that are excluded from the model (or from the sample). In other cases, a more complex model is of interest in which more careful scrutiny of the causal relationships between exposure and disease is the primary objective. In this scenario, where the appropriate factors (e.g., confounders, covariates, effect modifiers, etc.) are all

incorporated into the model, we believe that the assumption of constant odds ratios across clusters is appropriate since any cluster differences will be represented in these other factors. Many of these effect modifiers and other important factors will be measured in the NCS so that they can be controlled for in the modeling process. However, these more complex models were not explored in our power studies. It should be noted that incorporating effect modifiers into a model (e.g., estimating a different odds ratio for different groups of people), will likely reduce the power to detect odds ratios of a specified size within each group, because of a reduction in the sample size to estimate each group-specific odds ratio. Further research that investigates the power to detect specified odds ratios in the presence of effect modifiers may be a promising avenue for future study in the design of the NCS; however, in terms of comparing the different types of designs considered here, we envision similar results would hold.

The power to detect relationships between a health outcome and a continuous measure of exposure (e.g., biomarkers of pesticide concentrations) is also important. The power calculations above have considered evaluation of relationships only between binary exposure and outcome variables. Thus, further work on the power to detect relationships between a binary (or continuous) health outcome and a continuous risk factor may be an important avenue for future research in the design of the NCS. Nonetheless, in terms of comparison of the design approaches outlined in this paper, we envision that the results would be qualitatively similar (i.e., the more "optimal" designs in terms of a binary-binary relationships would likely remain "optimal" in terms of a binary-continuous relationships).

The large number of factors affecting the calculation of power (see above), each of them different for different hypotheses and different designs, makes power studies for the NCS as a whole relatively difficult. For this reason, in our limited examples we have generally focused on specific scenarios in an attempt to demonstrate reasonable approaches to the calculation of power, and to identify general conclusions in terms of what types of designs are most "optimal." This is not to say that we feel as though the design question is answered. Rather, the results presented here are meant to elucidate some of the difficulties in designing the NCS, and to illustrate the complexity of this issue.

10 RESULTS, CONCLUSIONS, AND RECOMMENDATIONS FOR FUTURE WORK

In the following sections we provide an overview of results and recommendations for future work relevant to the sampling design for the National Children's Study. Specifically, Section 10.1 provides a discussion of the results with a particular focus on the sample sizes and (model-based and weighted) power estimates associated with the different design options when assessing NCS hypotheses at different stages of life. Section 10.2 provides a discussion of the potential impact that various assumptions may have on the technical results that will be used to assess the performance of the different design options. Section 10.3 provides an assessment of how the different sampling design options meet the various different goals of the NCS (NCS goals are discussed in Chapter 1 and Appendix B1). Section 10.4 provides some overall conclusions, with Section 10.5 providing recommendations for future work that would benefit the choice of a final sampling design for the NCS.

10.1 DISCUSSION OF RESULTS

In this report we have considered and investigated the use of a family of designs in conducting the NCS. In very broad terms, this family of designs calls for selecting a portion of the cohort in a national probability-based sample (NPBS), and the remaining portion of the cohort using a Centers of Excellence approach. The rationale is that by combining these approaches to selecting the cohort, we can capitalize on the strengths associated with each approach while attempting to minimize their individual weaknesses. Thus, the family of designs is initiated by selecting a value for P_1 (i.e., the portion of the cohort selected through the NPBS approach). Once this fraction is selected, the NPBS portion of the cohort is selected in a two-stage clustered design where counties are the primary sampling units, and households are sampled within counties to identify women of child-bearing age. Note that other sampling frames can be considered for recruiting study participants within selected PSUs, such as a physicians office frame in rural areas as discussed in Chapter 3. The Centers portion of the cohort is further split into three components, where a fraction (P_2) is recruited from a probability-based sample of areas in proximity to the Centers, another fraction (P_3) is recruited from a probability-based sample of Center patients, and the remaining fraction (1- (P_2+P_3)) is recruited through a convenience or opportunity sample. Chapter 3 of this report provides a more detailed description of this family of designs and further discusses the rationale for its use.

Within this family of designs, there remain a large number of design possibilities. For example, what fraction of the cohort will be selected in the NPBS, how many PSUs will be utilized, and what fraction of the Centers cohort should be selected using probability-based sampling of the area in proximity to the Centers? By specifying answers to these questions, candidate designs can be identified for more careful study of their corresponding characteristics. In Chapter 5 of this report we outline the steps necessary in conducting the NPBS and Centers sampling approaches, and we specify a set of designs that are to be considered further. In particular, we allow P_1 to take values of 0.25, 0.50, and 0.75 (i.e., allowing 25, 50, and 75 percent of the cohort to be selected in a NPBS approach), we also allow P_2 to take values of 0.25, 0.50, 0.75, and we allow the number of PSUs utilized in the NPBS to take values of 50 and 100, giving us a total of 18 designs. (Note that the family of designs outlined above also allows a

portion of the Centers cohort to be selected as a volunteer sample. For the cost analysis we assumed between 1 and 4 percent of the Centers cohort would be selected as a volunteer sample, and for the power calculations for these designs we concentrate more directly on the probability-based sampling aspects of the family of designs and assume that the entire cohort will be selected in some probabilistic manner. See Chapter 9 for further discussion of this issue.)

In addition to these 18 designs, we consider five other designs where the value of P_1 is set to 0 (so that the entire cohort is selected through a set of purposively selected Centers), and the value of P_2 is set to 0 for two of the designs, and 0.25, 0.50, or 0.75 for the other three designs. To demonstrate the impact of allowing a large portion of the cohort to be study volunteers, one of the designs for which P_2 is set to 0 calls for a 50% volunteer sample. As in the other 18 designs described above, the remaining four designs concentrate on the probability-based sampling aspects and assume that the entire cohort is selected in some probabilistic manner from the candidate sampling frames (i.e., from Center patients or the Center geographic area). It should be noted that these additional five designs correspond to designs that limit the sampling frame population to only the population associated with Centers. [In particular, for the two designs with P_2 set to 0, the sampling frame population consists of patients of the set of purposively selected Centers, and for the other three designs (i.e., with P_2 greater than 0) the sampling frame population consists of individuals living in the geographic area and/or individuals that are patients of the set of purposively selected Centers.] In other words, the population represented by these designs may be significantly smaller than the population associated with the other 18 designs that include some portion of the cohort selected in a national probability-based sample (see Chapter 9 or Appendix A for a rationale as to why this may be a design approach worth consideration). Thus, direct comparison of the designs that involve some form of NPBS sampling to these designs must take into consideration the fact that the sampling frame populations may be fundamentally different.

Based on these design specifications, a total of 23 designs are considered in the cost and power analyses of Chapters 8 and 9. It should be noted that these specific designs are selected in order 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, and we do not assume that they include the "optimal" NCS design (see Section 10.2 for further discussion).

A final design characteristic that is needed in order to estimate costs and conduct power analyses for selected hypotheses is the retention rate (i.e., the percentage of the original cohort that continues to participate in the study over time) associated with a given design. Retention rates have an effect on cost estimates since the number of children remaining in the study highly influences the costs of data collection. For power calculations, retention rates are important, especially when evaluating hypotheses that can be tested only after health effects are assessed in later stages of life. Chapter 7 of this report provides further details of the retention rate issue, describes retention rates seen in other longitudinal studies, and outlines the retention rate assumptions that are utilized in the cost estimates and power analyses. In general, the retention rates assume that individuals recruited in the NPBS have the lowest retention rates, individuals recruited in the Centers area sample have slightly higher retention rates than those in the NPBS, individuals recruited through the Centers patient list have significantly higher retention rates, and

individuals that volunteer for the study have the highest likelihood of remaining in the study. See Chapter 7 for a more detailed description of these assumptions.

Using the retention rate assumptions and assuming that the initial sample size for all designs is 100,000, Figure 10-1 displays several graphs of the estimated cohort sample size as a function of life stage (i.e., the number of subjects that remain in the study as a function of their age). The top left panel of the figure depicts the sample size associated with designs for which P_2 (Center-area PBS) is set to 0.50, and P_1 (National PBS) takes values of 0.0, 0.25, 0.50, and 0.75 corresponding to the different lines in the graph (note that for the NPBS we assume the number of PSUs is 100 in these examples). The other panels of the figure correspond to fixing the value of P_1, and allowing P_2 to take values of 0.25, 0.50, 0.75, and 0.0 when P_1=0 corresponding to the different lines in the graph. (Note that in constructing these plots we have not included the design that includes some volunteer subjects.) As suggested in the figure, over the course of the study we envision that a significant portion of the cohort may be lost to follow-up, with the worst case scenario (bottom right panel) resulting in only around 30000 individuals remaining in the study through completion. Additionally, note that the values of P_1 and P_2 (and the assumptions about retention rates associated with different modes of sampling) can play a significant role in the number of individuals remaining in the study, especially as the study progresses. This characteristic will likely have cost implications and will have power implications for those hypotheses related to outcomes (and/or exposures) that are assessed during the later stages of the study (e.g., schizophrenia).

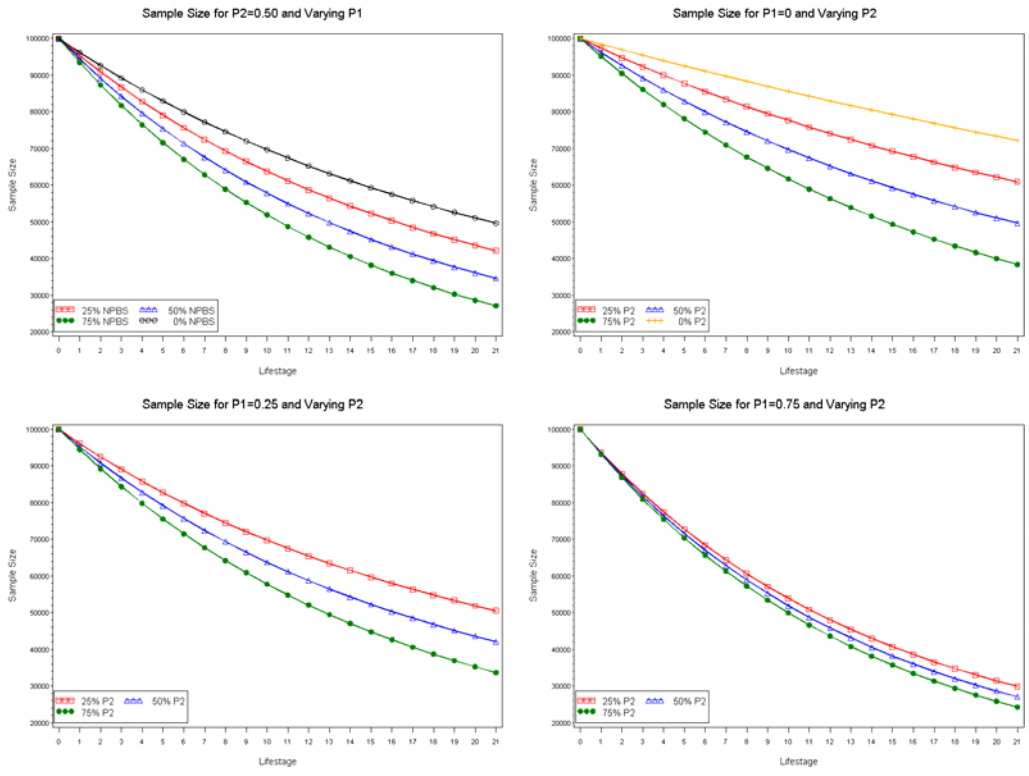

Figure 10-1. Estimated sample size as a function of life stage for different designs within the family of designs (assuming initial sample size of 100000).

Using the selected design specifications and the initial assumptions regarding retention rates (detailed in Chapter 7) and implementation costs (detailed in Chapter 8), power calculations were conducted in order to evaluate any important differences within the family of designs. In Chapter 8 of this report we focus on the issue of estimating costs for the study and we generally specify two approaches to estimating costs. The first approach assumes that each of the designs will initially recruit 100,000 subjects to participate in the study. Based on this assumption, implementation of the different designs will produce different cost estimates (i.e., some designs may be more expensive and other designs less expensive). The second approach assumes that the study has a fixed cost of approximately $2.7 billion, and thus the number of subjects that can be recruited and followed will depend on the costs associated with each design. In other words, designs that are generally less expensive (taking all costs into consideration) will have the ability to recruit a larger number of initial participants, whereas designs that are more expensive will necessarily recruit/follow fewer participants. However, it is generally the case that in the cost-constrained approach, the designs that were able to initially recruit a larger number of individuals at the beginning of the study had proportionally fewer study participants at the end of the study, again reflecting a potential tradeoff that would need to be considered if the NCS were designed to meet a fixed cost constraint.

Tables 10-1 and 10-2 display the estimated costs and initial sample sizes for the 23 designs considered under these two costing scenarios. In particular, the costs displayed in Table 10-1 are estimated under the assumption that all designs recruit an initial cohort of 100,000 individuals. Under this assumption, each of the designs has different total costs ranging from approximately $2.6B up to $3.7B. Not surprisingly, the 50 PSU designs are significantly less costly than their 100 PSU counterparts (i.e., for the same values of P_1 and P_2, the 100 PSU design is more expensive). Additionally, note that the designs that select a larger portion of the cohort in the national probability sample are generally less expensive than the other designs. While this may seem somewhat counterintuitive (e.g., under the assumption that recruitment for the national probability sampling is more expensive), these designs are less expensive primarily due to their lower assumed retention rates and consequentially their reduced cost of data collection. This same characteristic can be seen by fixing the proportion of the cohort selected in the NPBS and evaluating the cost of the study as the proportion of the cohort selected in the Center area sample increases. Again, the designs that have lower retention rates turn out to have lower costs, and generally reduced power for the model-based analyses.

On the other hand, the initial sample sizes displayed in Table 10-2 are estimated under the assumption that there is a total fixed cost for the study of $2.7 billion. Under this assumption, the 23 designs have different sample sizes for the number of live births in the NCS cohort. Note that the designs that select the largest portion of the cohort in the NPBS have the largest initial sample size. As above, these designs offer a lower cost due to their lower retention rates, and the resulting cost savings realized since fewer people are followed for the entire study period. Similar to Figure 10-1, Figure 10-2 displays several graphs of the estimated cohort sample size as a function of life stage assuming that there is a total fixed cost for the study of $2.7 billion. Note that the designs that begin with a larger number of people tend to end with fewer people (due to their lower retention rates). This highlights just one of the competing objectives at play in the NCS (if we assume a fixed cost). Large initial sample sizes (affordable

due to estimated lower retention rates and the corresponding estimated decrease in costs) may provide sufficient power for assessing hypotheses associated with diseases (and exposures) that are diagnosed early in life, but may be less optimal in terms of assessing hypotheses for diseases (and exposures) that are diagnosed later in the study (e.g., schizophrenia).

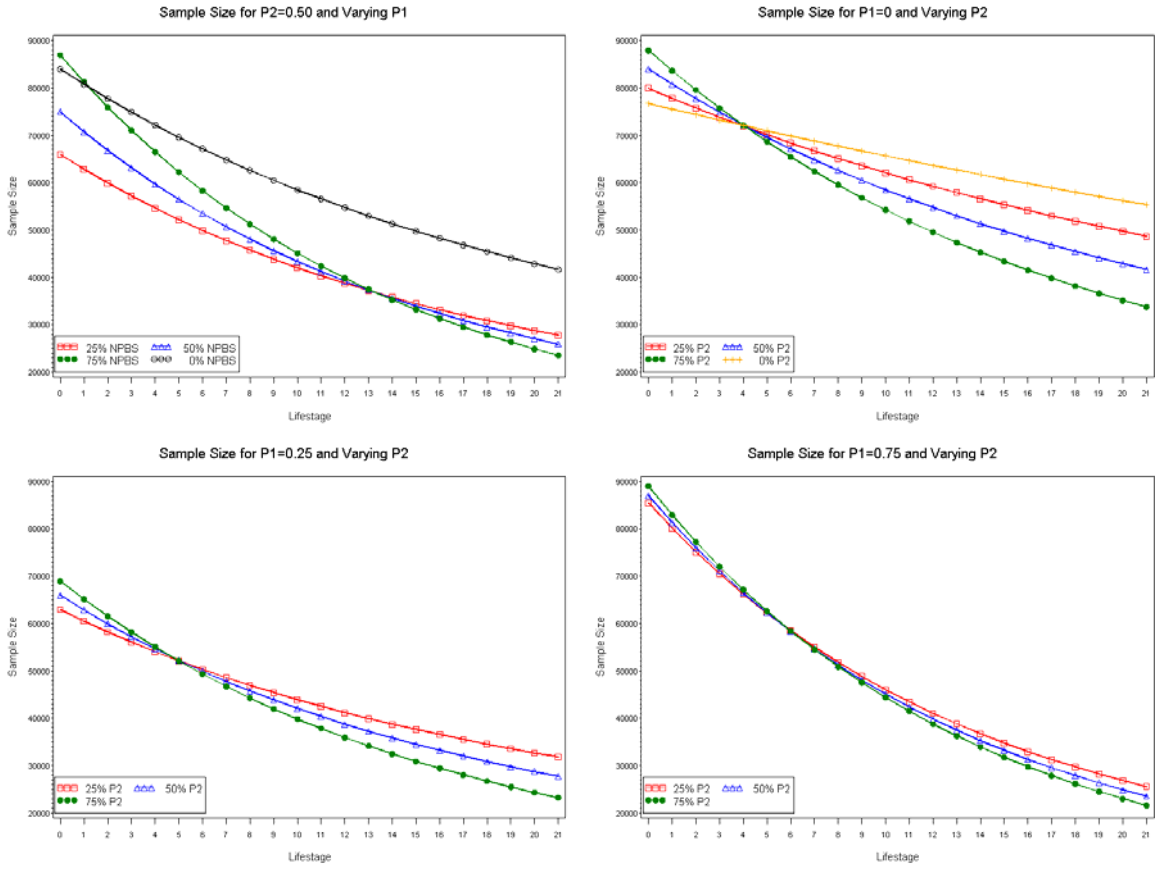

Figure 10-2. Estimated sample size as a function of life stage for different designs within the family of designs (assuming total study cost of $2.7B and 100 PSUs).

In addition to estimating the costs associated with each of the 23 designs, we also calculate the power of each design to detect relationships of interest. As mentioned in Section 9, for a study like the NCS, with multiple hypotheses and multiple inferences of interest, there are many ways to assess power (e.g., different statistical tests, alternative models, different inference goals), and there are many factors that influence the calculation of power (e.g., prevalence of the outcome, strength of the exposure/outcome relationship, etc.). Thus, the power calculations presented in this report (see Chapter 9) focus on a number of scenarios that were motivated by the core hypotheses of the study. In particular, a total of nine scenarios were investigated, and the power to detect the relationship of interest was calculated (via simulation) for varying degrees of the strength of that relationship (see Table 10-3 for the nine scenarios). These power calculations are adjusted for the possible clustering in the data resulting from the different designs, and were performed for both a model-based (i.e., unweighted) analysis and a weighted analysis.

Thus, for each of the 23 designs, Tables 10-1 and 10-2 also display a subset of the power results that were presented in Chapter 9 (Table 10-1 for the fixed initial sample size approach and Table 10-2 for the fixed cost approach). For each of the nine hypotheses, the table contains the power to detect an odds ratio of a specified size for both an unweighted (i.e., model-based) and weighted analysis. As discussed in Section 9, the odds ratio for each hypothesis was determined by inspecting the results and determining what odds ratio provides around 80 percent power in a weighted analysis. Note the following general conclusions (see Chapter 9):

- For unweighted (model-based) analyses, it is generally the case that the design that provides the largest available sample size at that life-stage corresponds to the design with the highest power.
- For the fixed sample size designs, there is relatively little difference in power from the weighted analyses when using a 50 PSU design versus a 100 PSU design (assuming the same proportion of the cohort is selected in the NPBS), suggesting that there is only a small effect of clustering for PSU sizes on the order of 100. The differences are even more subtle for unweighted (model-based) power results when comparing 50 and 100 PSU options for the national probability-based sample component. This agrees with some of our preliminary results on design effects for estimation of relationships (see Chapter 5), where it may be the case that clustering of the data has relatively low impact on estimating relationships of interest.
- On the other hand, for the fixed cost designs, the 50 PSU design provides greater power from both unweighted (model-based) and weighted analyses than the 100 PSU design (again, assuming the same proportion of the cohort is selected in the NPBS) due to its lower costs and resulting ability to follow a larger cohort of children. However, the 50 PSU design may pose other feasibility challenges with respect to recruiting a larger number of participants – especially in rural areas.
- Comparing the power for a weighted analysis to that for an unweighted analysis, many of the designs indicate a larger effect of unequal weighting (at least larger than the effect of clustering) in the cohort with designs that have a larger portion of the cohort selected in the NPBS having generally higher power when comparing designs that select some portion of the cohort in the NPBS (as expected).
- For less common outcomes, for outcomes assessed later in life, and for less common exposures (e.g., hypotheses 1.1a, 1.1b, 2.2, and 4.1), only stronger exposure/outcome relationships (i.e., only larger odds ratios) are detectable with sufficient power.
- In general, for the unweighted (model-based) analyses, the width of the range of power estimates associated with the different design options is proportional to the width of the range of sample sizes depicted in Figures 10-1 (fixed sample size designs) and 10-2 (fixed cost designs) at the time the health outcome is observed. Thus for the fixed sample size designs (constrained to have an initial sample size of 100,000 children), there is little difference among the different design options in unweighted (model-based) power at the beginning of the study – but larger differences in power among the designs later in the study (as a function of sample size). The fixed cost designs have a wider range of unweighted (model-based)

power estimates at the beginning of the NCS due to different assumed sample sizes – with the range narrowing somewhere during the middle years of childhood (when the sample sizes are somewhat convergent) – and the range widening again toward adolescence and early adulthood when the sample sizes diverge because of differences in retention.

- In general, for the weighted analyses, power is defined by the fraction of the sample that is included in probability-based samples of relatively unrestricted populations (e.g. $P_1 + P_2*(1-P_1)$).

Table 10-1. Summary table of cost estimates and power to detect relationships of interest for different hypotheses (see Table 10-3) and for the 23 design scenarios assuming a fixed sample size of 100,000.

Design	Number of PSUs	% of Cohort in NPBS	% of Cohort in C-Area	% of Cohort in C-Patients	Cost (millions)	Initial Sample Size	Hypothesis 1.1a OR = 1.75		Hypothesis 1.1b OR = 2.0		Hypothesis 2.2 OR = 1.75		Hypothesis 2.3 OR = 1.5		Hypothesis 3.1 OR = 1.2		Hypothesis 3.2 OR = 1.2		Hypothesis 4.1 OR = 1.75		Hypothesis 4.2 OR = 1.35		Hypothesis 5.1 OR = 1.2	
							Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.
A1	0	0.00	0.00	1.00	$3,473.4	100000	0.881	0.866	0.969	0.936	0.949	0.890	0.992	0.977	0.873	0.797	0.986	0.967	0.981	0.935	0.986	0.953	0.956	0.942
A2	0	0.00	0.00	0.50	$3,449.1	100000	0.919	0.604	0.967	0.765	0.948	0.643	0.991	0.823	0.866	0.588	0.989	0.830	0.976	0.830	0.988	0.831	0.950	0.752
B3	0	0.00	0.25	0.75	$3,325.1	100000	0.896	0.436	0.929	0.490	0.956	0.437	0.979	0.546	0.822	0.354	0.979	0.531	0.963	0.531	0.979	0.547	0.944	0.407
B4	0	0.00	0.50	0.50	$3,177.4	100000	0.910	0.615	0.890	0.635	0.938	0.675	0.970	0.785	0.820	0.498	0.983	0.746	0.948	0.718	0.979	0.728	0.930	0.643
B5	0	0.00	0.75	0.25	$3,029.7	100000	0.853	0.699	0.844	0.684	0.881	0.668	0.924	0.801	0.801	0.519	0.964	0.801	0.942	0.793	0.970	0.805	0.913	0.729
C6	50	0.25	0.19	0.56	$3,333.1	100000	0.917	0.606	0.860	0.433	0.916	0.582	0.961	0.560	0.815	0.458	0.979	0.702	0.952	0.729	0.976	0.691	0.935	0.579
C7	50	0.25	0.38	0.38	$3,222.2	100000	0.928	0.647	0.853	0.521	0.917	0.623	0.927	0.591	0.821	0.452	0.974	0.725	0.952	0.710	0.972	0.741	0.915	0.620
C8	50	0.25	0.56	0.19	$3,111.3	100000	0.908	0.642	0.819	0.515	0.904	0.599	0.895	0.552	0.771	0.484	0.969	0.724	0.931	0.669	0.969	0.734	0.895	0.605
D9	50	0.50	0.13	0.38	$2,982.8	100000	0.922	0.780	0.847	0.567	0.903	0.711	0.937	0.666	0.752	0.539	0.957	0.822	0.934	0.800	0.955	0.800	0.894	0.734
D10	50	0.50	0.25	0.25	$2,908.9	100000	0.902	0.783	0.789	0.636	0.903	0.779	0.901	0.719	0.779	0.596	0.966	0.852	0.923	0.834	0.963	0.874	0.878	0.734
D11	50	0.50	0.38	0.13	$2,835.0	100000	0.901	0.832	0.765	0.617	0.894	0.770	0.837	0.720	0.749	0.630	0.944	0.848	0.932	0.861	0.943	0.866	0.870	0.800
E12	50	0.75	0.06	0.19	$2,648.7	100000	0.883	0.755	0.755	0.625	0.912	0.753	0.833	0.669	0.728	0.594	0.933	0.848	0.893	0.793	0.919	0.849	0.840	0.726
E13	50	0.75	0.13	0.13	$2,611.6	100000	0.882	0.848	0.748	0.609	0.874	0.790	0.794	0.689	0.725	0.644	0.935	0.896	0.917	0.831	0.931	0.880	0.875	0.783
E14	50	0.75	0.19	0.06	$2,574.6	100000	0.853	0.876	0.695	0.618	0.880	0.789	0.789	0.726	0.710	0.615	0.922	0.899	0.906	0.847	0.932	0.908	0.839	0.798
F15	100	0.25	0.19	0.56	$3,675.2	100000	0.901	0.622	0.905	0.468	0.939	0.546	0.974	0.520	0.825	0.455	0.967	0.705	0.947	0.664	0.976	0.727	0.929	0.559
F16	100	0.25	0.38	0.38	$3,564.3	100000	0.935	0.637	0.879	0.468	0.921	0.536	0.944	0.546	0.811	0.469	0.973	0.721	0.956	0.667	0.969	0.711	0.925	0.593
F17	100	0.25	0.56	0.19	$3,453.4	100000	0.917	0.649	0.806	0.474	0.902	0.561	0.899	0.529	0.765	0.445	0.959	0.717	0.946	0.659	0.963	0.724	0.891	0.575
G18	100	0.50	0.13	0.38	$3,324.9	100000	0.926	0.771	0.847	0.604	0.901	0.730	0.907	0.701	0.812	0.608	0.964	0.872	0.942	0.838	0.961	0.853	0.910	0.720
G19	100	0.50	0.25	0.25	$3,251.0	100000	0.923	0.839	0.813	0.591	0.883	0.738	0.879	0.698	0.758	0.617	0.965	0.886	0.942	0.832	0.962	0.884	0.908	0.781
G20	100	0.50	0.38	0.13	$3,177.2	100000	0.921	0.828	0.763	0.642	0.880	0.730	0.850	0.693	0.764	0.611	0.960	0.876	0.935	0.847	0.961	0.868	0.890	0.763
H21	100	0.75	0.06	0.19	$2,990.8	100000	0.926	0.859	0.786	0.627	0.879	0.759	0.858	0.690	0.749	0.654	0.954	0.886	0.913	0.845	0.951	0.907	0.881	0.770
H22	100	0.75	0.13	0.13	$2,953.7	100000	0.914	0.870	0.720	0.643	0.880	0.842	0.832	0.737	0.747	0.667	0.945	0.918	0.928	0.880	0.957	0.910	0.882	0.836
H23	100	0.75	0.19	0.06	$2,916.7	100000	0.915	0.851	0.688	0.608	0.856	0.836	0.801	0.700	0.753	0.658	0.945	0.920	0.908	0.883	0.938	0.915	0.864	0.826

Table 10-2. Summary table of cost estimates and power to detect relationships of interest for different hypotheses (see Table 10-3) and for the 23 design scenarios assuming fixed total costs of $2.7B.

Design	Number of PSUs	% of Cohort in NPBS	% of Cohort in C-Area	% of Cohort in C-Patients	Cost (millions)	Initial Sample Size	Hypothesis 1.1a OR = 1.75		Hypothesis 1.1b OR = 2.0		Hypothesis 2.2 OR = 1.75		Hypothesis 2.3 OR = 1.5		Hypothesis 3.1 OR = 1.2		Hypothesis 3.2 OR = 1.2		Hypothesis 4.1 OR = 1.75		Hypothesis 4.2 OR = 1.35		Hypothesis 5.1 OR = 1.2	
							Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.
A1	0	0.00	0.00	1.00	$2,746.9	76750	0.828	0.774	0.894	0.866	0.882	0.865	0.974	0.944	0.729	0.722	0.954	0.904	0.939	0.895	0.960	0.916	0.892	0.840
A2	0	0.00	0.00	0.50	$2,705.8	75750	0.804	0.578	0.900	0.590	0.875	0.538	0.978	0.762	0.756	0.491	0.963	0.751	0.938	0.726	0.966	0.728	0.890	0.574
B3	0	0.00	0.25	0.75	$2,735.0	80000	0.828	0.364	0.887	0.379	0.888	0.380	0.956	0.453	0.757	0.252	0.948	0.509	0.929	0.493	0.950	0.541	0.872	0.391
B4	0	0.00	0.50	0.50	$2,728.5	84000	0.848	0.610	0.872	0.649	0.895	0.530	0.937	0.669	0.770	0.459	0.954	0.634	0.922	0.720	0.955	0.699	0.894	0.571
B5	0	0.00	0.75	0.25	$2,710.3	88000	0.843	0.669	0.826	0.595	0.900	0.685	0.886	0.721	0.744	0.534	0.946	0.800	0.903	0.768	0.942	0.799	0.851	0.659
C6	50	0.25	0.19	0.56	$2,728.4	77000	0.842	0.465	0.850	0.358	0.855	0.458	0.919	0.421	0.685	0.368	0.940	0.580	0.891	0.584	0.928	0.583	0.854	0.465
C7	50	0.25	0.38	0.38	$2,715.7	80000	0.851	0.524	0.771	0.407	0.864	0.474	0.896	0.465	0.743	0.386	0.934	0.632	0.928	0.580	0.935	0.619	0.851	0.506
C8	50	0.25	0.56	0.19	$2,705.3	83500	0.863	0.560	0.732	0.417	0.857	0.519	0.847	0.459	0.695	0.401	0.927	0.649	0.909	0.635	0.920	0.671	0.843	0.525
D9	50	0.50	0.13	0.38	$2,719.3	88500	0.870	0.695	0.842	0.473	0.868	0.687	0.885	0.592	0.744	0.535	0.937	0.789	0.907	0.773	0.942	0.792	0.840	0.646
D10	50	0.50	0.25	0.25	$2,714.5	91000	0.866	0.716	0.771	0.613	0.893	0.761	0.871	0.680	0.749	0.560	0.944	0.808	0.926	0.773	0.943	0.822	0.878	0.674
D11	50	0.50	0.38	0.13	$2,714.4	94000	0.893	0.771	0.714	0.561	0.842	0.731	0.803	0.686	0.732	0.582	0.935	0.834	0.892	0.774	0.933	0.841	0.877	0.737
E12	50	0.75	0.06	0.19	$2,713.8	104000	0.889	0.819	0.777	0.621	0.885	0.758	0.838	0.715	0.733	0.619	0.945	0.836	0.906	0.791	0.945	0.848	0.865	0.752
E13	50	0.75	0.13	0.13	$2,708.9	106000	0.876	0.863	0.753	0.624	0.889	0.848	0.875	0.724	0.734	0.655	0.938	0.900	0.907	0.866	0.937	0.896	0.874	0.799
E14	50	0.75	0.19	0.06	$2,694.4	107500	0.912	0.856	0.708	0.665	0.870	0.832	0.812	0.782	0.714	0.702	0.937	0.889	0.925	0.861	0.927	0.876	0.860	0.798
F15	100	0.25	0.19	0.56	$2,712.8	63000	0.756	0.398	0.713	0.274	0.773	0.320	0.872	0.329	0.656	0.291	0.885	0.466	0.870	0.404	0.893	0.479	0.771	0.371
F16	100	0.25	0.38	0.38	$2,715.5	66000	0.762	0.461	0.714	0.313	0.800	0.359	0.836	0.355	0.644	0.299	0.874	0.494	0.820	0.482	0.901	0.488	0.776	0.388
F17	100	0.25	0.56	0.19	$2,711.6	69000	0.801	0.455	0.638	0.332	0.746	0.382	0.774	0.357	0.648	0.312	0.884	0.522	0.834	0.497	0.896	0.510	0.758	0.420
G18	100	0.50	0.13	0.38	$2,711.8	73000	0.852	0.664	0.751	0.463	0.815	0.571	0.820	0.524	0.649	0.450	0.891	0.744	0.882	0.682	0.889	0.735	0.768	0.625
G19	100	0.50	0.25	0.25	$2,709.6	75000	0.848	0.705	0.692	0.477	0.798	0.613	0.801	0.566	0.650	0.497	0.912	0.746	0.871	0.729	0.910	0.768	0.805	0.638
G20	100	0.50	0.38	0.13	$2,697.0	77500	0.841	0.666	0.607	0.433	0.789	0.606	0.738	0.578	0.655	0.515	0.878	0.752	0.866	0.732	0.900	0.756	0.795	0.636
H21	100	0.75	0.06	0.19	$2,709.3	85500	0.864	0.798	0.707	0.555	0.796	0.777	0.780	0.643	0.678	0.595	0.911	0.858	0.903	0.852	0.909	0.859	0.835	0.744
H22	100	0.75	0.13	0.13	$2,702.8	87000	0.899	0.847	0.687	0.578	0.823	0.787	0.769	0.705	0.709	0.624	0.920	0.874	0.886	0.824	0.921	0.883	0.831	0.778
H23	100	0.75	0.19	0.06	$2,703.4	89000	0.871	0.830	0.650	0.591	0.844	0.765	0.707	0.660	0.691	0.630	0.911	0.882	0.894	0.857	0.909	0.871	0.823	0.743

Table 10-3. Nine scenarios investigated.

Hypothesis	Health Outcome (prevalence)	Risk Factor (prevalence)	Years of Follow-Up
1.1a	Central Nervous System Defects (0.60%)	Impaired Glucose Metabolism During Pregnancy (5%)	1
1.1b	Malformations of the Heart (0.60%)	Impaired Glucose Metabolism During Pregnancy (5%)	19
2.2	Cerebral Palsy and Autism (0.25%)	Prenatal Infection and Mediators of Inflammation (20%)	7
2.3	Schizophrenia (1%)	Pre/Perinatal Infection and Mediators of Inflammation (20%)	21
3.1	Increased Risk of Injury (10%)	Exposure to Neurotoxins or Behavioral Attributes of Childcare (5%)	10
3.2	Increased Risk of Injury (10%)	Exposure to Neurotoxins or Behavioral Attributes of Childcare (10%)	10
4.1	Development of Asthma (6%)	Respiratory Viral Infection or Maternal Stress During Pregnancy or Exposure to Air Pollution (1%)	10
4.2	Development of Asthma (6%)	Respiratory Viral Infection or Maternal Stress During Pregnancy or Exposure to Air Pollution (5%)	10
5.1	Obesity (15%)	Impaired Glucose Metabolism During Pregnancy (5%)	10

As a summary of the 23 designs over all nine hypotheses, Figures 10-3 and 10-4 display the average weighted and unweighted power (using the selected odds ratios) for each design. In particular, Figure 10-3 displays the power assuming that an initial sample size of 100,000 subjects is obtained for all of the designs. The top left panel displays the average power over all nine hypotheses for each of the 23 designs, the top right panel displays the power for the one hypothesis involving an early life stage outcome, the bottom left panel displays the average power among the six hypotheses assessed during childhood (ages 4 to 12), and the bottom right panel displays the average power among the two hypotheses assessed during adolescence or early adulthood (ages 13-21). Note that in general the designs associated with the highest power in an unweighted analysis (i.e., model-based) have the lowest power when conducting a weighted analysis. One exception to this is for the early life stage hypothesis where, because of the relatively short follow-up period, all 23 designs achieve approximately similar sample sizes, and, thus, have similar power in an unweighted analysis. Another exception to this generalization is for the design (design A1, represented by the square orange symbols) for which the sampling frame population is restricted to only current patients of the set of selected Centers . For this design, both the average weighted and average unweighted powers are relatively high.

However, recall that the designs involving no NPBS sampling (i.e., the solid black and orange symbols in the figure) correspond to a smaller sampling frame population. In other words, the population of inference for these designs is somewhat smaller than it is for the other designs. The acceptability of this sampling frame restriction must be considered when comparing the power under these designs to the power for designs involving at least some NPBS sampling. Additionally, as described in Section 5, we envision that it is possible to identify a more optimal approach to obtaining the Centers sample so that a more "self-weighting" design

could be obtained. To the degree that a more "self-weighting" design could be obtained, the weighted analysis power results would improve.

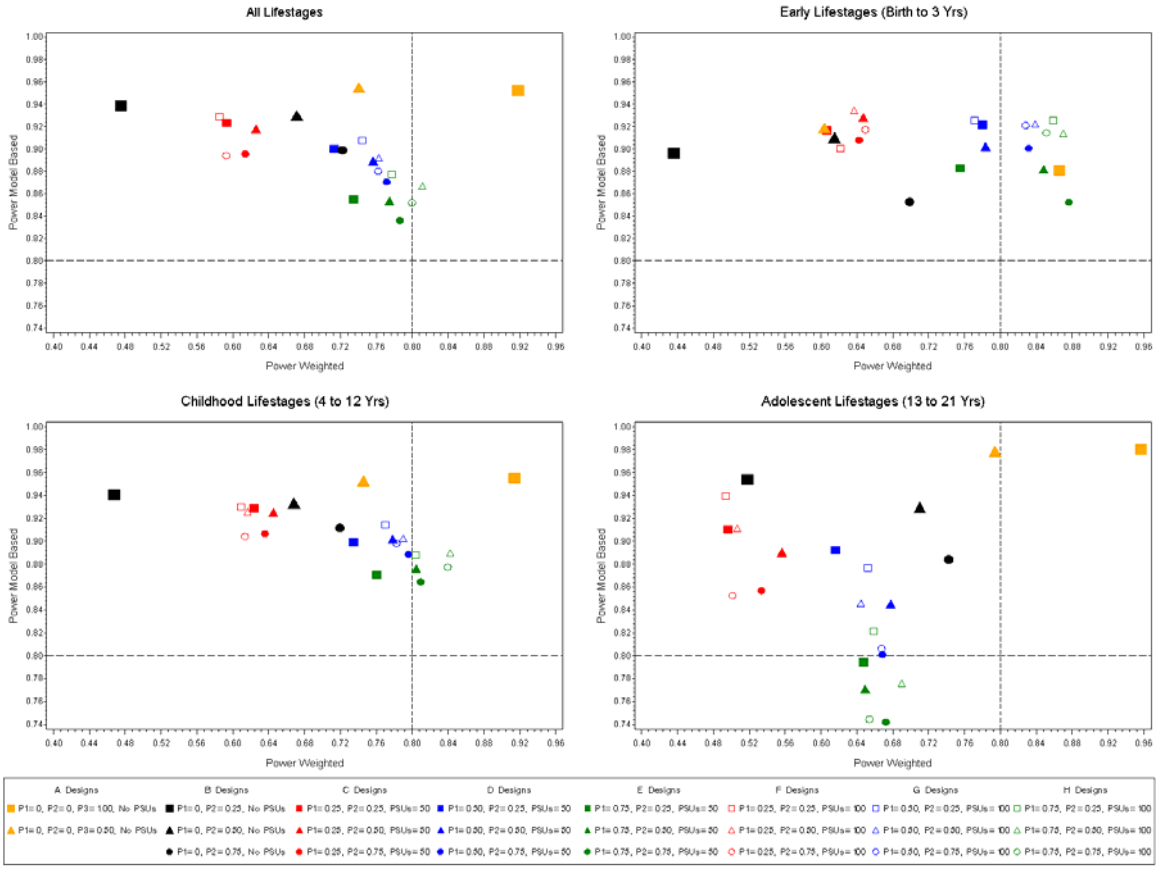

Figure 10-3. Average power (for unweighted and weighted analyses) to detect relationships of interest for hypotheses involving all life stages, early life stages, childhood life stages, adolescent life stages assuming fixed sample size of 100,000 children.

Figure 10-4, on the other hand, displays the same information but for the fixed cost approach to designing the study. Note that here we see the differences between the power for assessing the early life stage hypotheses versus the power for assessing the later life stage hypotheses. In particular, note that the designs associated with the highest power (in an unweighted analysis) for assessing the early life stage hypotheses (upper right panel) often end up being the designs associated with generally lower power for assessing the hypotheses requiring longer periods of follow-up. Thus, we again see the issue of competing objectives resulting in different design implications.

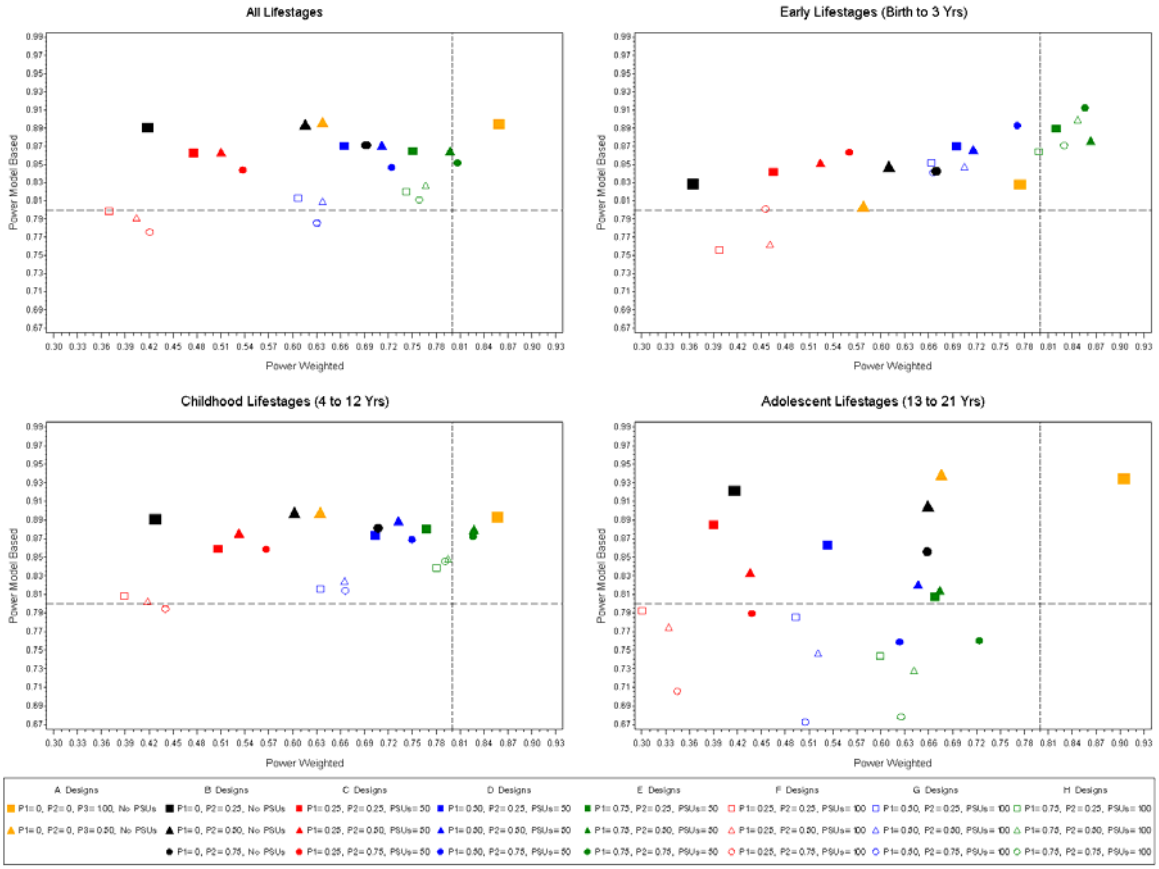

Figure 10-4. Average power (for unweighted and weighted analyses) to detect relationships of interest for hypotheses involving all life stages, early life stages, childhood life stages, adolescent life stages assuming fixed total study costs of $2.7 billion.

10.2 DISCUSSION OF THE POTENTIAL IMPACT OF ASSUMPTIONS ON THE RESULTS AND OTHER ADVANTAGES AND LIMITATIONS OF THE TECHNICAL APPROACH

In the previous section, we provide an overview of technical results corresponding to a range of design options that were developed and characterized throughout this report. However, many of these results are highly influenced by a number of critical assumptions that merit further discussion. In the following sections, we provide an open discussion related to each of these critical assumptions, as well as insight into how the results might change under different assumptions. Specifically, Section 10.2.1 provides a discussion related to the assumed recruitment and retention rates for different modes of sampling; Section 10.2.2 provides a discussion related to the cost estimates (which are also highly influenced by the assumed recruitment and retention rates); Section 10.2.3 provides a discussion of the power studies – including the selection of hypotheses and relationships that were studied and critical assumptions that were made in the power study simulations; and Section 10.2.4 provides a discussion of the parameters chosen, which define the range of design options that are presented in this report. In some cases, we would recommend sensitivity analyses or other additional statistical design work to determine the extent to which different assumptions might lead to different recommendations

on how to recruit study participants in the NCS. Specific recommendations for these sensitivity analyses are discussed further in Section 10.5.

10.2.1 Assumptions Related to Recruitment and Retention Rates

The first set of assumptions that warrant further discussion are those regarding the rates of recruitment and retention of study subjects. This topic has been the theme of many research studies and a large volume of published literature. In fact, there are a number of studies discussed in Appendix G that provide some limited insight into potential methods for recruitment and retention in the NCS, and offer some information that can be used to formulate estimates of the initial response and retention rates for the NCS. Chapter 7 describes one approach to estimating recruitment and retention rates for the NCS based on the information from these other studies. Unfortunately, as discussed in Chapter 7, these other studies do not involve the same scope, size, and design as that envisioned for the NCS, and, thus, estimating recruitment and retention rates based on these studies is problematic at best. Admittedly, it may be the case that recruitment and retention rates for the NCS will generally be higher than those observed in other studies (e.g., due to incentive programs, the important nature of the NCS, etc.), or it may be the case that recruitment and retention rates for the NCS will be lower than those observed in the other studies (e.g., due to subject burden, the length of the study, the methods of recruitment, etc.). We assumed a simple exponential decay model for retention rates experienced under different methods of recruiting study subjects into the NCS based on what was observed in historical studies, when clearly other methods are plausible. In addition, we assumed that there would be large differences in retention rates between study subjects that are recruited using probability-based sampling from relatively unrestricted populations compared to study subjects recruited using probability-based sampling from a much more restricted and convenient sampling frame or through convenience sampling.
However, due to the fact that

> (1) estimating recruitment and retention rates is not an exact science,
> (2) there is little or no information on observed recruitment and retention rates in a study similar to the NCS, and
> (3) the estimated recruitment and retention rates are drivers for the results in a number of other sections of this report (e.g., the cost estimates in Chapter 8 and the power calculations in Chapter 9),

in the following we discuss ways in which assumed recruitment and retention rates may have an impact on design decisions for the NCS and on the results presented in this report.

One important consideration with regard to recruitment, retention, and response rates is in how they affect the generalizability of probability-based samples. As discussed in Appendix A, the generalizability of probability-based samples (i.e., the possibility of basing statistical inferences on the random sampling mechanism) is compromised by less than perfect recruitment, retention, and response rates (all of which will occur in the NCS). Two primary options exist for dealing with less than perfect response rates. The first option is to perform a second near-perfect study of initial nonresponders. While less information is obtained about initial nonresponders than originally planned, this approach requires no assumptions for validity. The second option is

to assume that the subpopulation of responders is unbiased relative to the population of interest. For recruitment, retention, and response rates that are relatively high, this assumption may be sensible; however, as these rates decrease, this assumption becomes less and less reasonable. While methods such as multiple imputation can be used to adjust for potential biases introduced into the study sample from a limited amount of nonresponse while still allowing researchers to extrapolate study results back to the population of interest, these methods become far less effective when the response rates are low (due to either recruitment nonresponse or attrition). Thus, while probability-based samples offer the possibility of unbiased statistical inference with few assumptions, the reality of imperfect response rates imposes a need for either additional data on initial nonresponders or an assumption that responders are no different than nonresponders.

The estimated retention rates may also have significant impact on the cost estimates and power calculations presented in this report. In terms of the cost estimates for fixed sample size designs (i.e., where we assume that all designs recruit an initial cohort of 100,000 children), the assumed retention rates have some obvious implications, i.e., if the true retention rates are lower than the assumed retention rates, then costs will go down, and if the true retention rates are higher than the assumed retention rates, then costs will go up. Similarly, for the fixed cost designs, if the true retention rates are lower than the assumed rates then the opportunity to recruit a larger initial sample size for the study will have been missed (i.e., we probably wouldn't recognize the lower retention rate until it is too late to enroll additional participants in the study, and therefore the study may end up costing less than anticipated). Conversely, if the true retention rates are higher than the assumed rates in the cost-constrained approach, then study planners might need to reduce the amount of data collection within the larger than expected NCS cohort to remain within the fixed costs of the study. On the other hand, there may be more subtle implications when comparing the various designs if retention rates for the different methods of selection are changed. For example, if the assumed retention rates for the national probability sample are too low, then the costs associated with these designs will increase (or the allowable sample size or numbers/types of measures that can be performed for fixed costs of $2.7B will go down). Thus, comparison of the different designs in terms of their costs could be influenced by changing the retention rates.

In terms of power, ignoring the issue discussed above regarding the validity of inference in the presence of imperfect recruitment and retention rates, the general impact of differing retention rates is also easily understood. For the unweighted (model-based) analyses with fixed sample size designs (i.e., where we assume that all designs recruit an initial cohort of 100,000 children), power will increase (decrease) when any component of the NCS family of designs experiences higher (lower) rates of retention. For the unweighted (model-based) analyses with fixed cost (i.e., where we assume that all designs are constrained to the same budget of $2.7 billion), it is likely that higher retention rates would result in the necessity for reduced data collection on study subjects at later stages in the study with potential impact on the power to address hypotheses related to later life-stage priority outcomes. For the weighted analyses of study data intended to extrapolate to the reference population for the NCS (e.g., all children born in the U.S. during the NCS period of recruitment), the parameter estimates for the relationship between adverse health effects and exposure will be most highly influenced by the fraction of the cohort recruited via probability-based sampling from a relatively unrestricted target population – namely the P_1 fraction of the cohort recruited in a national probability-based sample, and the (1-

P_1)*P_2 fraction of the cohort recruited from a probability-based sample of MSAs surrounding the purposively selected Centers. Therefore, for weighted analyses with fixed sample size designs, power will increase (decrease) when either of these two components of the NCS family of designs experiences higher (lower) rates of retention. For the weighted analyses with fixed cost, it is likely that higher retention rates for either of these two components would result in reduced sample sizes and lower power in the beginning of the study and increased sample sizes and higher power later in the study.

To demonstrate the impact of assuming alternative retention rates on the cost estimates and power calculations, Tables 10-4 and 10-5 display the estimated costs, initial sample sizes, and estimated power for the 23 designs when assuming the **alternative retention rates** outlined in Chapter 7. Specifically, Table 10-4 corresponds to the case where all designs recruit an initial cohort of 100,000 individuals, and Table 10-5 corresponds to the case where total study costs are fixed at $2.7 billion, resulting in different sample sizes for the number of live births affordable under each design. The alternative rates correspond to increasing the longer-term retention rates for individuals selected in the NPBS, increasing the longer-term retention rates for individuals selected in the Center geographic area, and leaving the retention rates for Center patients and volunteer subjects unchanged (see Chapter 7 for more detailed discussion). As indicated above, for the fixed sample size scenario, the impact of these higher retention rates for designs that involve some NPBS and/or some Center area recruitment is an increase in power (especially for those hypotheses corresponding to later life stages) as well as an increase in cost. On the other hand, for the fixed cost scenario, the impact is a decrease in the initial sample size affordable under designs involving some NPBS and/or some Center area recruitment, and a decrease or an increase in the power depending on the life stage associated with a selected hypothesis. (Note that figures similar to those presented in the results of Section 9, but assuming these alternative retention rates, can be found in Appendix I.)

Figure 10-5 and 10-6 display the average weighted and unweighted power for each design under these alternative retention assumptions (similar to Figures 10-3 and 10-4). Figure 10-5 displays the power assuming an initial sample size of 100,000 subjects is obtained for all of the designs, and Figure 10-6 displays the power assuming a total fixed cost for the study of $2.7 billion. As in Figures 10-3 and 10-4, the top left panel of the figures displays the average power over all nine hypotheses, the top right panel displays the power for the one hypothesis involving an early life stage outcome, the bottom left panel displays the average power among the six hypotheses assessed during childhood (ages 4 to 12), and the bottom right panel displays the average power among the two hypotheses assessed during adolescence or early adulthood (ages 13-21).

These alternative retention rates could perhaps be considered an upper-bound for plausible retention rates associated with the NCS cohort; whereas the initial retention rates utilized in the other sections of this report may represent a lower-bound for plausible retention rates associated with the NCS. This being the case, the corresponding costs and power estimates can be considered to represent the range of costs and power associated with the different design scenarios. However, one important assumption that should be noted in these comparisons is the assumption that additional resources have not been committed to obtaining higher retention rates. In other words, it is likely the case that retention rates will in part depend on incentives, tracking

mechanisms, and other factors requiring study resources (perhaps significant resources). In these results we have not included these types of resource expenditures, but have simply assumed that all designs will incur similar costs in terms of tracking and retaining study participants. If different resource expenditures are necessary to achieve desired retention rates for different sets of individuals, then cost estimates will certainly increase or decrease accordingly.

Table 10-4. Summary table of cost estimates and power to detect relationships of interest for different hypotheses (see Table 10-3) and for the 23 design scenarios assuming a fixed sample size of 100,000 and using the alternative retention rates (see Chapter 7).

Design	Number of PSUs	% of Cohort in NPBS	% of Cohort in C-Area	% of Cohort in C-Patients	Cost (millions)	Initial Sample Size	Hypothesis 1.1a OR = 1.75		Hypothesis 1.1b OR = 2.0		Hypothesis 2.2 OR = 1.75		Hypothesis 2.3 OR = 1.5		Hypothesis 3.1 OR = 1.2		Hypothesis 3.2 OR = 1.2		Hypothesis 4.1 OR = 1.75		Hypothesis 4.2 OR = 1.35		Hypothesis 5.1 OR = 1.2	
							Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.
A1	0	0.00	0.00	1.00	$3,473.3	100000	0.881	0.866	0.969	0.936	0.949	0.890	0.992	0.977	0.873	0.797	0.986	0.967	0.981	0.935	0.986	0.953	0.956	0.942
A2	0	0.00	0.00	0.50	$3,449.2	100000	0.919	0.604	0.967	0.765	0.948	0.643	0.991	0.823	0.866	0.588	0.989	0.830	0.976	0.830	0.988	0.831	0.950	0.752
B3	0	0.00	0.25	0.75	$3,442.7	100000	0.896	0.436	0.962	0.520	0.938	0.485	0.993	0.622	0.865	0.364	0.986	0.571	0.966	0.587	0.984	0.572	0.961	0.456
B4	0	0.00	0.50	0.50	$3,412.7	100000	0.910	0.615	0.954	0.667	0.937	0.691	0.993	0.862	0.883	0.570	0.989	0.826	0.971	0.820	0.988	0.765	0.960	0.688
B5	0	0.00	0.75	0.25	$3,382.6	100000	0.853	0.699	0.954	0.843	0.947	0.716	0.990	0.909	0.874	0.632	0.985	0.856	0.977	0.831	0.984	0.884	0.955	0.741
C6	50	0.25	0.19	0.56	$3,556.7	100000	0.917	0.606	0.957	0.702	0.940	0.613	0.993	0.789	0.870	0.566	0.989	0.829	0.977	0.771	0.987	0.827	0.962	0.713
C7	50	0.25	0.38	0.38	$3,534.0	100000	0.928	0.647	0.964	0.726	0.948	0.662	0.991	0.852	0.880	0.578	0.989	0.815	0.982	0.784	0.989	0.847	0.964	0.722
C8	50	0.25	0.56	0.19	$3,511.4	100000	0.908	0.642	0.968	0.743	0.954	0.637	0.990	0.836	0.885	0.585	0.984	0.859	0.972	0.814	0.984	0.876	0.959	0.744
D9	50	0.50	0.13	0.38	$3,312.3	100000	0.922	0.780	0.945	0.849	0.925	0.798	0.988	0.927	0.864	0.735	0.984	0.922	0.962	0.877	0.985	0.929	0.941	0.839
D10	50	0.50	0.25	0.25	$3,297.3	100000	0.902	0.783	0.956	0.888	0.957	0.850	0.986	0.938	0.850	0.742	0.989	0.947	0.970	0.932	0.986	0.941	0.942	0.876
D11	50	0.50	0.38	0.13	$3,282.2	100000	0.901	0.832	0.951	0.906	0.939	0.862	0.990	0.952	0.865	0.758	0.984	0.955	0.976	0.902	0.984	0.944	0.950	0.867
E12	50	0.75	0.06	0.19	$3,084.1	100000	0.883	0.755	0.937	0.882	0.949	0.837	0.987	0.956	0.839	0.718	0.980	0.941	0.965	0.928	0.986	0.942	0.929	0.854
E13	50	0.75	0.13	0.13	$3,076.5	100000	0.882	0.848	0.935	0.909	0.942	0.870	0.987	0.961	0.839	0.737	0.983	0.945	0.963	0.938	0.977	0.950	0.939	0.892
E14	50	0.75	0.19	0.06	$3,068.9	100000	0.853	0.876	0.931	0.892	0.934	0.878	0.988	0.971	0.846	0.785	0.979	0.963	0.963	0.951	0.977	0.946	0.946	0.902
F15	100	0.25	0.19	0.56	$3,898.8	100000	0.901	0.622	0.967	0.661	0.954	0.605	0.993	0.791	0.871	0.575	0.989	0.837	0.980	0.779	0.986	0.829	0.964	0.710
F16	100	0.25	0.38	0.38	$3,876.1	100000	0.935	0.637	0.967	0.743	0.941	0.625	0.994	0.831	0.891	0.605	0.992	0.827	0.983	0.782	0.990	0.821	0.962	0.735
F17	100	0.25	0.56	0.19	$3,853.5	100000	0.917	0.649	0.962	0.720	0.931	0.657	0.990	0.831	0.882	0.573	0.989	0.844	0.974	0.809	0.989	0.824	0.944	0.735
G18	100	0.50	0.13	0.38	$3,654.4	100000	0.926	0.771	0.964	0.883	0.944	0.821	0.989	0.934	0.891	0.737	0.989	0.948	0.986	0.918	0.989	0.950	0.964	0.861
G19	100	0.50	0.25	0.25	$3,639.4	100000	0.923	0.839	0.960	0.906	0.937	0.875	0.990	0.951	0.880	0.764	0.990	0.950	0.977	0.928	0.992	0.959	0.957	0.895
G20	100	0.50	0.38	0.13	$3,624.4	100000	0.921	0.828	0.966	0.905	0.920	0.861	0.989	0.954	0.892	0.772	0.987	0.952	0.978	0.926	0.988	0.959	0.957	0.904
H21	100	0.75	0.06	0.19	$3,426.2	100000	0.926	0.859	0.962	0.910	0.933	0.874	0.990	0.976	0.881	0.791	0.988	0.971	0.980	0.949	0.987	0.964	0.949	0.912
H22	100	0.75	0.13	0.13	$3,418.6	100000	0.914	0.870	0.961	0.941	0.948	0.921	0.989	0.980	0.876	0.836	0.989	0.982	0.976	0.966	0.992	0.979	0.953	0.915
H23	100	0.75	0.19	0.06	$3,411.0	100000	0.915	0.851	0.968	0.946	0.925	0.899	0.991	0.983	0.861	0.818	0.988	0.971	0.974	0.965	0.990	0.976	0.947	0.929

Table 10-5. Summary table of cost estimates and power to detect relationships of interest for different hypotheses (see Table 10-3) and for the 23 design scenarios assuming fixed costs of $2.7B and using the alternative retention rates (see Chapter 7).

Design	Number of PSUs	% of Cohort in NPBS	% of Cohort in C-Area	% of Cohort in C-Patients	Cost (millions)	Initial Sample Size	Hypothesis 1.1a OR = 1.75		Hypothesis 1.1b OR = 2.0		Hypothesis 2.2 OR = 1.75		Hypothesis 2.3 OR = 1.5		Hypothesis 3.1 OR = 1.2		Hypothesis 3.2 OR = 1.2		Hypothesis 4.1 OR = 1.75		Hypothesis 4.2 OR = 1.35		Hypothesis 5.1 OR = 1.2	
							Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.	Unwt.	Wt.
A1	0	0.00	0.00	1.00	$2,746.9	76750	0.828	0.774	0.894	0.866	0.882	0.865	0.974	0.944	0.729	0.722	0.954	0.904	0.939	0.895	0.960	0.916	0.892	0.840
A2	0	0.00	0.00	0.50	$2,705.9	75750	0.804	0.578	0.900	0.590	0.875	0.538	0.978	0.762	0.756	0.491	0.963	0.751	0.938	0.726	0.966	0.728	0.890	0.574
B3	0	0.00	0.25	0.75	$2,745.4	77000	0.853	0.361	0.904	0.413	0.913	0.415	0.971	0.511	0.759	0.295	0.965	0.505	0.936	0.514	0.952	0.519	0.911	0.404
B4	0	0.00	0.50	0.50	$2,722.2	77000	0.823	0.586	0.928	0.645	0.898	0.572	0.970	0.765	0.800	0.538	0.963	0.760	0.946	0.723	0.971	0.803	0.891	0.662
B5	0	0.00	0.75	0.25	$2,721.0	78000	0.783	0.621	0.887	0.764	0.885	0.636	0.977	0.850	0.795	0.596	0.954	0.800	0.950	0.813	0.952	0.845	0.890	0.689
C6	50	0.25	0.19	0.56	$2,735.3	71000	0.813	0.464	0.902	0.510	0.885	0.478	0.965	0.642	0.733	0.438	0.956	0.688	0.930	0.629	0.949	0.697	0.881	0.532
C7	50	0.25	0.38	0.38	$2,719.3	71000	0.806	0.433	0.878	0.604	0.875	0.494	0.965	0.701	0.759	0.440	0.955	0.711	0.904	0.700	0.950	0.704	0.878	0.617
C8	50	0.25	0.56	0.19	$2,714.1	71500	0.766	0.499	0.865	0.594	0.882	0.519	0.958	0.686	0.762	0.479	0.945	0.710	0.937	0.692	0.934	0.715	0.865	0.578
D9	50	0.50	0.13	0.38	$2,711.8	77000	0.804	0.669	0.878	0.753	0.870	0.731	0.970	0.847	0.786	0.627	0.961	0.856	0.930	0.807	0.946	0.863	0.884	0.754
D10	50	0.50	0.25	0.25	$2,711.1	77500	0.802	0.677	0.898	0.833	0.895	0.768	0.966	0.864	0.779	0.629	0.968	0.900	0.923	0.851	0.952	0.863	0.911	0.764
D11	50	0.50	0.38	0.13	$2,704.9	77750	0.821	0.695	0.891	0.811	0.879	0.757	0.970	0.878	0.779	0.645	0.952	0.890	0.927	0.833	0.956	0.874	0.893	0.776
E12	50	0.75	0.06	0.19	$2,705.5	84000	0.851	0.762	0.921	0.851	0.855	0.786	0.976	0.928	0.760	0.681	0.961	0.907	0.947	0.876	0.951	0.923	0.886	0.817
E13	50	0.75	0.13	0.13	$2,709.8	84500	0.836	0.839	0.919	0.848	0.874	0.831	0.969	0.946	0.801	0.700	0.968	0.931	0.951	0.885	0.962	0.925	0.887	0.858
E14	50	0.75	0.19	0.06	$2,703.3	84500	0.850	0.808	0.917	0.863	0.875	0.850	0.975	0.953	0.793	0.738	0.962	0.934	0.947	0.942	0.951	0.936	0.897	0.854
F15	100	0.25	0.19	0.56	$2,721.6	58500	0.701	0.367	0.832	0.446	0.793	0.363	0.902	0.545	0.706	0.347	0.911	0.569	0.886	0.537	0.908	0.561	0.808	0.453
F16	100	0.25	0.38	0.38	$2,708.4	58500	0.781	0.388	0.832	0.419	0.794	0.374	0.938	0.557	0.676	0.350	0.914	0.575	0.869	0.545	0.913	0.590	0.796	0.462
F17	100	0.25	0.56	0.19	$2,706.1	59000	0.728	0.388	0.879	0.464	0.705	0.398	0.910	0.541	0.688	0.369	0.912	0.579	0.880	0.550	0.901	0.571	0.819	0.465
G18	100	0.50	0.13	0.38	$2,713.0	63500	0.796	0.582	0.839	0.691	0.817	0.581	0.925	0.805	0.720	0.533	0.932	0.815	0.905	0.772	0.923	0.811	0.857	0.700
G19	100	0.50	0.25	0.25	$2,714.2	64000	0.781	0.607	0.886	0.696	0.809	0.672	0.925	0.811	0.749	0.572	0.938	0.804	0.899	0.779	0.930	0.838	0.847	0.714
G20	100	0.50	0.38	0.13	$2,704.6	64000	0.765	0.599	0.851	0.737	0.796	0.622	0.923	0.789	0.723	0.560	0.935	0.827	0.908	0.817	0.942	0.827	0.843	0.729
H21	100	0.75	0.06	0.19	$2,706.1	69500	0.786	0.720	0.887	0.788	0.822	0.774	0.943	0.919	0.755	0.670	0.951	0.914	0.917	0.862	0.955	0.921	0.869	0.829
H22	100	0.75	0.13	0.13	$2,700.9	69500	0.803	0.764	0.879	0.846	0.827	0.766	0.948	0.914	0.755	0.711	0.955	0.928	0.946	0.883	0.951	0.923	0.864	0.823
H23	100	0.75	0.19	0.06	$2,706.2	70000	0.765	0.759	0.880	0.835	0.825	0.784	0.955	0.920	0.768	0.702	0.959	0.913	0.907	0.885	0.941	0.935	0.860	0.826

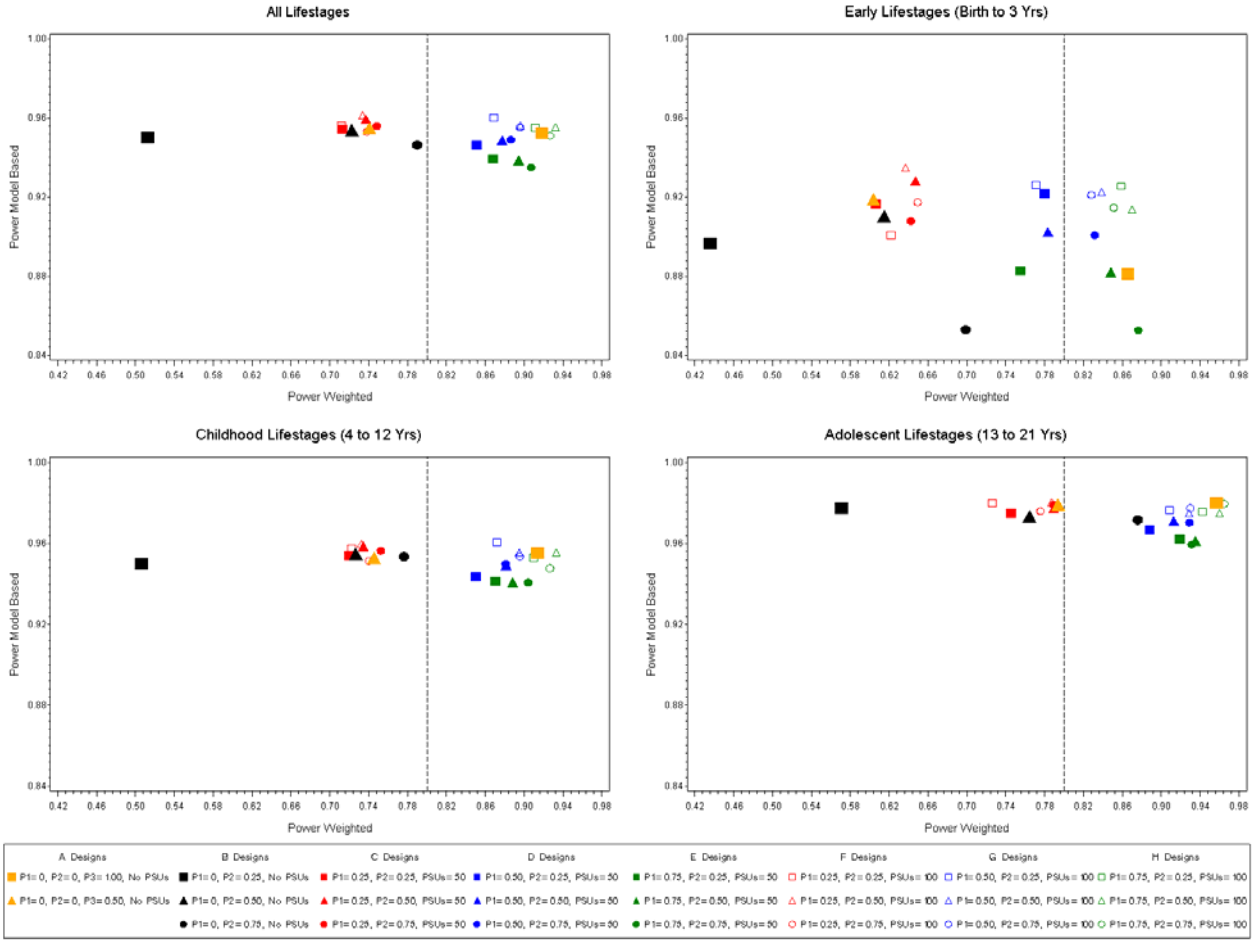

Figure 10-5. Average power (for unweighted and weighted analyses) to detect relationships of interest for hypotheses involving all life stages, early life stages, childhood life stages, and adolescent life stages assuming fixed sample size of 100,000 children and using the alternative retention rates (see Chapter 7).

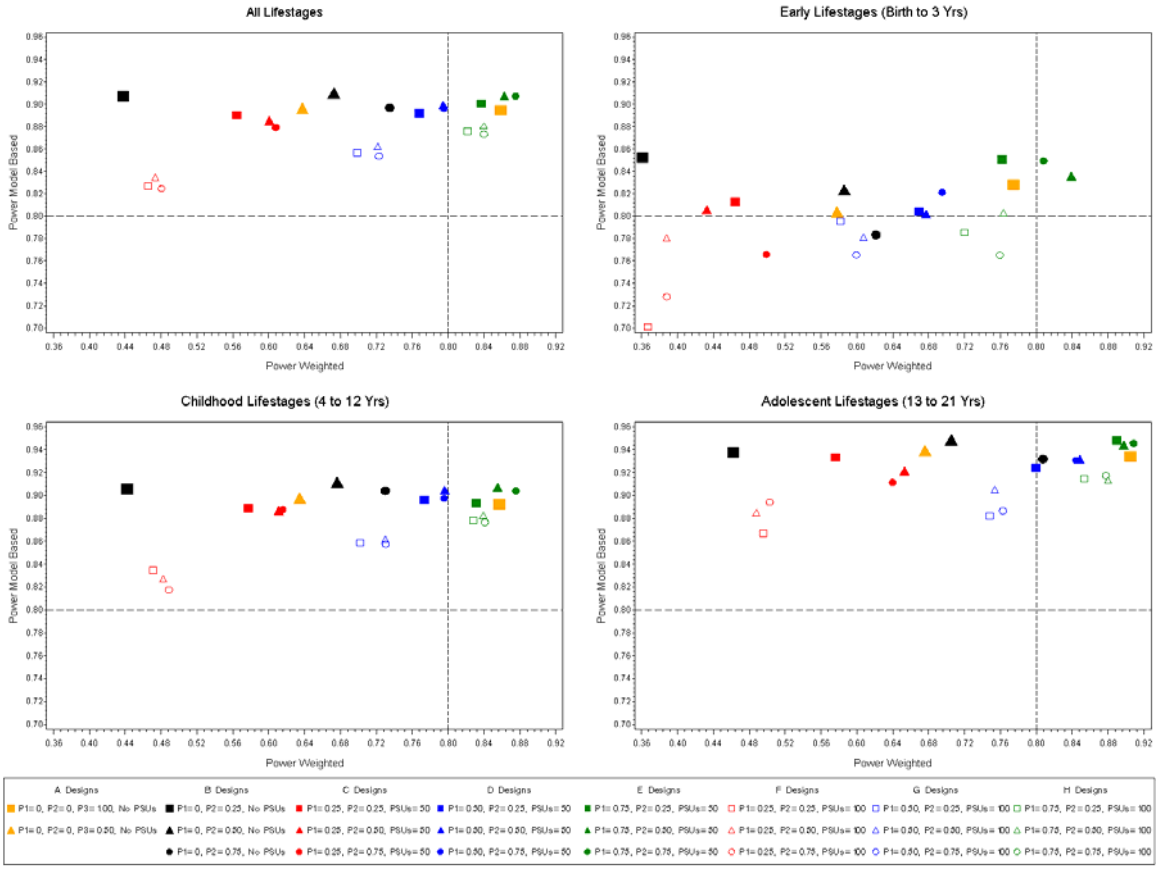

Figure 10-6. Average power (for unweighted and weighted analyses) to detect relationships of interest for hypotheses involving all life stages, early life stages, childhood life stages, and adolescent life stages assuming fixed costs of $2.7B and using the alternative retention rates (see Chapter 7).

10.2.2 Assumptions Related to the Cost Estimates

As discussed in Section 10.2.1 above, the cost estimates provided in Chapter 8 of this report are highly influenced by the assumptions related to recruitment and retention rates associated with each component in our family of designs – as a significant portion of the NCS operating expenses will be directly tied to the collection of data from study participants (see Appendix H). However, the cost projections for the study are also highly influenced by the assumptions that were made related to the costs associated with performing different aspects of the NCS. Chapter 8 and Appendix H provide details on the cost model (and associated assumptions) that was developed for providing cost projections for each design option discussed in this report. This cost model provides a starting point for discussion on the likely differences in cost that might be experienced based on the sampling design, and will require further scrutiny and refinement as plans for the NCS, such as specific data collection protocols, are further developed and specified.

Obviously, the cost estimates associated with the fixed sample size designs (i.e. designs that successfully result in an NCS cohort with 100,000 live births) are affected by the

assumptions described in Chapter 8 and Appendix H. In terms of the influence of the cost estimates on other technical results provided in this report, only the power estimates for designs constrained to a cost of $2.7 billion are affected. These power study results are sensitive to a number of cost assumptions, including:

- Costs of recruitment and retention under different modes of sampling
- Costs associated with operating Centers over the life of the study
- Costs associated with setting up and maintaining NCS infrastructure within PSUs (counties) selected as part of a National Probability-based Sample
- Per-subject costs associated with data collection at different stages of life.

It is difficult to project how changes to the above assumptions in the cost model would impact the power to address study hypotheses in the fixed-cost designs, other than to suggest that lower costs would likely lead to higher initial sample sizes and increased power.

Perhaps the cost assumption that has the greatest impact on the results provided in this document is that efforts and costs associated with retaining study subjects are consistent throughout the NCS cohort, regardless of the mode of sampling that is used for recruitment. This assumption is consistent with the lower assumed retention rates for women/children recruited through probability-based sampling of an unrestricted population. However, as discussed in Chapter 8, an alternative approach for the NCS might be to allocate proportionally higher resources for the retention of study participants that were recruited in this manner so that their retention rates are more similar to what is expected from the more convenient sampling approaches (probability-based sample of existing Center patients and convenience sampling).

It should be noted that the cost model for implementing the NCS under different design strategies is relatively easy to modify and update, and therefore, as discussed in Section 10.5, Battelle could readily assess how the costs associated with the fixed sample size designs or the initial sample size available in the fixed cost designs would likely change as a function of changes to any of the cost assumptions discussed in Chapter 8 and Appendix H.

10.2.3 Assumptions Related to the Power Studies

In the previous two sections, we discussed assumptions related to recruitment, retention, and cost – all of which have a significant impact on the power studies described in Chapter 9. However there are a number of other choices and assumptions that were made in conducting the power studies that merit discussion, including (1) the choice of hypotheses to explore, (2) the types of relationships studied, (3) critical assumptions that were integrated into the simulations that formed the basis of the power studies, and (4) the relevance of the odds ratios selected for presentation in the power study results. The following sections provide an overview of these four topics of discussion:

Choice of Hypotheses to Explore

Chapter 6 provides an overview of the core hypotheses of the NCS and our basis for selecting specific hypotheses for exploration in the power studies. In general, the nine

hypotheses that were chosen represent (1) all five priority outcomes, (2) those that will be most challenging to assess with high power, and (3) a range of lifestages when the health outcome of interest will be assessed. However, the following important factors should be noted:

- The core hypotheses for the NCS were reassessed during the time period when this design work was conducted. As a consequence, the two hypotheses related to injury that were selected for presentation in this report are under evaluation and will likely be changed or eliminated by the NCS Interagency Coordinating Committee. We chose to include the power study results from these two subhypotheses (3.1 and 3.2), because of their potential value in representing important avenues of research that could be supported by the NCS. However, these two hypotheses may not rise to the same level of importance as the other hypotheses that were selected for investigation.

- Our assessment of how challenging each hypothesis would be to address was based on a relatively simple inspection of the prevalence of the adverse health effects and primary measures of exposure that would be necessary to support each research objective. In some cases, as described in Chapter 9, either there was very little information regarding the prevalence of the primary risk factors or exposures, or the primary risk factor or exposure identified was not specific enough to provide accurate estimates of exposure (e.g., there are many candidate measures of exposure to air pollution that could be used to support the hypothesis that relates risk of asthma to early childhood exposure to air pollution.) In these cases, we tried to identify a conservative range of exposure prevalence to use in the power studies. For example, for the hypotheses related to asthma, we estimated that between 1% and 5% of the NCS study population would experience respiratory viral infection during pregnancy, maternal stress during pregnancy, and/or exposure to air pollution during early childhood.

- While all lifestages are represented in the group of hypotheses selected for the power studies, it should be noted that the adverse health effects with the lowest prevalence (autism, birth defects, and schizophrenia) are assumed to be observed in either very early stages of life (birth through 3 years) or in later stages of life (early adulthood – ages 18 to 21), while the adverse health effects with higher prevalence (injury, asthma, and obesity) were all assumed to be assessed during the middle of childhood (around age 10).

Types of Relationships Studied

As discussed in Chapters 6 and 9, the power studies that were pursued in this report focus on simple relationships that can be conceptually represented by a simple 2x2 contingency table that relates a simple binary measure of disease status to a simple binary measure of exposure. As discussed in Chapter 6, we believe that these simple logistic regression models lead to somewhat conservative estimates of power. In addition, while most of the hypotheses are written so that they could be supported by a one-sided statistical hypothesis test (e.g., impaired glucose metabolism during pregnancy is associated with an *increased* risk of congenital malformations of the heart), we chose to investigate power using a two-sided test – also providing a measure of conservatism in the power study results. Finally, while these power studies should be relatively

robust to the practice of controlling for the effects of important covariates and confounders – there will likely be substantially less power for assessing relationships between disease and exposure that are subject to important effect modifiers or interactions. For example, if a two-level genetic risk factor influences the relationship between disease and exposure with different odds ratios experienced for both groups, then assuming all participants have a measure of exposure, the effective sample size for assessing each odds ratio is reduced to the number of participants that are observed at each level of the genetic risk factor.

Assumptions of the Simulation Model

The simulation model that supports the power studies is described in detail in Chapter 9 and Section D.7 of Appendix D. The simulation model appropriately assigns sampling weights associated with study participants recruited under each mode of sampling within the family of designs, and also introduces within-cluster correlation in both the health outcomes and exposure variables. One critical assumption that is introduced in the simulation models that support the power studies is that of a constant odds ratio across all clusters. This is consistent with an assumption that the relationship between disease and exposure is constant regardless of the mode of sampling, or the location (cluster) where the study subject resides. That is not to suggest that the prevalence of the adverse health outcome or the primary exposure variables is constant from one cluster to the next, as the simulation model introduces intra-cluster variability in these measures.

The assumption of a constant odds ratio across clusters is consistent with our approach to designing this study with an initial emphasis on internal validity prior to assessing external validity. We took the point of view that for the clear majority of hypotheses that are currently under consideration, the relationship between adverse health outcomes and exposure would need to be transportable from one cluster to another to satisfy the check for internal consistency. If the relationship varies significantly from one cluster (or area) to another, it would most likely signify that an important explanatory variable or effect modifier was excluded from the model, and that the model fails the internal validity assessment.

Inevitably, analyses will be conducted on the NCS dataset in which the relationships vary from cluster to cluster (or subpopulation to subpopulation). We believe that the NCS will have similar overall power to detect marginal relationships across the cohort in this situation, but more importantly, that the NCS will have sufficient power to detect the true underlying relationships when the appropriate covariates and effect modifiers are added to the model. For example, consider the case in which we are assessing the relationship between a disease and exposure, and our power studies demonstrate that the study design will allow us to detect an overall odds ratio of 1.5. Suppose there is a fraction of the population who are genetically resistant to the disease (odds ratio of 1), and another fraction of the population that is susceptible (odds ratio much higher than 1.5) with exposure, and that the proportion of people with each genotype vary from cluster to cluster. If the genetic risk factor goes unmeasured in the NCS, we would expect to see intra-cluster variability in the odds ratio – leading us to the knowledge that something is missing from the model. We suspect that the NCS will still be able to detect a marginal odds ratio of 1.5 (even though the odds ratio varies from cluster to cluster), and that the study would also have sufficient power to detect the much higher odds ratio among people who are genetically susceptible to the disease.

Relevance of the Odds Ratios Presented

The odds ratios presented in the detailed tables in Chapter 9, as well as the summary tables (10-1 and 10-2) presented in this chapter were selected based on their ability to distinguish differences in the performance of the different sampling design options. The selected odds ratios in these tables range in values from 1.2 to 2.0, and generally result in a range of power estimates from model-based and weighted analyses across the design options. In reality, the true odds ratio for each of the hypotheses investigated is not likely to be well represented by the values listed in these detailed tables. In cases where the true odds ratio is measurably higher than what was assumed in the tables, all of the sampling designs should have reasonably high power for estimating the relationship. However, if the true odds ratio is measurably lower than what was assumed in the tables, only a select few (or none) of the design options may be able to detect the relationship with sufficient power. The power curves provided in Chapter 9 provide a basis for determining the relative performance of the different design options across a range of assumed odds ratios spanning 1.05 to 2.0.

Discussion of Sample Bias

In Chapter 2 and Appendix A of this report the issue of sample bias is discussed. In terms of sample bias, the primary concern for the NCS is whether the exposure/outcome relationships observed in the sample of individuals included in the cohort is biased relative to the reference population. For the realized sample of individuals, possible sources of bias include: (1) selection from a biased sampling frame (i.e., random or non-random selection from a set of elements that are different than the reference population), (2) utilization of biased selection mechanisms, and/or (3) bias introduced through non-response and recruitment failure. Additionally, when assessing relationships between exposures and outcomes (i.e., modeling the exposure/outcome relationship), there may be other sources of bias such as imperfect measures of exposure or failure to include important covariates and effect modifiers in the model (or failure to measure them). In general, all types of sampling (i.e., both random and non-random selection) have the potential to obtain a biased sample, some more so than others. Appendix A provides a more detailed discussion of the issue of sample bias, and the strengths and weaknesses of various sampling approaches as they relate to sample bias.

Admittedly, none of the results presented in this report have addressed the issue of sample bias. The various sampling frames were assumed to be unbiased, and the various sampling mechanisms were assumed to result in unbiased samples (e.g., random attrition and random recruitment failure was assumed). Additionally, accurate and unbiased measures of exposure and health outcome were assumed (see Appendix C for statistical methods that account for exposure measurement error). In other words, the power calculations described in Chapter 9 assumed that the relationships observed in the NCS cohort were the same as the relationships that would be observed in the larger reference population. That is not to say that sample bias is an unimportant issue for the NCS, as it is central to the ability to generalize results observed in the NCS cohort to the larger reference population. Thus, perhaps an important avenue for additional research is further investigation of the sources of sample bias, some determination of the potential bias that could be introduced through these various sources, and evaluation of the impact that these potential biases may have with regards to the study objectives.

10.2.4 Parameters Chosen that Define the Range of Design Options

As discussed previously, the family of designs concept allows us to cover a range of sampling design options for the NCS under a common hierarchy, with values of P_1, P_2, and P_3 (as seen in Figure 3-1 and Table 3-1 of Chapter 3) governing the fraction of the NCS cohort that are recruited under the following four modes of sampling:

- National Probability-Based Sample (NPBS)
- Probability-Based Sample of MSAs surrounding Purposively Selected Centers (Center-area PBS)
- Probability-Based Sample of Existing Patients from Purposively Selected Centers
- Convenience or Opportunity Sample (from around the Centers).

In addition to the choices of the fraction of the NCS cohort that are recruited under the above modes of sampling are the following assumptions:

- If the value of P_1 is greater than 0 (i.e., part of the NCS cohort is selected under a NPBS), then the number of PSUs selected into the study is nominally set at either 50 or 100. This allows us to investigate the advantages and limitations of having more or less PSUs involved in the NCS (impact on cost, power, etc.).
- The selection mechanism for PSUs in the NPBS (if $P_1>0$) is based on probability proportional to size sampling from among eight strata (four regions of the country crossed by Urban/Rural designation). However, depending on the fraction of the NCS cohort selected to be represented by areas in proximity to a university-based medical Center (which is likely to be highly concentrated in urban MSAs), further refinement of the selection algorithm of PSUs as part of the NPBS may include oversampling of rural areas so that they are proportionately represented in the NCS. Depending on the number of PSUs involved in the NCS and the fraction of participants recruited via a NPBS, we may also need to determine whether PSUs are defined as counties or groups of counties in close proximity for the rural portion (as the desired sample size from a rural PSU in some circumstances may be larger than the number of births expected to be successfully recruited from a single rural county).
- We assumed that the Centers have the ability to conduct data collection activities for a sample that includes approximately 2,000 live births (with a fraction of participants referred to the Center from a PBS of the surrounding MSA, and the remaining from either a PBS of Center patients or a convenience sample). In addition, we assumed in the cost and power studies that each Center would perform data collection on an equal number of study participants (which may not be optimal), and that the Centers would be located in metropolitan statistical areas, as designated by the US Census Bureau. It should be noted that NCS Centers of Excellence could be established in rural areas of the country as well – however, assuming that the Centers represent university-based medical and research facilities, in most cases this would entail the university establishing a center away from its main campus facilities.

10.3 EVALUATION OF HOW DIFFERENT SAMPLING APPROACHES MEET NCS OBJECTIVES AND STUDY DESIGN GOALS

The premise upon which this report is based is that there are multiple legitimate design options for the NCS, which can be discussed under a systematic framework which we have referred to as a family of designs. The bulk of the report to this point has been dedicated to providing estimates for some of the more difficult-to-assess and data-dependent properties of the design options including cost, recruitment and retention rates, and power. In this section we integrate those estimates into a broader evaluation framework for the design options that attempts to consider all the major objectives and constraints of the NCS – as presented in Chapter 1 of the report.

The objectives and constraints of the NCS effectively define the criteria by which alternative design options should be evaluated. There are three primary types of criteria that apply to any proposed design:

1. Ability to satisfy "givens" (constraints) required by the legislation or the government;
2. Scientific merit (validity, value, and feasibility); and
3. Cost.

Appendix B1 includes further discussion of specific criteria for evaluating the NCS related to each of the above categories. For the purposes of this section, the primary criteria that will be used in the overall evaluation of design options are introduced in Table 10-6, and consists of 16 separate criteria: three related to the givens of the study, 11 related to scientific merit, and two related to costs.

Under the family of designs framework presented in this report, three factors characterize the main differences in design options:

- Proportion (P_1) of the cohort in the national probability-based sample (NPBS),
- Proportion (P_2) of the remaining cohort (1- P_1) that is selected on a probability basis from the MSA surrounding a center[1],
- Number of PSUs in the national probability sample.

Table 10-4 includes an initial evaluation of three of the design options that were considered, selected to broadly span the differences in options. The selected designs were:

[1] Assuming that a relatively small portion of the cohort is recruited via a convenience or opportunity sample, the cost and power analyses results are highly dependent on the choice of the P_1 and P_2 parameters, and relatively robust to the specific choice of the P_3 parameter in the Family of Designs. In essence, within the (1-P_2) remaining fraction of center-based participants, the choice of P_3 helps to differentiate the number of participants recruited via a probability-based sample of patients already affiliated with the center and the number of participants recruited via a convenience or opportunity sample. While the choice of P_3 is important to consider when designing the study, it has relatively low impact on power and cost.

- C7 where 25% of the cohort is included in the NPBS (P_1=25%), 50% of the remaining cohort is chosen on a probability basis from the surrounding MSAs (P_2=50%), and 50 PSUs are chosen for the NPBS;
- D10 where P_1=50%, P_2=50%, and there are 50 PSUs; and
- H23 where P_1=75%, P_2=75%, and there are 100 PSUs.

An assessment of the criteria for all the design options is provided in the table in Appendix J. Additionally, Appendix J includes a table that assesses the cost, attrition, and statistical power criteria using the alternative retention rate assumptions in Chapter 7.

Table 10-6 includes only criteria for which there are judged to be significant differences across the design options. For example, all proposed design options are considered observational studies that are national in scope and address the same set of core hypotheses proposed by the ICC. Therefore these design objectives are considered met by all of the design options and are not specifically evaluated. It is also assumed that each design option will meet all human subject requirements. This assumes that a reasonable process for obtaining an overall approval for the common procedures and data collection elements of the study will be identified and that completely separate IRB approvals for all aspects of the study will not be required for each Center in a Centers-based design. If this is not the case, then multiple differing IRB requirements could become a significant disadvantage to a design with multiple Centers.

In addition, all design options are considered to be basically equivalent in their ability to handle cohort mobility during the course of the study. Under any of the designs, there will need to be a separate procedure and implementation process to track and maintain contact with subjects who move out of the PSU or Center under which they were originally recruited, in particular when they do not move within the confines of another PSU or Center. It is possible that there could be some efficiencies in using an organization that is implementing the NPBS (especially for designs with 100 PSUs) for the tracking and maintenance of participants who move outside the boundaries of any established PSU or Center, but a basically separate process will need to be established and funded to track movers regardless of the design option chosen.

Following is a discussion of the differences in the three example design options for each of the criteria in Table 10-6. Each design option receives a rating for each criterion, either on a qualitative basis (for example as excellent, good, fair, or poor) or quantitatively using statistics developed in this report, according to the rationale described below.

10.3.1 Givens

The first "givens" criterion relates to the ability of the NCS to engage and benefit local communities. We assume local community engagement and benefit is most likely to be accomplished by the involvement of Centers already established in the community. Thus design options with higher Center involvement receive a higher rating, even though there may be community involvement in PSUs selected as part of the NPBS.

The second "givens" criterion captures the design requirement that the study provide access to infrastructure that allows specialized medical measures and ease of access/collection of

biological samples at birth. We assume that Centers are much better positioned to provide this access compared to randomly selected PSUs within a NPBS. Therefore design options that have a higher percentage of Center involvement rate higher relative to this criteria.

The third "givens" criterion corresponds to the design requirement that the NCS enroll study participants as early as possible in pregnancy, with perhaps a subset of study participants being recruited prior to conception. Each sampling approach within the Family of Designs has distinct advantages and limitations with regard to meeting this criterion. For example, the probability-based sampling approaches that recruit from a large population rely upon a household model of recruitment within selected PSUs or MSAs surrounding the centers. This model for recruitment provides a mechanism for initial recruitment of age-eligible women who are not yet pregnant into the study, although this may not be a particularly resource efficient method of recruitment. Conversely, the probability-based sample of patients already affiliated with the centers and other center-based convenience samples can potentially identify women in early stages of pregnancy in a very resource-efficient manner through existing relationships and the pre-natal care that is already provided at the centers. However, it needs to be determined whether this mode of recruitment will provide the NCS with access to patients early enough in pregnancy to satisfy the goals of the study. Women affiliated with the centers can also be identified, recruited, and followed prior to conception – using similar methods as those proposed for the household model with initial contact limited to age-eligible women rather than all households. There is no doubt that recruitment of women in early stages of pregnancy (and in some cases prior to conception) will be one of the most challenging aspects of implementing the NCS. However, we conclude that all of the sampling design options explored in this report have the ability to satisfy this criterion – and that no sampling approach offers clear and distinct advantages over another with respect to this "given".

Table 10-6. Evaluation of Three NCS Study Design Options Relative to Givens, Scientific Merit, and Cost

Category	Criterion	Description	C7: P1:25% P2-50% 50 PSUs	D10: P1:50% P2-50% 50 PSUs	H23: P1:75% P2-75% 100 PSUs
GIVENS	1	To what degree does the study involve the local community, leverage community involvement, and help local community public health efforts? To what degree does the study provide flexibility to conduct special studies, particularly related to topics of community interest? (Rated as: Excellent (Exc.); Good; Fair; or Poor)	Good	Fair	Poor
	2	To what degree does the study provide access to infrastructure that allows specialized measures? To what degree does the study provide for ease of access/collection of biological samples at birth? (Rated as: Excellent (Exc.); Good; Fair; or Poor)	Good	Fair	Poor
	3	Prenatal recruitment should occur as early in pregnancy as possible, with perhaps a subset of study participants being recruited prior to conception.	Good	Good	Good
SCIENTIFIC MERIT: External Validity	4	Does the design clearly specify a sampling frame from which the study population is drawn, and can statistical methods be used to generalize results and characterize uncertainty for the sampling frame population? Are statistical adjustments available to account for non-response, non-random selection, or other deviations from specified probability selection from the sampling frame for the design? (Rated as: Excellent (Exc.); Good; Fair; or Poor)	Fair	Good	Exc
SCIENTIFIC MERIT: Diversity	5	To what degree does the design allow recruitment of a diverse population – geographic, ethnic, socio-economic, and other factors determined by study leaders. To what degree does the design ensure that target populations or exposures are represented in the study, and that heterogeneity in the study population is captured? To what degree does the design maximize the probability of observing the type and range of risk characteristics and exposures of interest, or maximize the probability of observing an outcome of interest? (Rated as: Excellent (Exc.); Good; Fair; or Poor)	Fair	Good	Exc
SCIENTIFIC MERIT: Internal Validity	6	To what degree does the design support collection of information on a wide range of potentially confounding factors that might influence associations of interest? To what degree does the design help in obtaining representative, accurate, and effective measures of exposure or outcome on the sampled population? (Rated as: Excellent (Exc.); Good; Fair; or Poor)	Good	Fair	Poor

Category	Criterion	Description	C7: P1:25% P2-50% 50 PSUs	D10: P1:50% P2-50% 50 PSUs	H23: P1:75% P2-75% 100 PSUs
SCIENTIFIC MERIT: **Power**	7	What is the average power across all hypotheses specified in Table 10-3 for the design option to detect the target associations of interest in an unweighted analysis when initial sample size is fixed at 100,000? Top percent shown is the average power. Bottom number is how many of the design options had less power (e.g. >90% indicates 90% of the design options had less power)	92% (>65% of other designs)	89% (>35% of other designs)	85% (>4% of other designs)
	8	What is the average power across all hypotheses specified in Table 10-1 for the design option to detect the target associations of interest in an unweighted analysis when total overall study costs are fixed? (e.g. >90% indicates 90% of the design options had less power)	86% (>57% of other designs)	87% (>74% of other designs)	81% (>22% of other designs)
	9	What is the average power across all hypotheses specified in Table 10-1 for the design option to detect the target associations of interest in a weighted analysis when initial sample size is fixed at 100,000? Top percent shown is the average power. Bottom number is how many of the design options had less power (e.g., >90% indicates 90% of the design options had less power)	63% (>28% of other designs)	76% (>61% of other designs)	80% (>89% of other designs)
	10	What is the average power across all hypotheses as specified in Table 10-1 for the design option to detect the target associations of interest in a weighted analysis when total overall study costs are fixed? Top percent shown is the average power. Bottom number is how many of the design options had less power (e.g. >90% indicates 90% of the design options had less power)	51% (>22% of other designs)	71% (>56% of other designs)	76% (>78% of other designs)
	11	What is the estimated rate of attrition for the design option in terms of number of participants remaining at the end of the study?. Bottom number is how many of the design options had fewer remaining participants (e.g. >90% indicates 90% of the design options had fewer participants remaining in the study at the end)	49,311 (>70% of other designs)	41,774 (>48% of other designs)	31,717 (>0% of other designs)
SCIENTIFIC MERIT: Resource for the Future	12	What aspects of the study design make it particularly strong in terms of serving as a resource for future studies and assessment of hypotheses identified in the future? (Not Rated Due to Significant Tradeoffs as Discussed in Text)	?		
	13	To what degree does the study design option maximize the chance of obtaining standard measures across the entire cohort throughout the life of the study? (Rated as: High, Medium, or Low Difficulty)	High Difficulty	Medium Difficulty	Low Difficulty
SCIENTIFIC MERIT: Access to the Data	14	To what degree does the study design allow for effective and timely access to the study database by a variety of researchers? (Rated as: Excellent (Exc.); Good; Fair; or Poor)	Good	Fair	Poor
COSTS	15	What are the estimated costs of the study, assuming a fixed initial sample size of 100,000? Top number is the estimated cost for the design option. Bottom number is how many of the other design options had higher costs (e.g., >90% indicates 90% of design options cost more)	$3.2B (>57% of other designs)	$2.9B (>17% of other designs)	$2.9B (>22% of other designs)
	16	What are the cost impacts in terms of reduced initial sample size when the overall cost of the study is fixed? Top number is the initial sample size. Bottom number is how many of the design options had lower initial sample size (e.g., >90% indicates 90% of the design options have less participants included initially when overall total costs are fixed)	80,000 (>39% of other designs)	91,000 (>78% of other designs)	89,000 (>74% of other designs)

10.3.2 Scientific Merit

External Validity: Assuming that the most straightforward path to external validity, the ability to extend study results to a broader population, is through statistical inference based on the sampling design, we rate designs highest for this criterion that have the strongest probability basis (i.e., that have the most participants selected on a probability basis and largest reference population).

Diversity: Diversity can be achieved either by specifically targeting certain subgroups of the population, for example, a highly exposed or low-SES neighborhood, urban or rural areas, or by achieving heterogeneity in the sample. Any of the designs could be structured to oversample and target specific populations of interest and to achieve diversity in this way. However, the designs that select from a large population on a probability basis, either NPBS or the MSA area coverage, are assumed to provide better heterogeneity than designs which focus on a Center's patients. Therefore we rate achievement of the diversity criterion higher as the percentage of participants restricted to Centers' patients decreases.

Internal Validity: We assume, at this point, that fewer resources will be required to recruit and maintain Center patients than would be required for a probability sample of participants. We also assume that Center patients will be more amenable to higher burden, with potentially higher rates of retention and more opportunity for detailed and precise assessment. With these assumptions, we rate designs with a higher percentage of Center patients higher in their ability to satisfy the internal validity criterion.

Power: A wide variety of power estimates for the different design options are presented in this report. Table 10-6 includes four measures of power – power using a weighted and an unweighted analysis with both fixed initial sample size (varying cost) for the study as well as fixed cost (variable initial sample size) for the study. For each of these four measures, the table provides the average of the power estimates for the specified relationships of interest (see Tables 10-1 and 10-2) across all hypotheses, as well as the percent of the other design options that had lower average power (thereby allowing a ranking of design options with respect to power).

While there are obvious tradeoffs between design options *across* criteria in Table 10-6 that we will discuss below, there are also significant tradeoffs between design options *within* the power criterion. In general, for the unweighted analyses, power is higher on average for the designs with less attrition and larger sample size (i.e., designs with less probability basis). The opposite is generally true for the weighted analyses where designs with more probability basis typically have higher power. However, for the weighted analysis there is also a significant tradeoff between sample size and the proportion of the study participants recruited using probability-based sampling from large unrestricted populations. Therefore, for some hypotheses, especially those where the outcome cannot be measured until well into the study, the increase in power in the weighted analysis due to participants selected on a probability basis from a large population may eventually be washed out by the reduced power resulting from a decrease in sample size with the assumed greater attrition for the probability sample. An example of this can be seen in Table 9-7 where the power to detect undesirable outcomes of pregnancy diagnosed

later in life in a weighted analysis with 50 PSUs peaks with P_1=50%, and is lower with P_1=25% or P_1=75%.

Figures 10-7 and 10-8 provide a graphical illustration of the tradeoff in power for the weighted and unweighted analyses, for both a fixed initial sample size of 100,000 and a fixed-cost (variable initial sample size), respectively. In these figures, the average weighted and unweighted power (for a specific relation of interest) for each design option is plotted on the same graph and connected by a straight line. The more sloped the line, the greater is the discrepancy between the weighted and unweighted results. The higher the line, the greater is the average power. In some manner, these graphs could be interpreted as providing a basis for choosing a design based on the value that one places on power from unweighted analyses (or internal validity of the study) versus power from weighted analyses (or external validity of the study). In other words, one could calculate a weighted average of the two types of power estimates to serve as a basis for selecting a design. For example, if a value judgment is made that internal validity is three times as important as external validity, one could inspect to see which design options have the highest weighted average of power at a point that is about 75% of the way toward the unweighted (model-based) side of the x-axis on these figures. It should be noted that this represents only one of many ways that could be used to assess tradeoffs between weighted and unweighted power. Other methods might examine the number of hypotheses estimated to have power above a specified level or the estimated variability for the relationship between outcome and exposure associated with each design. All approaches will be affected by the choice of odds ratio, and the amount of emphasis given to weighted or unweighted analyses.

In summary, each design option has mixed results relative to power. If one places a higher emphasis on the validity of a weighted analysis, then the design options with a higher probability component perform better. On the other hand, if one expects to use unweighted analyses, then the reverse is often true. An important point to note, however, is that designs that integrate a mixture of sampling approaches have the potential to raise power across the different types of analyses, as indicated by the results listed in Table 10-6 (Scientific Merit Criteria 7-10) for design option D10.

Attrition: Due to assumptions discussed in Chapter 7, design options with a higher probability component are estimated to have higher attrition rates, as captured in Table 10-6 as the estimated number of participants remaining in year 20 of the study, and the percentage of the other design options that had less remaining participants.

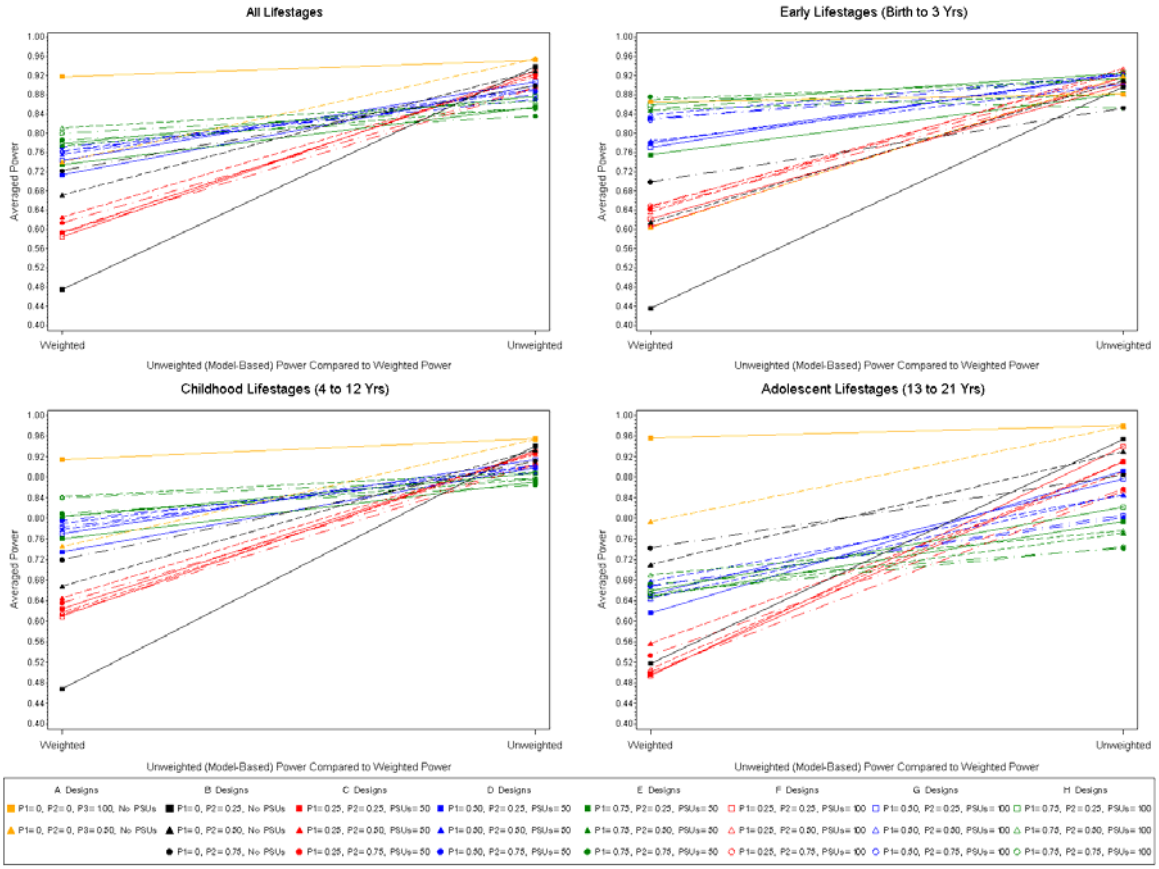

Figure 10-7. Example trade-offs in power between unweighted (model-based) and weighted analyses for sampling designs with fixed initial sample size.

Resource for the Future: There are significant tradeoffs between the different design options that make it difficult to assign a rating for this criterion. Most notably, designs with a higher probability component offer the important advantage of being able to generalize associations investigated and detected in the future based on the probability sampling basis. Given that we cannot know *a priori* all the important measures needed to support unspecified future hypotheses, this property provides a significant protection for exploring a wide range of possible hypotheses in the future. On the other hand, designs with a higher probability component could significantly decrease the usefulness of the NCS as a future resource if these designs result in fewer archived samples or, more importantly, in fewer participants remaining in the study in the out years to be available for measurement and outcome-dependent studies. Therefore, no ratings have been assigned for this criterion.

Standardization Across the Study: While standardization can be achieved in theory for any of the chosen designs, the degree of difficulty will be related to the number of separate organizations that are chosen to implement the study. Assuming that one or few organizations will be responsible for the NPBS, the level of difficulty will be less for designs that recruit a higher fraction of participants through the NPBS. As the number of independent organizations (Centers) increases, there will be more requirements for training, protocol audits, IT control, etc.

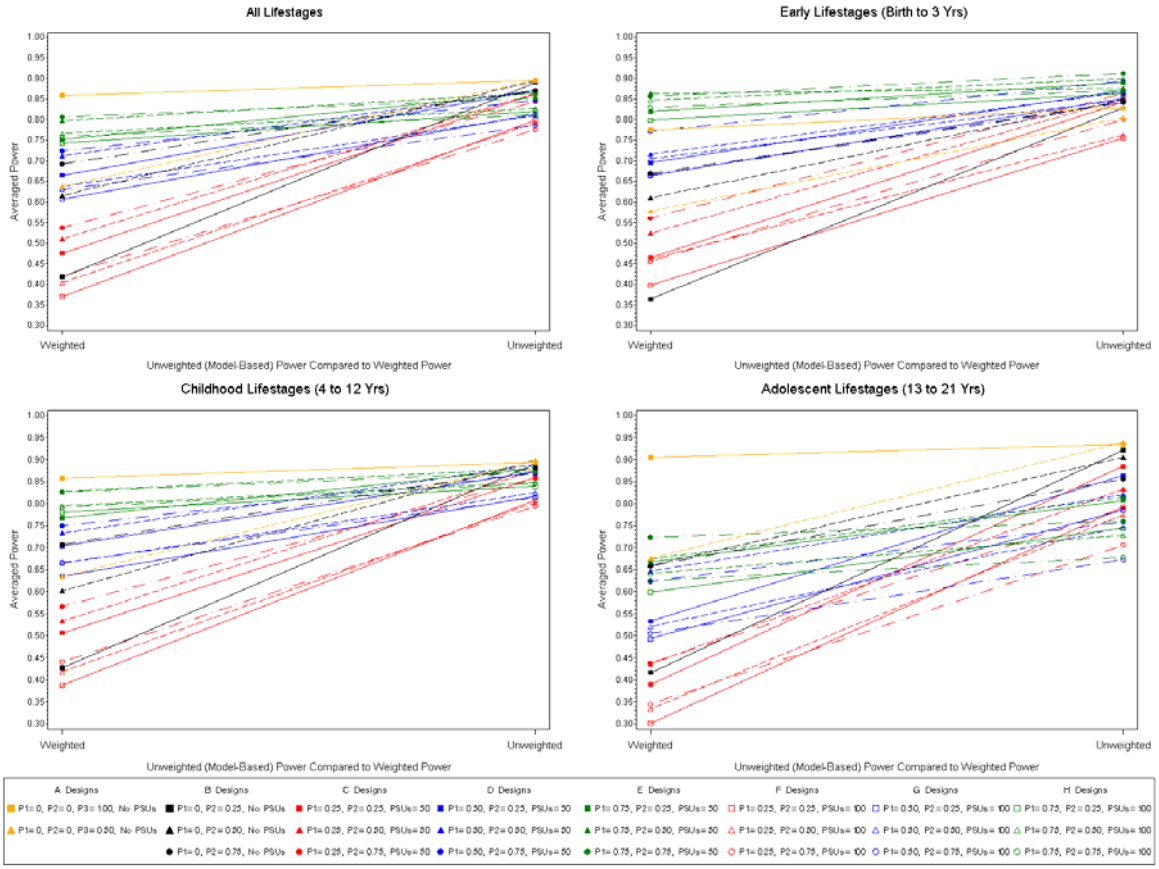

Figure 10-8. Example trade-offs in power between unweighted (model-based) and weighted analyses for sampling designs with fixed costs of $2.7 billion.

Access to Data: Timely access to the data by a variety of researchers will depend to a large degree on the efficiency of the NCS Information Management System and database. However, we assume that if a Centers of Excellence approach is taken to study implementation, then researchers at each Center would have more immediate access to their own data. Therefore, we rank designs with more Centers higher relative to timely access to data by a variety of researchers. Access to the complete dataset should be the same under any of the design options, assuming efficient data management protocols are established.

10.4 CONCLUSIONS

In looking across all the information presented in this report on the characteristics of various design options, and in evaluating that information relative to study objectives, the first and most important conclusion is that significant tradeoffs appear inevitable. These tradeoffs are reflected in Table 10-6, for example, the design with the highest rating for external validity (H23) has the lowest rating for community involvement. There is no single design that leaps out as clearly the best choice from all perspectives. Therefore, the ultimate choice of design cannot be separated from value judgments related to the importance of the different, and sometimes competing, study objectives. The information in this report, however, allows decision makers to understand those tradeoffs in detail, and therefore to be able to make informed decisions when choosing one design over another by understanding what is being gained and what is being lost.

In light of the study givens, it appears that a final design will need to include a significant Center component to meet the community and specialized measure requirements. On the other hand, it appears that including a probability component offers many advantages related to scientific merit and external validity. Therefore, a hybrid approach within the family of designs that incorporates both sampling approaches seems highly desirable.

Hybrid approaches identified in this report, most notably designs that include from 25% to 50% of the sample being conducted as a NPBS, with the remaining percentage covered under centers with the probability component to be negotiated, do offer an attractive balance, achieving power for external validity that appears reasonable for many hypotheses, and that still allow significant community involvement and ability to recruit highly motivated participants. Thus, a hybrid design is possible which is both acceptable and defensible across multiple objectives.

Finally, in reviewing the technical information in this report, the uncertainty associated with expected recruitment and retention rates is one of the most significant limiting factors in more precise estimates of the value of the different designs, leading to a high recommendation in Section 10.5 for pilot work to better understand this issue.

10.5 RECOMMENDATIONS FOR FUTURE WORK

In the short term, there will be much value in conducting updated cost and power analyses, and perhaps sensitivity analyses, as a function of changing assumptions on recruitment, retention, and costs of implementing the NCS. In addition, the study planners and the expert

panel reviewing this body of work are likely to suggest useful alternatives to the parameter values and other components that define the design options presented in this report. Examples include the following parameter values, as discussed in Section 10.2.4: the values of P_1, P_2, and P_3; number of PSUs in a NPBS; fraction of the NPBS that concentrates in rural areas; and number of participants that can be serviced by Centers. In addition, alternatives for sampling frames and organizational structures that can be used, and for the assumptions used for retention rates and study costs may be provided for consideration. The investigation of additional hypotheses – including hypotheses that would be supported by different or more complex models of the relationship between adverse health effects and exposure, could also be integrated into this design work in an efficient manner. In particular, the hypothesis-specific power studies that were developed for this report do not rise to a level of complexity that allows for careful consideration of the impact of strategies for oversampling of specific subgroups of the population of interest. As discussed in Section 5.4 of Chapter 5, oversampling of specific subpopulations in the NCS may benefit power to address some hypotheses, while detracting from the power to address other hypotheses. Careful investigation of this issue may be an obvious follow-up activity for the design work that was presented in this report. It should be noted that the infrastructure that has been established to support this sampling design project (as well as the previous design project supported by EPA's National Exposure Research Laboratory discussed in Appendix C) will easily accommodate these types of short-term analysis activities – in which the cost model and power studies can be reparameterized, upgraded, and recalculated rather efficiently.

Our additional recommendations for future work revolve around four major themes that would likely help optimize the choice of a final sampling strategy for the National Children's Study:

(1) Determine the data collection protocol necessary to support the main research objectives of the NCS

(2) Determine the likely recruitment and initial retention rates associated with the four modes of sampling within the family of designs, as a function of the burden that is likely to be imposed by the data collection protocol

(3) Refine the final choice of sampling design based on the results of the above two efforts as well as the guidance received from the expert panel.

(4) Develop statistical tools for the analysis of data from the NCS under the complex design, and for the design of nested substudies within the NCS cohort.

The following sections provide some additional detail on these suggestions for additional future work.

10.5.1 Defining the Core Sampling Protocol

If a family of designs concept is to be realized, with a fraction of study participants being recruited through probability-based sampling of a relatively unrestricted population base, it will be very important to define a core data collection protocol that keeps the burden on study participants to a reasonable level. In Chapter 3, we discuss the notion of having a core data collection protocol, which is standardized across all study participants and provides coverage across the main research objectives of the NCS. This data collection protocol may exclude certain difficult to obtain, burdensome, or expensive measurements under the assumption that easier to obtain (yet less accurate) surrogate measures are obtained for all study participants. A discussion of a series of statistical design principles for obtaining efficient exposure information for the NCS is contained in Appendix C, and should provide a reasonable starting point for ways to think about leveraging detailed exposure assessment information collected on a small subset of NCS participants to appropriately model exposure in the larger NCS cohort.

Appendix F, which contains a White Paper on Measures for NCS Core Hypotheses, may also provide a starting point for the determination of an appropriate data collection protocol for participants of the NCS. This White Paper provides a systematic review of the key measures and methods for obtaining these measures that are necessary to support each of the current NCS hypotheses, as a function of life-stage. This appendix also provides aggregated tables that suggest the data collection burden that might be necessary to support the NCS as a function of lifestage and method for obtaining the data (e.g., blood sample, urine sample, other physical sample, medical record review, interview/questionnaire, physical exam, and direct observation).

Under the assumption that some of the design concepts contained in Appendix C are integrated into the data collection protocol for the NCS, we suggest that the NCS consider initiating pilot studies to

1. Assess the relationship (through literature reviews, analyses of existing datasets, and/or conducting small scale measurement studies where necessary) between detailed exposure assessment measures and less burdensome surrogate measures of exposure (so that the surrogate measures of exposure can be used for the majority of NCS participants),
2. Characterize the amount of within-person temporal variability and trends over time in key measures of exposure and primary risk factors (leading to a potential reduction in the number of data collection events necessary to support the NCS research objectives), and
3. Assess the relationship between prospective and retrospective measures of exposure and risk (also potentially leading to a reduction in the number of data collection events), with special emphasis on assessing preconception and early-pregnancy exposures.

The results of these pilot studies will hopefully lead to reductions in the minimal data collection protocol that can be made with confidence. These pilot studies may also influence the sampling design protocols in other ways. For example, the fraction of the NCS cohort that must be recruited in the preconception phase may be substantially reduced if temporal variability in

key exposure measures is well understood, or if it can be shown that retrospective measures of preconception exposure that are assessed in early stages of pregnancy are sufficient to support certain preconception hypotheses.

10.5.2 Establishing Realistic Recruitment and Retention Rates for the NCS

Once the data collection protocol for the NCS is better defined (as discussed above in Section 10.5.1), it is important to pilot test recruitment and initial response rates as a function of mode of sampling within the family of designs and the burden imposed by different modes of data collection (both the core data collection protocol, as well as add-on data collection that collects more detailed information, possibly targeting certain sampling frames or with increased incentives). This pilot test should be done in as realistic a setting as possible, with appropriate informed consent procedures and a planned schedule of incentives corresponding to the planned data collection activities over the life of the study. Initial recruitment rates should be assessed, along with an assessment of the overall enthusiasm of the study participant (women in early stages of pregnancy or preconception). Once initially recruited, two to three waves of realistic data collection activities should be applied to this pilot study sample, so that response rates for the first phases of the study can be identified. If possible, it would be useful to assess the retention rates through birth and slightly beyond (e.g. through 6 months), so that we can gauge whether the retention rates change as a function of switching the focus of data collection activities from the pregnant woman to her child. In addition, methods for maximizing retention should be carefully investigated, and their use should pilot tested here. Costs of recruitment and data collection activities under each sampling approach should be carefully monitored as well (e.g., statistics on the number of households visited, number of scheduled appointments for data collection that are missed, number of participants that complete only part of the data collection protocol, etc.)

This pilot study is a critical element to finalizing the sampling design for the NCS. As discussed in Appendix A, there are advocates of probability-based sampling who expect that initial recruitment rates for the NCS using PBS in an unrestricted population will be between 70 and 90 percent, while others believe that it will in fact be much lower. The fact is that there is currently no empirical evidence suggesting what the recruitment and retention rates will be for a study as complex and burdensome as the NCS. For example, if the initial recruitment rates are very low (e.g., 30 percent) from the NPBS, study planners may want to think more carefully about the value of pursuing a NPBS (perhaps they could establish data collection activities in purposively selected rural areas that are operated by the Centers, such as the current NIEHS/EPA funded study of migrant farm workers in the Salinas Valley Center operated by the University of California at Berkeley).

10.5.3 Refining the Final Choice of Sampling Design

Once data from the above two efforts is available, we believe that study planners can confidently proceed with activities to optimize the sampling design with an integrated plan for data collection activities for the NCS. This optimization needs to simultaneously address cost, retention, and power to assess NCS core hypotheses. While it is conceivable that, as a result of

this initial design work (with review, input, and modifications incorporated from the suggestions of the expert panel, ICC, NCSAC, etc.), study planners will be able to move forward with a consensus strategy for the NCS sampling design. The final NCS design would benefit tremendously from the more accurate estimates of recruitment, retention, and costs of study implementation that are experienced in the pilot study described above. The cost model can be properly updated to take these improved estimates (and made more realistic as the study protocol and structure are developed further) into account, and the power studies can be based on more realistic retention rates (and cost constraints).

Assumptions related to intra-cluster correlations for exposures and outcomes may also be explored further and integrated into the design to improve the accuracy of the power studies for key NCS hypotheses[0]. Further work could also be conducted to assess the ability for NCS candidate designs to address more complex relationships that include major confounding variables and effect modifiers, and the impact of including these variables on power and validity. Methods to improve and equalize weighting for the different sampling approaches must also be considered in optimizing the design for the NCS (e.g., adjust numbers of patients required from Centers or expand/contract the populations represented). We also need to consider expanding coverage to rural areas more formally, which may be accomplished through a stratified over-sampling approach as part of a national probability-based component of the study or by using center-based outreach programs (e.g., cooperative extension programs) or health care clinics located in underserved areas that are affiliated with purposively selected Centers.

After completing the above listed work, the conceptual sampling design for the NCS, including final specification of the fractions of the NCS cohort recruited under the different sampling approaches, the number of PSUs, etc. can then be selected to allow the NCS to achieve the goals discussed in Section 10.3 in an "optimal" manner. With a final conceptual plan for sampling design available, based on more defensible estimates, NICHD and its consortium partner agencies can then proceed with developing the specific implementation plan for the NCS – with the selection of PSUs and Centers consistent with this design, enumeration of the appropriate sampling frames, etc.

10.5.4 Development of Statistical Tools to Support the NCS

We conclude our discussion of recommended future work in support of the sampling design for the NCS with a list of statistical methods development work that would be useful and perhaps necessary if the NCS were to move forward with a complex design that involves multiple sampling approaches:

- ***Tools to Efficiently Calculate Sampling Weights When Extracting NCS Data:*** As discussed in Appendix E, the NCS will be a complex longitudinal study that experiences different types of nonresponse (failed recruitment, wave nonresponse, item nonresponse, and attrition). In addition, a nontrivial fraction of NCS participants will physically move away from the area in which they were originally recruited – which will affect the assessment of clustering in the analysis phase. It will not be sufficient to develop a static series of sampling weights and cluster designation that can be used for the analysis of

data from the NCS – rather, a dynamic tool will likely need to be developed so that appropriate weights and clustering designation can be assigned in an appropriate manner for the intended statistical analysis. This tool can be integrated into the NCS Information Management System so that appropriate sampling weights and cluster identifiers are provided to researchers when data are extracted in support of specific research objectives.

- ***Diagnostic Tools to Verify Assumptions on Internal and External Validity:*** There are many assumptions related to internal and external validity that should be assessed while analyzing data from the NCS. For example, as discussed in Section 10.2.3, the power studies that were pursued in this design work were based on an assumption of a constant odds ratio across all clusters. This translates into an assumption that the relationship between adverse health outcome and exposure is transportable and consistent across all areas – and has direct relevance to the assumption of internal validity of the model. Similarly, weighted analyses of NCS data are not necessarily representative of the population of interest – especially in the presence of imperfect recruitment and response. We believe that in addition to developing statistical analysis tools that appropriately model NCS data while taking into consideration the complex design and data collection protocol, a series of diagnostic tools should be developed to help researchers assess their statistical models of the NCS data. For the first example, a tool could be developed to provide empirical estimates of the variability in cluster-specific odds ratios as a measure of internal validity. For the second example, information on study participants who failed to provide data in support of the specific research objective (due to initial failed recruitment, item nonresponse, or attrition) could be analyzed to assess whether there were any detectable biases in the resulting sample worth reporting.

- ***Statistical Analysis Tools for Estimating the Exposure-Outcome Relationships for the NCS:*** Innovative statistical methods can greatly enhance the power to detect relationships in the NCS. Examples include statistical design for add-on data collection activities, in which more detailed exposure assessment information is collected from small subsets of NCS participants; advanced statistical analysis tools that correctly assess the relationship between disease and exposure, even when the majority of NCS participants have imprecise measures of exposure (i.e., surrogate measures of exposure); and development of optimal outcome dependent design strategies for conducting resource efficient nested case-control studies from within the NCS cohort. A strong basis for these methods has already been established (Appendix C); however most of this methods development work was conducted under a simplifying assumption of a simple random sample. Additional work will need to be conducted to update these statistical design and analysis tools so that they will take the complex sampling design into consideration. The goal of this work should be to provide a series of statistical design and analysis tools that can be used to draw appropriate inferences from the NCS data.

11 REFERENCES

Chapter 1

Branum, A.M., et al. (2002). "The National Children's Study of Environmental Effects on Child Health and Development," *Environmental Health Perspectives,* November 21, 2002.

ICC Hypotheses Subcommitte (Draft). November 25, 2003. "National Children's Study: Improving Children's Environmental Health with High Quality Science - Formulating Hypotheses and Study Design". Presented to the NCS Advisory Committee on December 15, 2003.

The Children's Health Act of 2000. Public Law 106-310, 106th Congress, Section 1004, October 17, 2000.

"The National Children's Study: Scientific and Business Plan for the Research Investment," a collaborative study of the Department of Health & Human Services (NICHD, NIEHS, CDC), the U.S. Environmental Protection Agency, and other Departments and Agencies, November 2002.

Westat. "Sampling Strategies for the Proposed National Children's Study," technical report to the National Center for Health Statistics at the U.S. Centers for Disease Control and Prevention, October 25, 2002.

Chapter 2

Deming, W.E. (1953). *On the distinction between enumerative and analytic surveys*, J. Amer. Stat. Assoc., vol.48, pp.244-255

Hahn, G.J., and Meeker, W.O. (1993). *Assumptions for statistical inference*. The American Statistician 47:1-11.

Westat. "Sampling Strategies for the Proposed National Children's Study," technical report to the National Center for Health Statistics at the U.S. Centers for Disease Control and Prevention, October 25, 2002.

Westat. "Sampling Strategies for the Proposed National Children's Study," technical report to the National Center for Health Statistics at the U.S. Centers for Disease Control and Prevention, October 25, 2002.

Hartley, H.O. (1962). Multiple Frame Surveys. *Proceedings of the Social Statistics Section of the ASA Meetings,* pp. 203-206.

Lohr, S.L. and Rao J.N.K. (2000). *Inference from Dual Frame Surveys*, J. Amer. Stat. Assoc., vol. 95, pp. 271-280.

Chapter 3

National Children's Study Advisory Committee: Final Minutes from the National Children;s Study Assembly Federal Advisory (NCSAC) Meeting held on June 5-6, 2003 in Baltimore, Maryland.

Westat.(2002). "Sampling Strategies for the Proposed National Children's Study," technical report to the National Center for Health Statistics at the U.S. Centers for Disease Control and Prevention, October 25, 2002.

Chapter 4

Centers for Disease Control. (1992). "Sample Design: Third National Health and Nutrition Examination Survey. Series 2: Data Evaluation and Methods Research." National Center for Health Statistics, Centers for Disease Control and Prevention, U.s. Department of Health and Human Services, DHHS Publication No. (PHS) 92-1887.

Chapter 5

Branum, A.M., Collman, G.W., Correa, A., Keim, S.A., Kessel, W., Kimmel, C.A., Klebanoff, M.A., Longnecker, M.P., Mendola, P., Rigas, M., Selevan, S.G., Scheidt, P.C., Schoendorf, K., Smith-Khuri, E., Yeargin-Alsopp, M. (2002). *The National Chilren's Study of Environmental Effects on Child Health and Development*. Environmental Health Perspectives, November, 2002.

Centers for Disease Control. (1992). "Sample Design: Third National Health and Nutrition Examination Survey. Series 2: Data Evaluation and Methods Research." National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, DHHS Publication No. (PHS) 92-1887.

Centers of Disease Control. (1999). Table 2-1. Live Births by County of Occurrence and Place of Residence, by Race of Mother: United States, each State and County, and Specified Urban Places of 10,000 or More. Available online at http://www.cdc.gov/nchs/data/statab/t992x01.pdf

Diggle, P., Heagerty, P., Liang, K-Y., Zeger, S. (2002). Analysis of Longitudinal Data. Oxford University Press.

Kalton, G.,, Brick, J.M., and Le, T. (2003). Household Surveys in Developing and Transition Countries: Design, Implementation and Analysis, United Nations.

Sampling Strategies for the Proposed National Children's Study, WESTAT, 2002.

Sample Design: Third National Health and Nutrition Examination Survey, Vital and Health Statistics, Series 2, No. 113, 1992.

U.S. Environmental Protection Agency. (1995a). "Report on the National Survey of Lead-Based Paint in Housing: Base Report." Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, EPA 747-R95-003, April 1995.

U.S. Environmental Protection Agency. (1995b). "Report on the National Survey of Lead-Based Paint in Housing: Appendix I: Design and Methodology." Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, EPA 747-R95-004, April 1995.

Westat. (2002). "Sampling Strategies for the Proposed National Children's Study," technical report to the National Center for Health Statistics at the U.S. Centers for Disease Control and Prevention, October 25, 2002.

Chapter 6

Centers for Disease Control and Prevention. Health, United States, 2003. http://www.cdc.gov/nchs/.

Chapter 7

Bender, B., Ikle, D., DuHamel, T., Tinkelman, D. (1997). Retention of Asthmatic Patients in a Longitudinal Clinical Trial. *Journal of Allergy and Clinical Immunology*. Vol 99. No. 2.

Booth, A., Johnson, D. (1985). Tracking Respondents in a Telephone interview Panel Selected by Random Digit Dialing. *Sociological Methods and Research.* 14:53-64.

Brick, M., Collins, M., Davies, E., Chandler, K. (1997). Feasibility of Conducting Follow-up Surveys in the National Household Education Survey. *U.S. Department of Education. National Center for Education Statistics, NCES 97-335. Washington, DC.*

Buck, G., Lynch, C., Stanford, J., Sweeney, A., Schieve, L., Rockett, J., Selevan, S., and Schrader, S. (2003). Prospective pregnancy study designs for assessing reproductive and developmental toxicants. *Environmental Health Perspectives*. Accepted.

Callahan, M.A., Clickner, R.P., Whitmore, R.W., Kalton, G., Sexton, K. (1995). Overview of Important Design Issues for a National Human Exposure Assessment Survey. *J. Exposure Anal. and Env. Epid.* 5, no. 3: 257–282.

Cox, B., Massey, J., O'Connor, D. (2000). An Investigation of Response Rates in Random Digit Dialed (RDD) Telephone Surveys. *Personal Correspondence.*

Drew, J., Choudhry, G., Hunter, L. (1988). Nonresponse Issues in Government Telephone Surveys. *In R.M. Groves, D.P. Biemer, L.E. Lyberg, J.P. Massey, W.L. Nicholls, and J. Waksberg (Eds.), Telephone Survey Methodology,* 233-246. New York: John Wiley and Sons.

Robertson, et.al. 1999. The National Human Exposure Assessment Survey (NHEXAS) Study in Arizona-Introduction and Preliminary Results. *Exposure Anal. and Env. Epid.* 5, No. 3:

Strauss, W., Nishioka, M., Reiches, N., Ma, J., Lehman, J., Brown, V., Ryan, L., Spiegelman, D., Fitzmaurice, G., Loecke, D., Harezalk, J., Kalton, G., Brock, D., Choudhry, H., Clickner, R., Chu, A. (2003). Development of Exposure Assessment Study Design for the National Children's Study. *Final Report, Task Order 19. Contract no. 68-D-99-011, submitted to USEPA, Office of Research and Development.*

Turner, S., and Le Souef, P. (2003). Is Patient Dropout From a Longitudinal Study of Lung Function Predictable and Reversible? *Pediatric Pulmonology.* 35:29-33.

Whitmore, R.W., Byron, M.Z., Clayton, C.A., Thomas, K.W., Zelon, H.S., Pellizzari, E.D., Lioy, P.J., and Quackenboss, J.J. (1999). Sampling Design, Response Rates, and Analysis Rates for the National Human Exposure Assessment Survey (NHEXAS) in EPA Region 5. *J. Exposure Anal. and Env. Epid.* 9: 369–380.

Chapter 9

Heagerty, P.J., Zeger, S.L. (2000). Marginalized multilevel models and likelihood inference. Statistical Science 15: 1-19.

Hosmer, D.W. and Lemeshow, S. (2000). Applied Logistic Regression (2^{nd} ed.). John Wiley and Sons, Inc. New York.

Liang, K.Y. and Zeger S.L. (1986). Longitudinal data analysis using generalized linear models. Biometrika. 73:13-22.

Zeger, S.L. and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 42:121-130.