Sample size calculation

The cohort study: We used two sources to estimate the incidence of oesophageal cancer in the area: the 1971 cancer registry data, which showed age-standardised rates for oesophageal cancer of 109/100 000 among men and 174/100 000 among women, and a much smaller active surveillance of 4192 adults in 1996 that reported age-standardised rates of 144/100 000 among men and 49/100 000 among women.[1;2] By using the age-specific esophageal cancer incidence rates in each of these two reports and the expected population structure for our cohort of 50 000, the expected number of cases among men and women per year and over 5 years is shown in table 1. As indicated, an average 5-year follow-up of the cohort of 50,000 individuals, which will take until 2012, will lead to between 500 and 1200 cases of esophageal cancer. If the study is continued until 2015, then we expect 400 000 person-years of follow-up and between 800 and 1920 ESCC cases in this cohort. For the purpose of power calculations, we adopted the conservative estimate of 500 cases occurring within our cohort over a median 5-year follow-up period. The initial follow-up of the cohort subjects during the pilot phase found a rate of 1.92 ESCC cases per 1000 person-years, equivalent to 480 cases in a 5-year follow-up of 50,000 individuals, similar to this 500-case estimate. Conducting nested studies for biological samples with 500 cases and 1000 controls, at a significance level of 0.05, will result in a 90% power to detect increased risks of ≥ 1.5 for common exposures (around 20%) and increased risks of ≥ 2.0 for uncommon exposures (around 5%) (Table 2).

Table 1: Expected number of ESCC cases in a cohort of 50,000 men and women, using rates from two previous studies in Golestan

| Age group | Mahboubi et al., 1973 | | Saidi et al., 2000 | |
	Age-Standardized Rates/10^5 Men: Women	Expected cases Men: women	Age-Standardized Rates/10^5 Men: Women	Expected cases Men: women
40-49	285: 350	16.9: 36.7	0:0	0: 0
50-59	550: 750	34.1: 75.5	0:59	0: 5.8
60-69	575: 880	27.4: 29.0	215:199	10.0: 6.8
70+	650: 820	14.4: 16.5	3225: 740	71.4: 14.9
Total/year		92.7: 157.7		81.4: 27.5

Total in 5 years		463.5: 788.5		407.0: 137.5

Table 2: Number of cases required to provide a 90% power to detect a main effect with a two-sided 0.05 level test, assuming two controls per case

	Proportion of population exposed			
Relative risk	5%	10%	20%	50%
1.5	1673	904	532	389
2.0	506	278	170	137
2.5	264	147	93	81
3.0	171	97	62	58

The case-control study: Calculations were done for various exposure rates. If the exposure rate among controls is 50%, $\alpha = 0.05$ (two-sided), control/case ratio equal to 2:1, 300 cases will provide us with 99% power to detect an odds ratio of 2.00 and a power of 81% to detect an odds ratio of 1.50. If the proportion of a specific exposure is 10% among our controls, 300 cases and 600 controls will provide us with 80% power to detect an odds ratio of 1.80. Therefore, we chose to recruit approximately 300 cases and 600 controls.

Reference List

1. Mahboubi E, Kmet J, Cook PJ, Day NE, Ghadirian P, Salmasizadeh S. Oesophageal cancer studies in the Caspian Littoral of Iran: the Caspian cancer registry. _Br.J.Cancer_ 1973;**28**:197-214.

2. Saidi F, Sepehr A, Fahimi S, Farahvash MJ, Salehian P, Esmailzadeh A _et al_. Oesophageal cancer among the Turkomans of northeast Iran. _Br.J.Cancer_ 2000;**83**:1249-54.