We made a systematic literature review and found that many authors forgot about the fact that the diagnostic tests are not perfect, that is, the proportion of test positives (the so-called apparent prevalence) may differ from the true prevalence.

A shocking example selected from many similar ones:

Moujaber et al. (Int. J. Inf. Dis., 12, 500–504) studied Helicobacter pylori infection in Australia. They reported apparent seroprevalence rates by age group. The ELISA test they applied had 96.4% sensitivity and 92.7% specificity. The table shows that correcting for Se/Sp would have resulted in quite different estimates, leading to different conclusions.

<table>
<thead>
<tr>
<th>Age group</th>
<th>No. subj</th>
<th>No. pos.</th>
<th>App. prev.</th>
<th>95% CI</th>
<th>True prev.</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4</td>
<td>151</td>
<td>6</td>
<td>4.0%</td>
<td>1.5 - 8.4%</td>
<td>0%</td>
<td>0 - 1.2%</td>
</tr>
<tr>
<td>5-9</td>
<td>150</td>
<td>9</td>
<td>6.0%</td>
<td>2.8 - 11%</td>
<td>0%</td>
<td>0 - 4.1%</td>
</tr>
<tr>
<td>10-14</td>
<td>301</td>
<td>25</td>
<td>8.3%</td>
<td>5.4 - 12%</td>
<td>1.1%</td>
<td>0 - 5.3%</td>
</tr>
<tr>
<td>15-19</td>
<td>300</td>
<td>30</td>
<td>10.0%</td>
<td>6.8 - 14%</td>
<td>3.0%</td>
<td>0 - 7.5%</td>
</tr>
</tbody>
</table>

While Rogan and Gladen (1978) described a method for correcting the point estimate, there was a great uncertainty among the authors how to correct the confidence interval. We found papers reporting the true prevalence but giving a confidence interval for the apparent prevalence at the same time.

**Message**
- When estimating prevalence, never forget that the diagnostic test is not perfect, thus the proportion of test positives may considerably differ from the proportion of truly diseased.
- Be aware that sensitivity and specificity must be taken into account when estimating prevalence.
- There are valid methods – for point estimates as well as for confidence intervals – to estimate the true prevalence by correcting for the sensitivity and specificity of the diagnostic procedure.

When sensitivity and specificity are not known a priori but estimated from a sample, the uncertainty of their estimates will increase the variance of the prevalence estimate, and this must be taken into account in the confidence interval construction.

If the estimated sensitivity and specificity were regarded as fixed known values, the confidence interval for the prevalence would be unduly short, not maintaining the nominal level.

We described in Reiczigel et al. (2010) how to construct exact confidence intervals for the true prevalence under the assumption that sensitivity and specificity were known values.

The construction is based on the relation between the apparent and true prevalence, illustrated in the figure below.

But this confidence interval may not be valid if sensitivity and specificity are not known exactly but estimated from samples. Therefore we have worked out a new CI for that case as well. Since an exact CI seemed to be computationally infeasible, we proposed an approximate interval. The new method is described in Lang and Reiczigel (2014).

**References**

This research was supported by the Hungarian National Research Fund (OTKA K 108571) and by the Research Faculty Grant 2014 of the Szent István University, Faculty of Veterinary Science.