The ABC of reporting statistical analyses in the BJD: Always Be Clear

Loes Hollestein[1,2], Jo Leonardi-Bee[3], Serigne Lo[4], Saharon Rosset[5], Noam Shomron[6]
[1] Department of Dermatology, Erasmus MC University Medical Center, Rotterdam, The Netherlands
[2] Department of Research, Netherlands Comprehensive Cancer Center (IKNL), Utrecht, The Netherlands
[3] Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK
[4] Melanoma Institute Australia, University of Sydney, Sydney, NSW, Australia
[5] Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel
[6] Faculty of Medicine and Edmond J Safra Center for Bioinformatics, Tel Aviv University, Tel Aviv, Israel

The British Journal of Dermatology (BJD) editorial board now includes five associate editors who review the methodological reporting of articles submitted to the BJD. In order to assure successful publication of high quality research submitted to the BJD it is of utmost important that methods and results are described in sufficient detail by authors. The BJD author guidelines, editorials on reporting and reporting guidelines all contribute to achieving this goal [1-3]. Here, we would like to highlight a few general and study-specific issues.

Always Be Clear about the study design

The design is the most essential part of conducting a study. Most flaws in clinical research are a result of poor planning. Inappropriate study design cannot be fixed during the analyses phase and does therefore not lead to an answer to the research question. The statistical analyses are dependent on correct study design. A clear and detailed research question should lead to determining the appropriate outcome measure. When designing the study, the outcome measure and how and when the outcome should be measured, determines the type of analyses that will be conducted at the end of the study.

Always Be Clear about the purpose of each analysis

The aim of describing the methods is that the knowledgeable reader with access to the original data has sufficient detail to verify the reported results [4]. It is thus not sufficient to sum up all statistical analyses used, but the authors need to specify which test was used for which purpose. For example: 'a two-sample independent t-test was used to compare continues data between 2 groups', tells the reader what can be read from any statistical textbook, but doesn't tell the reader for which comparison the authors used a t-test and if an independent sample t-test was thus the appropriate test. The author needs to specify for which study specific aim the statistical test was used, e.g. 'a two-sample independent t-test was used to compare the body weight between patients and controls'.

Always Be Clear when reporting the results

Results should be described in such a way, that the results can be incorporated into other analyses [4].

To re-use reported data for meta-analyses, requires that numerators and denominators of percentages, risks, odds and hazards ratios are reported. Also, reporting p-values only is not sufficient: descriptive statistics of the groups, sample size, effect size and a measure of precision (standard error or 95% confidence interval) should be reported. In addition, actual p-values should be reported, rather than referencing p-values to a critical value (e.g. p=0.006 instead of p<0.05).

Clinical Trials: Always Be Clear about the analysis plan and interpretation of results

For sake of transparence, authors should develop a statistical analysis plan, a prospective document that specifies how trial data will be analysed [5]. The analysis plan could either be a stand-alone document or be integrated into the protocol. The document should be signed off before blindness had to be broken, and before any analysis has started. Any deviations from the signed-off plan should be clearly explained.

Adherence to the CONSORT reporting guidelines facilitates transparency of reporting, genuine and accurate interpretation of results [6]. In most of the clinical trial papers submitted to BJD in 2017-2018, the CONSORT 2010 checklist, which is only appropriate for the individually randomised, two groups, parallel trials, was completed[6]. Extended CONSORT checklists have been developed for 1) other designs such as cluster randomized trial, non-inferiority and equivalent trial, or pilot study and 2) other interventions including non-pharmacologic treatment interventions, herbal medicinal interventions or acupuncture interventions. Given the continuous evolving nature of the CONSORT guidelines, authors are strongly recommended to visit the CONSORT website to be sure the most up-to-date checklist is used.

Epidemiology: Always Be Clear about assumptions in regression analyses

Regression analyses are frequently used in epidemiological studies. Authors need to state in the methods, how the model assumptions were tested, if they were met and (if applicable) what has been done when the assumptions were not met. For linear regression analyses these assumptions include: independence of observations, linear relation between covariate and outcome, normal distribution of the residuals (differences between the observed and predicted values), the same standard deviation of the residuals across every value of the predictor (homoscedasticity) [7]. The linearity assumption also applies to other regression models (e.g. Logistic, Cox, Poisson). Non-linear relationships can be included in regression models by including non-linear functions (most accurate) or categorize continues variables (simple, easy to interpret). When a Cox proportional hazards model is applied to survival data, the proportional hazards assumption should be tested [8].

Genomics: Always Be Clear about your analyzed DNA and RNA data given the very 'noisy' background

Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS) methods are used to detect causative DNA mutations in an individual. These procedures provide a long list of variants in the DNA of the tested individual. The list of variants is dependent on: (i) the depth of the run. For example, an average of 100X coverage over the DNA would generate more variants than a 20x coverage read; (ii) the similarity to the reference DNA. For example, an African American DNA aligned against a European reference DNA would generate a vast amount of non-matching variants. These and other parameters affect the probability of locating a detrimental causative mutation and should therefore be reported in the methods.

Changes in cellular RNA levels possesses information on the state and function of samples. Most commonly used is the expression analysis of mRNA using sequencing technologies (RNA-seq), which allows biologists to identify gene and transcript expression under two or more conditions, and also to identify new genes and new splice variants. RNA-seq experiments must be analyzed with robust, efficient and statistically principled algorithms. Fortunately, the bioinformatics community has developed mathematical and statistical tools for RNA-seq processing presented as downloadable software or online tools, which must be specified in the methods section. Many other sophisticated procedures prior to data generation increase biological insight. For example, CLIP (cross-linking immunoprecipitation) is used in order to analyze protein interactions with RNA [9]. These should be clearly described in the manuscript.

Systematic Reviews and Meta-Analyses: Always Be Clear about the analysis methods for pooling studies

The choice over using a fixed or random effects models need to be declared within the methods section. Random effects models should be the default as this model allows for the magnitude of effect for the treatment or association to vary by study and patient level characteristics. The fixed effect model should only be used where there is a strong rationale that there could be a single magnitude of effect for the treatment or association being investigated.

It is important that the variation between studies, known as heterogeneity, is fully investigated [10]– two methods commonly used are subgroup analysis and sensitivity analysis. Subgroup analysis is used to determine whether the pooled effect size varies between subgroups, for example high versus low methodological quality, using a statistical test. In contrast, a sensitivity analysis compares the magnitude of the pooled effect between a restricted set of studies, for example where poorer methodological quality studies are excluded, and the magnitude of the pooled effect from the overall meta-analysis. It is important that the characteristics to be investigated are biologically plausible and specified *a-priori*. The results from subgroup and sensitivity analyses should be only treated as exploratory since there is the potential for an association to be found purely by chance or due to other confounding factors.

Translational Research: Always Be Clear about accounting for multiplicity in hypothesis testing

Most studies do not test a single hypothesis, but rather cast a wide net for of interesting associations. It is absolutely not valid to perform all tests at the nominal level of significance required (which is typically 0.05) and declare as significant associations that pass this threshold. This invalidity is easy to state and explain, since performing 50 tests at level 0.05 essentially guarantees some findings will be made, even if there is nothing to find (from the definition of what test level means). What should be done to correct it may be more scenario-dependent and in some cases can be hard to figure out. The simplest adjustment is the Bonferroni correction (dividing the alpha-level by the number of tests). In some cases the Benjamini-Hochberg (BH) type methods for controlling the false discovery rate (FDR) can also be appropriate [11]. More advanced aspects include accounting for exploratory analysis, and dividing the tests into "primary" and "secondary" analyses, which can drive complex, but still valid, testing schemes [12].

Always Be Clear about the use of reporting guidelines

Reporting guidelines include many of the aforementioned statistical and methodological items, which should be described in your manuscript [3]. All recommended reporting guidelines can be found in the BJD author guidelines [2]. The use of reporting guidelines will contribute to manuscripts with transparently reported methods and results, which are always clear!

References

1 Hollestein LM, Nijsten T. Guidelines for statistical reporting in the British Journal of Dermatology. *Br J Dermatol* 2015; **173**: 3-5.

2 Wiley. Author Guidelines of the British Journal of Dermatology, http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2133/homepage/ForAuthors.html [accessed on 30 January 2018]

3 EQUATOR network. Enhancing the QUAlity and Transparency Of health Research (EQUATOR), http://www.equator-network.org/ [accessed on 14 April 2015]

4 Lang T, Altman D. Basic Statistical Reporting for Articles in Biomedical Journals: The Statistical Analyses and Methods in the Published Literature'or 'The SAMPL Guidelines', http://www.equator-network.org/wp-content/uploads/2013/03/SAMPL-Guidelines-3-13-13.pdf [accessed on 30 January 2018]

5 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). E9: Statistical principles for clinical trials, http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html [accessed on 5 april 2018]

6 Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; **340**: c869.

7 Barker LE, Shaw KM. Best (but oft-forgotten) practices: checking assumptions concerning regression residuals. *Am J Clin Nutr* 2015; **102**: 533-9.

8 Kleinbaum DG, Klein M. *Survival Analysis: A Self-Learning Text*. New York, USA: Springer Science, 2005.

9 Zhang Z, Xing Y. CLIP-seq analysis of multi-mapped reads discovers novel functional RNA regulatory sites in the human transcriptome. *Nucleic Acids Res* 2017; **45**: 9260-71.

10 Dias S, Sutton AJ, Welton NJ et al. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Making* 2013; **33**: 618-40.

11 Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J Clin Epidemiol* 2014; **67**: 850-7.

12 European Medicines Agency (EMA). Guideline on multiplicity issues in clinical trials (draft), http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/03/WC500224998.pdf [accessed on 10 April 2018]