ADDENDUM

Statistical approach

Introduction

In order to establish the statistical approach applied in this thesis, the data were firstly examined for characteristics that are known to affect the suitability of specific tests. These included sample size, normality, the characteristics of the intervals between measurements, and missing data, as well as the occurrence of the multiple comparisons problem (Midway et al., 2020; Mishra et al., 2019).

Sample size. The sample size has implications on the power of an analysis. Depending on the analysis applied and the level of significance searched, a minimum sample size will be required. Accordingly, formulas are available that allow to calculate the minimum sample size based on the power aimed, provided a level of significance (the alpha) and effect size are given (Jones et al., 2003). The effect size, in its turn, is usually anticipated based on previous studies. Against the desired effect size, the sample may be constrained by a number of factors, such as cost, time, and convenience of the data collection. A situation where the availability of the data may limit the sample to a number that may compromise statistical hypothesis testing is the study of rare conditions. In order to undermine this problem, some strategies may be used. An important one is to favour longitudinal analysis and allow participants to stay in the study as long as possible (Day et al., 2018). This increases the amount of data collected and allow the use of repeated measures analysis, which may result in a reduction of 30% in sample size requirements as opposed to change in score analysis (Fitzmaurice et al., 2008). Avoiding dichotomising continuous variables is also recommended, as rates may be more sensitive than means (Day et al., 2018). In conformation to these recommendations, in the present thesis, besides the transversal study, a longitudinal study was performed with the PCA patients and data were collected until the patients could produce reliable tests. For the purpose of the longitudinal analysis of some functions in which a floor effect was evident (detailed in Chapters 6 and 7), an epoch was selected in which change could still be detected.

The multiple comparison problem. When multiple inferences are made from the same dataset, the chance of finding a positive result due to reasons such as random sampling error is significantly increased, so that individual p-values will no longer reflect the chance false positive. This is called the multiple comparison problem. The solution is usually to reduce the alpha and increase the significance threshold (Midway et al., 2020). Amongst the several strategies developed for this purpose is the Bonferroni correction, which divides the individual significance level by the number of inferences made (Armstrong, 2014). It is known to be very conservative, as it leads to a high rate of false negatives. In contrast, liberal multiple comparisons tests have less strong adjustment. When choosing a multiple comparisons-test it is paramount to decide whether the data are parametric or non-parametric and whether the comparisons are planned or unplanned. For planned comparisons of parametric data, Bonferroni or sequential Bonferroni may be appropriate, but this is not the case of planned comparisons of non-parametric data - for which other tests, such as the common Mann-Whitney-Wilcoxon U test are recommended – neither of non-planned comparisons (Midway et al., 2020). A consequence of applying multiple comparisons tests is to increase the number of false negatives. Depending on the cost of missing a possible positive result, adjustment for multiple comparisons may rather not be performed, provide the risk of a false positive is acknowledged (McDonald 2014). Moreover, established approaches for multiple testing correction are not applicable to linear mixed models, although strategies have been developed to this problem in the context of the millions-of-markers genome wide associated studies (Joo et al., 2016). In the present study, correction for multiple comparisons was not applied.

Normality. When choosing the statistical approach to be applied to a specific analysis, it is paramount to know the structure of the data, i.e., whether the data follows a normal or Gaussian distribution, for this has consequence to the suitability of statistic tests. If the data are not normally distributed, statistical procedures such as correlation, regression, and t tests are not suitable (Ghasemi and Zahediasl, 2012), whereas rank-based non-parametric methods may be used (Pek et al., 2018). In this cohort, data corresponding to visual field sizes and cognitive scores were not normally distributed, as checked with Kolmogorv-Smirnov and Shapiro-Wilk tests in SPSS 25. The statistical approach was set as following.

Cross-sectional study:

To investigate the association between the cloverleaf effect - a measure of reliability of visual fields - and the presence of hemianopia the Spearman's rank correlation coefficient was used. The Wilcoxon signed rank test was used for comparison of the sizes of visual fields to both static and kinetic stimuli, as well as to the comparison of visual fields acquired on the Octopus at two different target-velocities. Comparison of the dissociation to the target-velocities between PCA patients and controls was performed using Mann-Whitney U Test. For the correlation of neglect and hemianopia Pearson's chi-squared test was applied. The latter test was performed in Stata16, whereas the remainder of the tests applied in the cross-sectional study used SPSS 25.

Longitudinal study:

The endeavour of this analysis was to model the progression the visual fields over time, showing intrasubject trajectories, according to the hemifield (whether in the more or less affected baseline at baseline), as well as differences between subjects. The data presented a further challenge represented by the occurrence of multiple missing data points, a common problem in longitudinal research with human subjects (Krueger and Tian, 2004). These problems have to be dealt with in a sequential manner. One such solution is mixed models, which consider, simultaneously, fixed and random effects.

Multilevel mixed-effect models are models that vary at more than one level (Bryk et al., 2002) (hence also called “hierarchical models”, “nested models”, random-effects model). Of note, longitudinal information may be seen as a special case of hierarchical model, with repeated measures nested within subjects (Maia et al., 2010). In this study, in the first level are the hemifields with repeated measures, and in the second level are the hemifields with a single predictor, the hemianopia status (e.g., the more or the less affected hemifield). The latter is fixed, whereas time is random.

The same approach was applied to the longitudinal study on cognitive scores, which were set as dependent of time and hemianopia status.

Bibliography

Armstrong, R.A., 2014. When to use the Bonferroni correction. Ophthalmic Physiol Opt. 34**,** 502-8.

Bryk, Stephen W. Raudenbush, Anthony S., 2002. Hierarchical linear models: applications and data analysis methods (2. ed., [3. Dr.] ed.). Thousand Oaks, CA [u.a.]: Sage Publications.

Day, S., et al., 2018. Recommendations for the design of small population clinical trials. Orphanet J Rare Dis. 13**,** 195.

Fitzmaurice, G., et al., 2008 (editors). Longitudinal Data Analysis. Boca Rhaton: Chapman & Hall/CRC Press.

Ghasemi, A., Zahediasl, S., 2012. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. 10**,** 486-9.

Jones, S.R., Carley, S., Harrison, M., 2003. An introduction to power and sample size estimation. Emerg Med J. 20**,** 453-8.

Joo, J.W., et al., 2016. Multiple testing correction in linear mixed models. Genome Biol. 17**,** 62.

Krueger, C., Tian, L., 2004. A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. Biol Res Nurs. 6**,** 151-7.

Maia, J. A. R, et al., 2010. The challenge of longitudinal information: a" guided tour" on hierarchical modeling, tracking and missing information with a set of data from Muzambinho study. Brazilian Journal of Physical Education and Sport 24(3):413-431.

McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland, 254-260.

Midway, S., et al., 2020. Comparing multiple comparisons: practical guidance for choosing the best multiple comparisons test. PeerJ. 8**,** e10387.

Mishra, P., et al., 2019. Selection of appropriate statistical methods for data analysis. Ann Card Anaesth. 22**,** 297-301.

Pek, J., Wong, O., Wong, A.C.M., 2018. How to Address Non-normality: A Taxonomy of Approaches, Reviewed, and Illustrated. Front Psychol. 9**,** 2104.