

# Statistical critique of a scientific paper

## Applied Biostatistics – Second Assignment

Lucia Montero Sanchis

Juraj Korcek

### 1 Briefly give the biomedical background for the paper. What question/hypothesis is being investigated?

The paper tries to find the risk factors associated with central lymph node metastases (CLNM) occurrence in patients suffering from papillary thyroid carcinoma (PTC). According to the authors, the need to identify the risk factors for CLNM in PTC patients arises from the fact that the ultrasonography, the main diagnostic tool for deciding the stage of PTC before surgery, has quite poor performance in detecting CLNM while the neck dissection approaches (CCND, PCND, CCLND), despite being way more accurate, are invasive and can cause additional medical complications. Therefore, a method that could reliably predict prior to surgery whether CLNM has occurred or not would be very valuable.

### 2 What data are collected (include how many individuals, what variables)?

Data was collected for 202 patients with PTC who had undergone thyroid surgery in one hospital between 2014 and 2016. From those, patients who were not first primary PTC or had other head and neck malignant were not considered, and 180 patients were considered for the analysis.

The target variable is CLNM, which is a binary variable (yes/no).

The explanatory variables include:

- **tumor size** - defined as the largest diameter, more or less (and equal) 10 mm; binary factor. It was measured by ultrasound.
- **multifocality** - whether there is more than one foci<sup>1</sup>; binary factor
- **Hashimoto's thyroiditis (HT)** - yes/no; binary factor. Identified by definitive histopathology.
- **lymphovascular invasion (LVI)** - yes/no; binary factor. Identified by definitive histopathology.

---

<sup>1</sup>Foci are cells located in a specific organ of the body that are notably different from the surrounding cells.

- **extrathyroidal extension (ETE)** - yes/no; binary factor. Identified by definitive histopathology.
- **age** - less or more (and equal) 45 years; binary factor
- **sex** - male/female; binary factor

### 3 What analyses were carried out? Are these analyses appropriate for the problem?

First, the numerical variables age and tumor size are transformed into binary variables by choosing 45 years and 10 mm as respective thresholds. This might imply a loss of information, especially since the reasons for choosing two bins for each variable and such thresholds are not explicitly stated – we can see however that in the case of age it seems like the average age was chosen.

The univariate and multivariate logistic regression were used to determine the significance of each factor in predicting CLNM. The Chi-squared test ( $\chi^2$ ) was used to check whether the distribution of the binary factors in CLNM patients corresponds to that of PTC patients. The significance level was set to  $p < 0.05$ .

The logistic regression would have been a valid approach if continuous independent variables were used instead of binary ones where possible. Logistic regression analysis are usually used when the input variables are quantitative. Since our input variables are not quantitative, this analysis does not seem to make much sense. When using multivariate logistic regression, the explanatory variables are assumed to be independent. Generally an independence test is carried out beforehand in order to verify this assumption, which is not mentioned in the paper. However, in this case the results were roughly the same as for the univariate logistic regression, so we can assume that there were no strong dependencies between the explanatory variables.

The Chi-squared test is indeed an appropriate tool for checking if a categorical variable is a factor for suffering from CLNM. If the given variable has the same distribution in CLNM patients as it has in PTC patients then one can conclude, given statistical significance, that the variable is not a factor in suffering from CLNM and vice versa.

### 4 What other analyses should have been done (or might have been done but not shown)? Explain.

As mentioned in the previous point, one possible option could have been doing logistic regression on continuous variables instead of binary factors.

Fisher's exact test could have been used as well to analyse the association of the proportions of input and output variables.

## 5 Is there any mention of power of the analyses? How would you go about trying to estimate power?

There is no mention of power of the analyses throughout the article.

In order to estimate the power of Chi-squared test we need to know following quantities:

- test size / significance level  $\alpha$
- sample size  $n$
- effect size  $d$
- degrees of freedom  $df$

The significance level has been fixed by the authors at 0.05. The sample size is 180. The minimum effect size that we would like to detect is up to us to choose. However, it is important to keep in mind that the smaller the effect size detectable, the smaller the power of the test. When it comes to degrees of freedom, as the distribution is fully specified by the NULL hypothesis (i.e. we do not have any parameters) the  $df$  can be set to  $\#classes - 1$ . In our case the number of classes is 2 and so the number of degrees of freedom is  $df = 1$ . Then we can apply `pwr.chisq.test[1]` from R's `pwr` package to estimate the power of the test given some effect size we choose.

When it comes to verifying whether the coefficients of the logistic regression are significantly different from 0, the Wald test has been most probably used[2]. Unfortunately, we do not know how to estimate its power. Very likely the authors have not provided us with enough information to be able to do so anyway.

## 6 What conclusions do the authors draw? Are these conclusions substantiated by the results? Explain.

The authors conclude that multifocality, HT and LVI are the risk factors predicting CLNM in PTC patients, which seems to be a sound conclusion according to the results of both univariate and multivariate logistic regression analysis.

They also state that, unlike the other studies on this matter, this study did not identify sex, age and tumor size to be sufficiently predictive. However, from our point of view, this is caused by a couple of deficiencies in this study.

One of the issues that we have identified is that the age is split only into 2 bins that are very broad which can cause that the effect of younger age on the CLNM stays hidden.

When it comes to sex and tumor size, we believe that the sample sizes for one of the factor levels are too small ( $\sim 30$ ) which results in high probability that there is a difference only because of chance which might cause the corresponding p-values to be artificially high.

The claim that the rate of CLNM was higher in multifocal patients seems to be confirmed by the results, as the Chi-squared test comparing multifocality distributions of PTC patients and PTC patients with CLNM is below the significance level of 0.05.

## References

- [1] R Documentation - `pwr.chisq.test` function. <https://www.rdocumentation.org/packages/pwr/versions/1.2-2/topics/pwr.chisq.test>. Accessed: 19.05.2018.
- [2] IDRE Stats - Logistic Regression — SPSS Annotated Output. <https://stats.idre.ucla.edu/spss/output/logistic-regression/>. Accessed: 19.05.2018.