Sample size calculation

The Schlesselman case control study formula was used for sample size calculation [26];

$$n = \frac{\left(z_\alpha \sqrt{2\bar{p}\bar{q}} + z_\beta \sqrt{p_1 q_1 + p_0 q_0}\right)^2}{(p_1 - p_0)^2}$$

Where,

$$\bar{p} = \frac{p_1 + p_0}{2}$$

$$\bar{q} = 1 - \bar{p}$$

$$q_1 = 1 - p_1$$

$$q_0 = 1 - p_0$$

n = number of subjects in each group

z_α = Corresponding to α (level of significance (95%)) = 1.96

z_β= Corresponding to β (Probability of type II error. Power of study is 80%, 1- β = 0.2)

p_0 = proportion of exposure among control groups (*prevalence of the polymorphism in general population without PCOS).

The exposure rate of allele frequency of SNPs among controls (30%–44%) was based on literature from other countries as no studies have been done in Sri Lanka [10, 30-32].

The estimated proportion of 38% was used in sample size calculation.

R = odds ratio associated with exposure

The literature indicated odds ratio (OR) of occurrence of any SNP among PCOS to be in the range of 1.1–2.1 [10, 30-32]. This study used an OR of 2.7 in its sample size calculation because Sri Lankans with anovulatory PCOS manifest severe symptoms at a younger age, with greater IR and a higher prevalence of metabolic syndrome than white Europeans [2].

Therefore, association of kisspeptin with PCOS is likely to be much higher among Sri Lankan PCOS women.

P1 = Proportion of exposure among cases was calculated using the following formula**;**

$$p_1 = \frac{p_0 R}{[1+p_0(R-1)]}.$$

The number of consecutive subjects with confirmed PCOS commencing from adolescence was 50, to which 10% was added to make allowance for non compliance / drop outs etc. Therefore, the number of cases was 55. By selecting double the number of controls per cases, the final total number of controls was 110.