Incidence of hypospadias varies considerably across countries, ranging from 4 to 43 cases per 10,000 births. Environmental factors might explain these differences. The classical approach is to use case-control studies and questionnaires to identify these factors. However, this approach suffers from the unavoidable arbitrariness of the definition of controls and from recall bias.

New opportunities for environmental epidemiology arise with the increasing availability of public databases informing on local environmental exposures. These public databases can be mapped, using geographic information systems (GIS) to the addresses of geolocalized patients of interest. This can be used to compare environmental exposures by mapping the addresses of the cases and of the controls to the environmental geographic databases of interest.

A more straightforward approach is to choose randomly a set of places on the map, then compute the exposure at these places, and compare the exposure of the cases of interest to this reference. In other words, the idea is to define “virtual controls” instead of “physical controls” to evaluate the reference exposure. This is appealing because:

1. It avoids the costs of “physical controls” whose choice bears some arbitrariness
2. The choice of virtual controls (defined by a given algorithm) is more flexible

Common sense tells however that these virtual controls cannot be just randomly sampled on the map, as they would in this cases most likely fall in uninhabited places, while cases would be in majority in high population density spots. Instead of selecting arbitrarily a unique set of virtual controls, we sampled different algorithms for an optimal choice of “virtual controls”.

**OBJECTIVES**

To describe our methodology as a proof-of-concept of a “virtual controls” approach to search for environmental markers (factors) of a disease, applied here to hypospadias.

**METHODS**

1. The starting point is to define the population in which the virtual controls will be sampled. This population must be identical to the population from which cases are extracted. The basic idea is that, if the controls would have got the disease, they would have had the same probability of being selected than cases.

When the cases are recruited by clinical centers, the base population is harder to define because it depends on location and characteristics of these clinical centers. Our cohort included 8766 cases of hypospadias coming from 16 specialized centers. We defined several different populations from which we draw the “virtual controls”.

For example:
- In squares around each case
- In circles containing 1% of the cases around each clinical center.

We also considered in another sampling only the population living in rural locations. (thanks to INSEE definition [2])

**RESULTS**

For every set of “virtual controls”, we compared the distance to the closest vineyards between cases and controls. Here are the results for one set of “virtual controls” sampled in rural locations inside circles including 90% of rural cases around each clinical center.

<table>
<thead>
<tr>
<th>Range (m)</th>
<th>Cases</th>
<th>Controls</th>
<th>Odds ratio</th>
<th>p-value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-250</td>
<td>126</td>
<td>169</td>
<td>1.95</td>
<td>0.001</td>
<td>1.04</td>
</tr>
<tr>
<td>250-500</td>
<td>201</td>
<td>252</td>
<td>1.64</td>
<td>0.001</td>
<td>1.14</td>
</tr>
<tr>
<td>500-1000</td>
<td>252</td>
<td>397</td>
<td>1.40</td>
<td>0.0132</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Table 1 – Odds ratios comparing expositions to vineyards in the different rings for each virtual controls drawn in rural locations inside circles including 90% of rural cases around each clinical center.

**CONCLUSION**

The results are consistent with an association between hypospadias and vineyards. Nevertheless, we obtain similar results with other types of CLC, like forest. Further investigation is needed. We are also going to crosscheck those results with other databases, in particular the “parcel” database.

Computational optimization is needed to reduce computation times.

**REFERENCES**