POPULATION UNDER SURVEILLANCE

When planning a surveillance program, it is important to consider the population size needed to obtain reasonably precise estimates. In this context, it would be necessary to take into account the most basic comparisons of rates in order to achieve the aims of the program. In general, this would concern evaluations of changes in rates over time and of population differences in rates. Two different approaches to determine the required population size are presented below, the first based on a hypothesis testing approach and the other on a confidence interval approach. The calculations are illustrated by a worked example.

Hypothesis testing approach

Under the hypothesis of a given annual percent change in the attack rate, this approach allows to calculate the necessary population size based on a Poisson probability function where the minimal number of events to be registered per year is given by the following relation:

Number of events per year = X / k =

$= 2 / k^3 * [(\phi^{-1} (1- \alpha/2) + \phi^{-1} (1-\beta)) / (t / 100)]^2$

where

X = indicates the number of events over k years;

α= significance level; 1-β= statistical power;

t = indicates the attack rate percent change per year;

ϕ^{-1} = is the inverse of the Poisson probability distribution

[http://en.wikipedia.org/wiki/Poisson_distribution].

For example, for an 80% probability (1-β) of detecting a 2% change in event rate per year over 5 years significant at the 5% level (α, two tailed test), the annual number of events needed is approximately 300:

Number of events per year = X / k =

$= 2 / 5^3 * [(1.96 + 0.84) / (2 / 100)]^2 = 314$

To give an example, the following table shows the numbers of events to be collected per year for an 80% probability of detecting a 2% or 1% change in attack rate per year over 10 years, significant at the

5% level (two tailed test), for men and women ages 45-74, for Coronary and Cerebrovascular events separately. In the table, to give an example, population sizes estimated for a low CVD incidence country (Italy) and a high CVD incidence country (Finland) are given. Coronary and Cerebrovascular attack rates used for the calculations derive from the Italian Progetto CUORE [URL http://www.cuore.iss.it/], and the Finnish National Cardiovascular Disease Register [Laatikainen T, et al. National Cardiovascular Disease Register, statistical database. URL http://www.ktl.fi/cvdr/].

In table 5, the column 'Events' shows the number of events to be collected per year to satisfy the chosen parameters; the two columns beside indicate the country specific crude attack rates used for estimating the minimal numbers; the next column shows the number of men and women to be taken under surveillance in the country specific population, calculated on the basis of events to be collected and country specific attack rates; following, the required total population size based on the number of men and women respectively, using the European standard population structure is reported; the last column shows the correspondent total population size to monitor after 10 years, under the assumption of a constant decrease, in order to maintain statistical power.

Confidence interval width approach

An alternative approach to the hypothesis testing for estimating the population size to monitor is based on the confidence interval width: the requirement could be to have a confidence interval that is not too wide. Given that the purpose of the surveillance is to estimate attack rate and change in attack rate over time rather than testing a predefined hypothesis, this approach might be appealing. It is mainly based on the balance between two competing parameters: the confidence level and the interval width. If the confidence level is increased, the interval width will also increase, which means less information about the true rate. Given the confidence level and the interval width, it is possible to determine the related minimal population size. In a large population or for incidence rates not too small, the Poisson probability distribution can be approximated by the Normal distribution; in this case, estimation of the minimal population size (N) can be calculated using the following relation:

$$N >= (2z_{\alpha/2})^2\, p(1-p) / w^2$$

where

p = attack rate estimate;

$p(1-p) = \sigma$ = standard deviation estimate;

α= significance level; in this context a factor specified by the confidence level, e.g. α=0.05/2 would correspond to a 95% confidence interval;
z = refers to the use of the standard Normal distribution for deriving probabilities;
w = the chosen absolute interval width.

For example, in a large population with an attack rate of 44.1 / 10,000, given the significance level of 5% (α, two tailed test), and an absolute interval width of 20% of the attack rate, the minimal population size needed is approximately 87,000:
$N >= (2*1.96)^2 *0.00441*(1-0.00441) / (0.00441*20/100)^2 >= 86,727$

Estimating the population size needed for monitoring time trends in event rates is important and the results may limit the number of possible areas able to produce stable trend estimates. What matters is the annual number of events, and not the population size; in high attack rate countries, smaller populations can be studied and in low attack rate areas larger ones would be needed. The limitations of using less than ideal sizes of populations for study could be reduced by:
i) accepting a higher threshold for the annual rate of change than those used in the example of 2% per year. This would be relevant to areas with low but rapid rates;
ii) increasing alpha and beta to lower the sample size. This would lower the power below 80% and/or increase α, the significance level, from 5% to 10%;
iii) pooling:

(a) results from age groups down to 25 (small effect on numbers);
(b) results from the age groups beyond 74 (large effect);
(c) combining data from both sexes (moderate effect);
(d) combining data from two or more geographically separate areas within one country establish trends, while studying them separately for other purposes;
(e) combining data within collaborative projects for centres in different countries, matched for certain characteristics such as initial event rates, risk factor trends, socio-economic characteristics, or health services.

While pooling data will increase numbers, it may conceal important information.
It is recommended that the minimum period of observation is one complete calendar year because of possible seasonal variations.

TABLE 5 Minimal size of low and high risk population under surveillance required for fatal and nonfatal coronary and stroke events, ages 45-74 years

		Attack rate (x 10,000)		Male and Female population required according to gender specific attack rates		Total population required using EU standard population structure		Total population required after 10 years under the assumption of continuous attack rate decrease	
Attack Rate percent variation (t %)	*Events*	*Men*	*Women*	*Male population*	*Female population*	*Total pop based on MEN*	*Total pop based on WOMEN*	*Total pop based on MEN*	*Total pop based on WOMEN*
2%									
Total Coronary Events Attack rates									
Italy	314	44.1	12.8	71,192	245,277	444,948	1,532,984	544,563	1,876,191
Finland	314	272.7	116.9	11,512	26,846	71,948	167,789	88,056	205,354
Total Cerebrovascular Accidents Attack rates									
Italy	314	33.5	20.3	93,718	154,658	585,737	966,611	716,873	1,183,017
Finland	314	112.0	61.2	28,044	51,317	175,276	320,730	214,517	392,536
1%									
Total Coronary Events Attack rates									
Italy	1256	44.1	12.8	284,767	981,110	1,779,791	6,131,937	1,967,964	6,780,251
Finland	1256	272.7	116.9	46,047	107,385	287,794	671,157	318,222	742,116
Total Cerebrovascular Accidents Attack rates									
Italy	1256	33.5	20.3	374,872	618,631	2,342,949	3,866,443	2,590,663	4,275,232
Finland	1256	112.0	61.2	112,177	205,267	701,104	1,282,921	775,229	1,418,560