Appendix.

Supplementary information about the performance of focused Scan test is showed in this appendix. Several focused cluster detection methods have been developed to identify clustering of high-incidence areas around the main polluting foci (e.g. Ramis et al. 2011). In this paper a spatial focuses-local version of the Scan test (Kulldorff, 1997; Kulldorff and Nagarwalla, 1995) was used. Unlike the typical focused-local spatial test, the focused-local spatial scan test (Puett et al., 2010) identifies the foci and the distance from foci that is most likely to reject the null hypothesis of equal risk in all regions.

In our case, the local-focused scan statistic is used to determine high-risk areas around a pre-determined set of points (pollutant sources). To understand the operation of the statistic, consider a study area A that contains a total of N regions. We assumed that the incidence of cancer in each region is distributed according to a Poisson model (Kulldorff, 1997). We select a priori set of focused point F_i (i=1,…,m) knowledge about the location of the hypothesized cluster.

Suppose Z a set of regions whose centroids are less than k kilometres from the foci F_i (circular windows). If the window includes the centroid of a specific census track, then this census track is included in the window. For each window, the spatial scan statistic tests the null hypothesis of equal risk of childhood cancer incidence for all regions against the alternative hypothesis that there exists an elevated risk of childhood cancer incidence within the scan window when compared with areas outside the window.

The following hypothesis can then be tested:

**H_0**: The underlying risk is equal inside and outside for all windows

**H_A**: There is at least one window Z around a foci F_i for which the underlying risk is higher inside the circle as compared to outside.

Set c(Z) be the observed number of cases in circle Z and n(Z) be the expected number of cases in circle Z under the null hypothesis. Let \( L_{H_A}(Z) \) be the likelihood under the alternative hypothesis that there is a cluster in circle Z, and let \( L_{H_0} \) be the likelihood under the null hypothesis. It can then be shown that:

\[
Z = \frac{L_{H_A}(Z)}{L_{H_0}} = \left( \frac{c(Z)}{n(Z)} \right)^{c(Z)} \left( \frac{n(Z)}{c(Z)} \right)^{n(Z)}
\]

Details, including derivations as a likelihood ratio test, have been given elsewhere (Kulldorff, 1997). The local-focused scan test is defined as

\[
\sup_{Z} \frac{L_{H_A}(Z)}{L_{H_0}}
\]

Where \( \Theta \) is the previous predetermined set of all possible circular windows centred in some foci F_i with radius less than k kilometres.

The region Z which the likelihood ratio is highest is named the Most Likelihood Cluster (MLC). The significance of MLC is obtained by Monte Carlo techniques. We use Monte Carlo simulation generated 9999 random replications of the data set to obtain the statistical stability and the Gumbel approach (Abrams et al., 2010). The procedure is implemented in SaTScan free software (v.9.3) (http://www.satscan.org/).

An illustrative example
An example can illustrate the operation of the test and the kind of results it yields. Assume, as shown in Figure 1a, that we investigate high incidence risk around three (m=3) focus \( F_i \) in the area \( A \) divided in \( N=50 \) regions.

To calculate the value of statistic \( \Lambda \) moving windows are used to scan this area. These windows are generically named \( Z \), and are centred in any Foci \( F_i \) with different radius. Figure 1b show examples of different windows. For each window \( Z \) the observed number \( c(Z) \) of cases is counted and compared with the expected number \( n(Z) \) using the ratio of likelihood \( \lambda_Z \). The statistics \( \Lambda \) is the supreme of all \( \lambda_Z \) and identify the window that comprises the MLC. In this example the statistic identified cluster centred on foci \( F_1 \). Figure 1c show the localization of MLC usually name \( Z^* \).

The \( p \)-value of MLC is obtained for permutational bootstrapping. The value of \( \Lambda \) is compared with 9999 random permutations of cases and population of each region (Figure 1d)