Rare Disease Analyser

User Manual

Authors:
Konstantin Emil Thiel (main developer)
Johan Verbeeck (GPC expert)

1. Introduction and Requirements

The Rare Disease Analyser (RDA) is a web-application for statistical analysis of **longitudinal data** collected in studies with a small number of participants, a situation frequently encountered in rare disease research. Currently, the RDA features two statistical analysis and testing methods:

(i) Generalized Pairwise Comparisons: **GPC** [1]
(ii) Nonparametric Analysis of Longitudinal Data in Factorial Experiments: **nparLD** [2]

In a simulation study with only 6-7 subjects in 2 treatment groups and 4 different time points, both methods were found to control type-I error rates well [3]. The RDA provides GPC and nparLD for **univariate 2-group comparisons** of **ordinal** and **metric** variables. The study design can be either a standard 2-group **parallel design** or a 2-group **cross over design** with **2 study periods**, where participants switch treatment group after the first study period. The RDA computes p-values for these 2-group comparisons. For details regarding input data, null-hypotheses, and interpretation of the results, please read the subsequent sections of this user manual.

2. Input Data and Data Upload

The first step in using the RDA is data upload in the second tab of the app. Input data must be provided as a file of **comma-separated values** with at least **4 columns**. Each row in the file contains one observation collected in a longitudinal study (long data format). The file must not contain row numbers, but it must contain **column names in the first row**. Moreover, input data has to be complete for the nparLD, i.e., no missing values should occur, otherwise, the analysis method might fail. Additional input data specifications are given in the subsequent sections on the analysis methods.

In Figure 1, we uploaded a dataset and selected a data separator, i.e., the character separating individual columns, as well as the decimal separator, i.e., the character separating the integer from the decimal part in floating-point numbers. Once the correct data separator and decimal separator are selected, the dataset appears as a table below the Data Upload panel. Importantly, we now choose the correct variables (columns) in our file below the **Variable Selection** header:

- **Outcome**: An ordinal or metric variable, which is the outcome of interest to be compared between the 2 treatment groups.
- **Group Factor**: A variable that distinguishes the 2 treatment groups. Hence, there must be exactly 2 different values in this column.
- **Time Factor**: An integer-valued variable that distinguishes the different time points at which data was collected.
- **Subject**: A variable that uniquely identifies subjects in the dataset.

Note that the different values of the Time Factor must apply to all subjects in the dataset. Once everything is set-up, we proceed by choosing a Statistical Analysis in the third tab of the RDA. In the subsequent sections, the statistical analysis methods are described in detail.

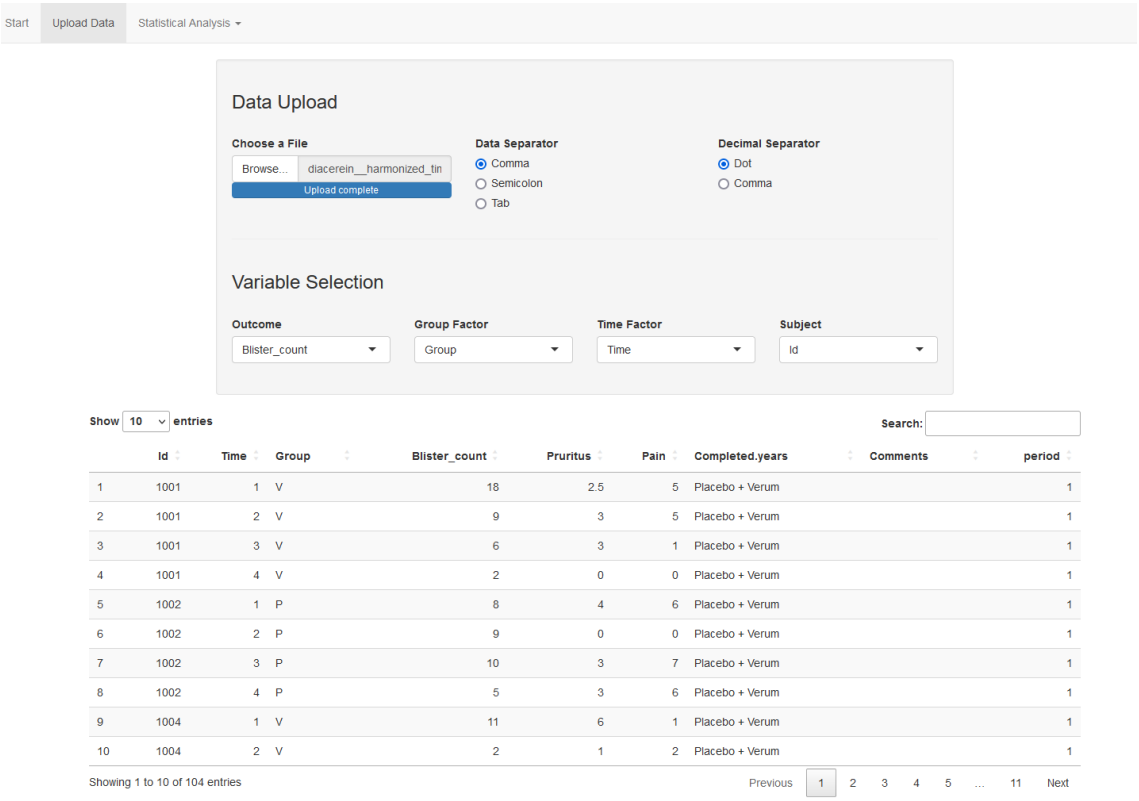


Figure 1: Successful upload of a dataset with columns "Id", "Time", "Group", "Blister_count", "Pruritus", "Pain", "Completed.years", "Comments", and "period".

3. nparLD

nparLD is a **nonparametric** method to analyse longitudinal data using so-called **relative effects**. These relative effects are probabilities of observations from a certain time point and/or group tending to be larger/smaller than observations from some reference distribution [2]. When there is no tendency, then the relative effect equals 1/2. In our context, reference distributions are those obtained by averaging over all time points, over both groups, or over all time points and groups, respectively. Using estimators of these distributions, nparLD provides tests for (i) a **group effect**, (ii) a **time effect**, and (iii) and a group-time **interaction effect**. Frequently, the latter is of particular interest when treatment groups are compared over time.

To use nparLD in the RDA, we choose the respective page in the dropdown menu that appears when clicking on the Statistical Analysis tab. Note that the dataset must contain at least two rows for each subject in your study (collected at different time points). On the nparLD panel, we choose the study design that best describes the dataset. In case of a cross-over study, the RDA expects another variable in the dataset that **distinguishes both treatment periods**, the period variable. In case of a single trial period, e.g., in a classical parallel design study, the period variable is not required.

In Figure 2, we apply nparLD to a single trial period dataset. The RDA plots relative effects of the outcome variable selected earlier on the left hand side and prints respective tests statistics and p-values

on the right hand side. Note that statistical tests are based on the so-called ANOVA-type test statistic. This test statistic is known for its **good type-I error control**, even in small-sample scenarios. The first row of the test statistics table refers to the group effect, the second row to the time effect, and the third row to the interaction effect. In our case, there is a significant interaction effect at the 5% level. This interaction effect is reflected by differing time trends of the relative effects in our treatment groups "P" and "V".

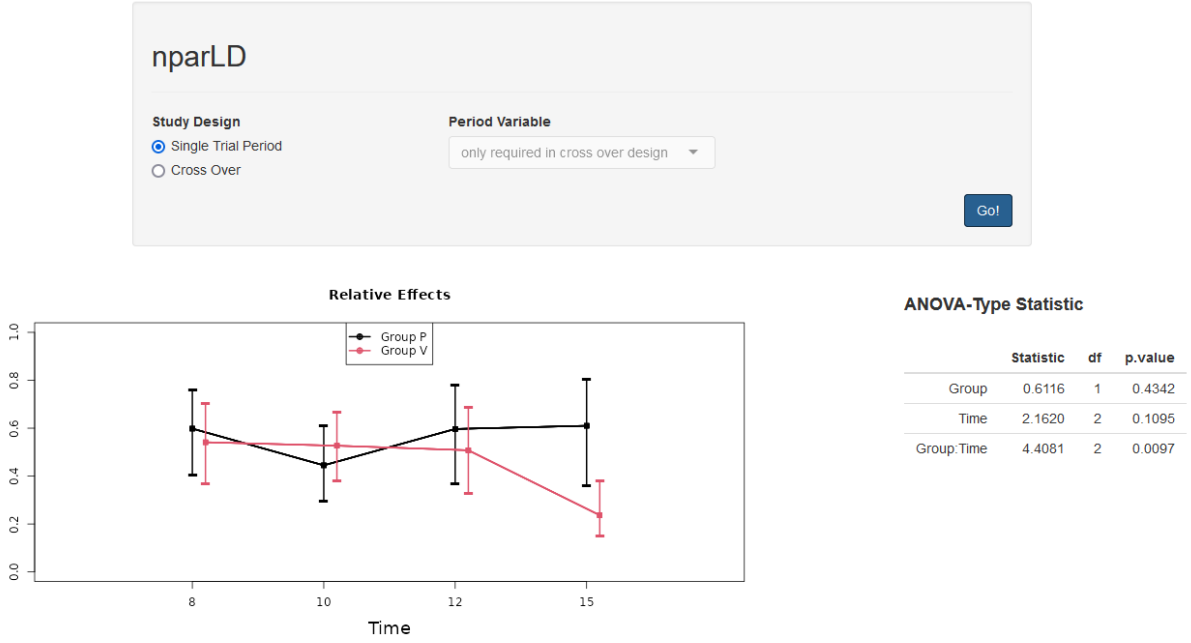


Figure 2: Application of nparLD to a dataset with a single trial period. The RDA plots relative effects of the outcome variable for all time points and treatment groups alongside with individual 95% confidence intervals. ANOVA-type statistics are provided for small-sample inference. In this case, the outcome variable Is called "Blister_count" for period 2 in the example dataset.

The RDA uses the nparLD tests implemented in the respective R software package (version 2.2) [4]. In cross-over designs, the RDA simply conducts the analysis for both treatment periods separately (results are displayed one below the other). This is because the deployed nparLD version cannot handle simultaneous analysis of cross-over studies. Hence, the RDA cross-over feature can be regarded a convenient way of analysing data from such a study using only one web browser window. Note that adjustments for multiple testing must be done by users separately, i.e., these adjustments are not performed by the RDA.

4. GPC

Generalized pairwise comparisons is a flexible non-parametric method, proposed for the analysis of multiple outcomes in a two-arm clinical trial [1]. As a longitudinal outcome is a specific form of multiple outcomes, GPC can be applied to it as well. The method extends the Mann-Whitney test to pairwise comparisons of multiple outcomes and allows combining any number and type of outcomes. The method forms all possible pairs of subjects with one subject from each treatment arm and decides which of the subjects has the most favourable outcome according to a pre-defined algorithm. This algorithm can evaluate all pairs on each of the multiple outcomes (non-prioritised GPC) or evaluate the pairs based on a prioritised list of the multiple outcomes (prioritised GPC), meaning that if it cannot be decided

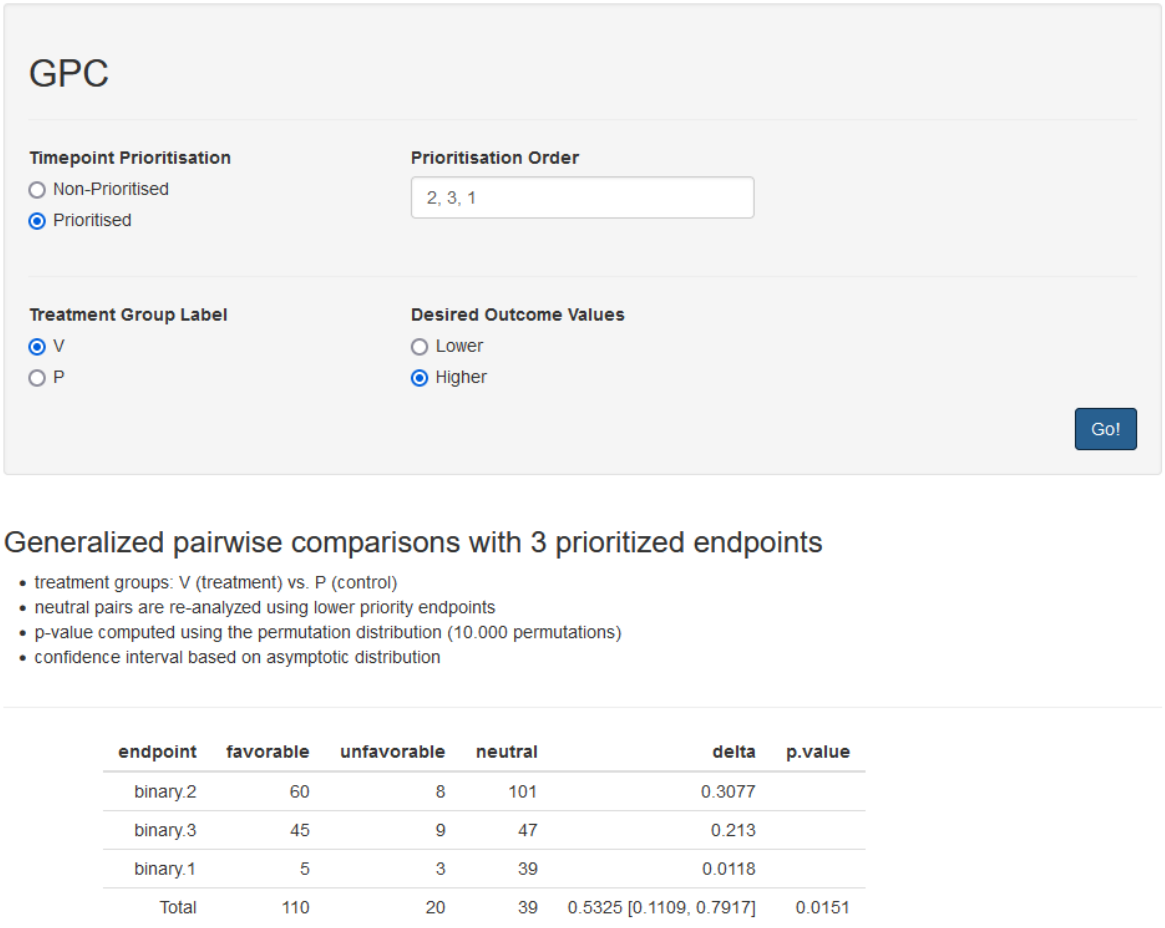
which of the subjects has the most favourable outcome on the highest priority outcome, the comparison moves to the next outcome in the priority list.

While several hypothesis tests for detecting a treatment difference have been proposed for GPC [7], only the permutation test controls the type I error in very small samples [8]. The size of the treatment effect in GPC is expressed by the net treatment benefit, which corresponds to the net probability of a favourable outcome for a patient in the experimental arm compared with a control patient. It takes values between -1 and 1, where positive values indicate a beneficial treatment effect, negative values reflect harm, a no treatment effect results in a value of 0. The confidence interval of the net treatment benefit is based on asymptotic u-statistics, which is equal to a bootstrap, but remains suboptimal for very small samples (<30 subjects).

Other variations of the GPC algorithm exist, such as the matched GPC, which compares the outcomes within the same subject. Although it seems natural to analyse data from a cross-over design trial with a matched GPC, it has been shown that ignoring the matched design still leads to asymptotically valid results [9]. There is no requirement for data to be complete in a GPC analysis. To use GPC in the RDA for longitudinal outcomes, we select it in the Statistical Analysis dropdown menu. The GPC method in the RDA can handle both single trial period studies and cross-over studies. However, when working with cross-over data, we have to ensure that time points are named equally in the dataset for both trial periods. That is, in the time variable, the first time point in period 1 must have the same label as the first time point in period 2, and so on.

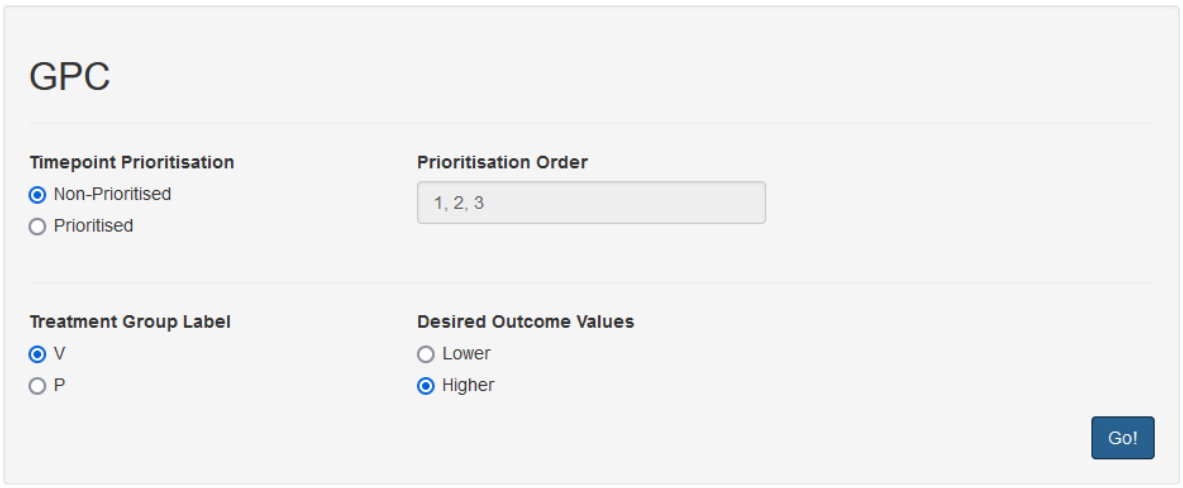
In order to proceed with the analysis, we select whether we want to prioritise time points. If we select prioritisation, then the order of the timepoints need to be specified in the Prioritisation Order box by descending order of priority, i.e. the first timepoint receives the highest priority and the last timepoint the lowest priority. The timepoints entered in the box should match the values in the dataset and should be separated by a comma followed by a single space. Next, we determine the Treatment Group Label in our earlier selected Group Factor. Recall, that there must be exactly two labels and here we tell the RDA which of these two labels denotes the treatment group (as opposed to the control group). Finally, we indicate if higher values of the outcome variable are most favourable or lower values by selecting the appropriate option under Favourable Outcome Values. When opting to perform a non-prioritised GPC, the Prioritisation Order box cannot and should not be completed.

The GPC version implemented in the RDA are based on the R software package “BuyseTest” (version 2.3.11) [5]. Confidence intervals are based on asymptotic u-statistics and p-values on a permutation test. While the permutation test is exact for small samples, at this stage, the confidence intervals are suboptimal for small samples (<30 subjects) and should be interpreted with caution. Figures 3 and 4 demonstrate the output of the RDA for a prioritised GPC and a non-prioritised GPC respectively, on the deployed example dataset for the dichotomized blister outcome (40% reduction yes/no) on 3 timepoints. The tables provide detail on how many pairs contributed at each timepoint to the overall net treatment benefit. For instance, of the 169 pairs, 60 resulted in a favourable outcome for the treatment arm based on the first-ranked timepoint in the prioritised GPC, while 8 resulted in a favourable outcome for the control arm and 101 pairs resulted in a neutral comparison, i.e., a favourable outcome could not be determined based on this outcome. The 101 neutral pairs are then compared on the next ranked timepoint, which assigns an additional 54 pairs to either the experimental or control arm. Per timepoint the net treatment benefit (called delta) is derived as the difference between the favourable and unfavourable pairs divided by the total number of pairs. All elements in the tables are additive and sum to the final analysis in the bottom row. For the prioritized GPC, the net treatment benefit is 0.5325 [0.1109, 0.7917], which is significantly different from zero (p= 0.0151) at the 5% level.



endpoint	favorable	unfavorable	neutral	delta	p.value
binary.2	60	8	101	0.3077	
binary.3	45	9	47	0.213	
binary.1	5	3	39	0.0118	
Total	110	20	39	0.5325 [0.1109, 0.7917]	0.0151

Figure 3: Prioritised GPC applied to a dataset with outcome variable "binary" recorded at time points 2,3, 1 (in prioritisation order).



Generalized pairwise comparisons with 3 endpoints

- treatment groups: V (treatment) vs. P (control)
- all pairs are compared for all endpoints (full GPC)
- p-value computed using the permutation distribution (10.000 permutations)
- confidence interval based on asymptotic distribution

endpoint	favorable	unfavorable	neutral	delta	p.value
binary.1	36	10	123	0.1538	
binary.2	60	8	101	0.3077	
binary.3	90	12	67	0.4615	
Total	186	30	291	0.3077 [0.0803, 0.5046]	0.0177

Figure 4: Non-Prioritised GPC applied to a dataset with outcome variable "binary" recorded at 3 different time points.

References

[1] J. Verbeeck *et al.*, 'Generalized pairwise comparison methods to analyze (non)prioritized composite endpoints', *Stat. Med.*, vol. 38, no. 30, pp. 5641–5656, Dec. 2019, doi: 10.1002/sim.8388.

[2] K. Noguchi, Y. R. Gel, E. Brunner, and F. Konietschke, 'nparLD : An *R* Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments', *J. Stat. Softw.*, vol. 50, no. 12, 2012, doi: 10.18637/jss.v050.i12.

[3] M. Geroldinger *et al.*, 'A neutral comparison of statistical methods for analyzing longitudinally measured ordinal outcomes in rare diseases', *Biom. J.*, vol. 66, no. 1, p. 2200236, Jan. 2024, doi: 10.1002/bimj.202200236.

[4] K. Noguchi, M. Latif, K. Thangavelu, F. Konietschke, Y. R. Gel, and E. Brunner, 'nparLD: Nonparametric Analysis of Longitudinal Data in Factorial Experiments'. Aug. 07, 2022. Accessed: Apr. 13, 2024. [Online]. Available: https://cran.r-project.org/web/packages/nparLD/index.html

[5] B. Ozenne, J. Péron, and E. Cantagallo, 'Package "BuyseTest"'. Mar. 2022. Accessed: Apr. 13, 2024. [Online]. Available: https://cran.r-project.org/src/contrib/Archive/BuyseTest/BuyseTest_2.3.11.tar.gz

[6] B. Ozenne, E. Budtz-Jørgensen, and J. Péron. 'The asymptotic distribution of the net benefit estimator in presence of right-censoring'. *SMMR*, vol. 30, no. 11, pp. 2399–2412, 2021, doi: 10.1177/09622802211037067.

[7] J. Verbeeck, B. Ozenne, W. Anderson. 'Evaluation of inferential methods for the net benefit and win ratio statistics'. *J Biopharm Stat*, vol. 30, no 5, pp.765-782, 2020.

[8] W. Anderson, J. Verbeeck. 'Exact permutation and bootstrap distribution of generalized pairwise comparisons statistics'. *Mathematics*, vol. 11, p. 502. 2023.

[9] J. Verbeeck, et al. 'How to analyze continuous and discrete repeated measures in small sample cross-over trials?' *Biometrics*, vol. 79, pp. 3998-4011, 2023.