Persons using assistive technology may not be able to fully access information in this file. For assistance, e-mail niddk-cr@imsweb.com. Include the Web site and filename in your message.

APPENDIX A

SAMPLE SIZE DETERMINATIONS

The CPC Study was designed based on the assumption that studying the associations between many possible variables is of primary interest. For example, the subjects can be cross-classified according to presence or absence of a given symptom (S), such as pain exceeding a certain threshold, and presence or absence of a laboratory-based marker (R) such as elevated white count in EPS. The variable of interest can be a risk factor *(e.g.* alcohol consumption), another symptom (*e.g.* pain during urination), or the same symptom at a later point in time. For example, consider the binary cross-classification illustrated in Table 1.

Table 1. Cross-classification of Variable of Interest by Presence or Absence of Binary Symptom

	VARIABLE OF INTEREST **(R)**		
SYMPTOM	R_1 (present)	R_2 (absent)	
S_1 (present)	$P_{1\|1}$	$p_{2\|1}$	n_1
S_2 (absent)	$P_{1\|2}$	$p_{2\|2}$	n_2

Notation

Let the first subscript in $p_{i|j}$ represent the level of the column variable, and the second subscript represent the level of the row variable. That is, let $p_{1|1}$ represent the probability that the variable of interest is present, given that the symptom is present, and $p_{1|2}$ represent the probability that the variable of interest is present, given that the symptom is absent. Let n_1 represent the number of subjects with the symptom present, and n_2 represent the number of subjects with the symptom absent. Finally, let θ represent the true underlying odds ratio.

Assumptions

The following assumptions were made in arriving at the required overall sample size:

1. The proportion of subjects ($p_{1|2}$) with the risk factor, given the symptom is absent, is 0.5. (This estimate is conservative, because it maximizes the binomial variance).

2. At least 80% power is desired to test $\theta = 1$ against $\theta = 2.0$ and 2.5. The odds ratios correspond to the following proportions:

Θ	$p_{1\|2}$	$p_{1\|1}$
1.0	0.50	0.50
2.0	0.50	0.67
2.5	0.50	0.71

3. Two-sided tests with significance level 0.05 are desired.
4. The existence of clustering due to homogeneity of patients within Clinical Centers relative to simple random sampling requires a 1.5-fold increase in the necessary sample size to adjust for the clustering.

Results

Based on the above assumptions, but ignoring attrition, Table 2 summarizes the overall sample size ($n_1 + n_2$) required for a **binary risk factor** under specified study design parameters, given the assumed distribution of the symptom prevalence.

Table 2. Sample Size Requirements for Detecting Specified Odds Ratios Under Selected Study Design Characteristics: Type I Error Rate of 5%, Power of 80%

Underlying Odds Ratio (OR)	Percent Allocation for n_1:n_2	Sample Size for Symptom Present (n_1)	Sample Size for Symptom Absent (n_2)	Total Sample Size ($n_1 + n_2$)
OR= 2.0	10:90	110	918	1,028
	20:80	123	479	602
	30:70	140	335	474
	40:60	165	242	407
	50:50	197	197	393
OR= 2.5	10:90	69	684	753
	20:80	78	320	398
	30:70	90	207	297
	40:60	105	156	261
	50:50	126	126	252

Adjusted for clustering affect within Clinical Research Centers. The original sample size was expanded by a factor of 1.5 to adjust for the clustering.

Similarly, based on the above assumptions, but ignoring attrition, Table 3 summarizes the overall sample size ($n_1 + n_2$) required for a **continuous variable risk factor** under specified study design parameters, given the assumed distribution of the symptom prevalence.

Table 3. Sample Size Requirements for Detecting Specified Effect Sizes (Standard Deviation Units in Mean Difference) Under Selected Study Design Characteristics: Type I Error Rate of 5%, Power of 80%

Percent Allocation For n_1:n_2	0.2 s.d. Units	0.3 s.d. Units	0.4 s.d. Units	0.5 s.d. Units
10:90	3,273	1,455	819	525
20:80	1,842	819	462	295
30:70	1,403	624	352	226
40:60	1,228	547	308	197
50:50	1,178	524	296	190

Adjusted for clustering affect within Clinical Research Centers. The original sample size was expanded by a factor of 1.5 to adjust for the clustering.

TABLE #4

Distribution of Baseline Screening Visits, Follow-up Patient Contacts and Clinical Research Center (CRC) Workload by Study Months Under Specified Assumptions* (35 new patients/year/CRC)

Budget Year	Study Months	Baseline Screening	Follow-up Patient Contacts at Selected Months**						CRC Workload***			ANNUAL TOTALS
			6 (TI)	12 (CV)	18 (TI)	24 (CV)	30 (TI)	36 (CV)	Clinic Visits	Telephone Interviews	Total Contacts	
2	1-3	8.75							17.50	0.00	17.50	
	4-6	8.75							17.50	0.00	17.50	
	7-9	8.75	7.88						17.50	7.88	25.38	
	10-12	8.75	7.88						17.50	7.88	25.38	85.75
3	13-15	8.75	7.88	7.09					24.59	7.88	32.46	
	16-18	8.75	7.88	7.09					24.59	7.88	32.46	
	19-21	8.75	7.88	7.09	6.38				24.59	14.25	38.84	
	22-24	8.75	7.88	7.09	6.38				24.59	14.25	38.84	142.61
4	25-27	8.75	7.88	7.09	6.38	5.74			30.33	14.25	44.58	
	28-30	8.75	7.88	7.09	6.38	5.74			30.33	14.25	44.58	
	31-33	8.75	7.88	7.09	6.38	5.74	5.17		30.33	19.42	49.75	
	34-36	8.75	7.88	7.09	6.38	5.74	5.17		30.33	19.42	49.75	188.66
5	37-39		7.88	7.09	6.38	5.74	5.17	4.65	17.48	19.42	36.90	
	40-42		7.88	7.09	6.38	5.74	5.17	4.65	17.48	19.42	36.90	
	43-45			7.09	6.38	5.74	5.17	4.65	17.48	11.55	29.02	
	46-48			7.09	6.38	5.74	5.17	4.65	17.48	11.55	29.02	131.85
	CRC Total	105.00	94.50	85.05	63.79	45.93	31.00	18.60	359.58	189.29	548.87	
	Combined (6 CRCs)	630.00	567.00	510.30	382.73	275.56	186.00	111.60	2157.46	1135.73	3293.19	

*Baseline Screening of 8.75 new CPPS patients per quarter (35 per year) for 3 years of recruitment.

**TI: Telephone Interviews at 6,18,30 months; CV: Clinic Visit at 12,24,36 months.

***Total Clinic Visit workload includes 2 baseline screening visits for each new CPC patient.