Project Number 289706
Start date of the project: 01/12/2011, duration: 54 months

Deliverable 9.5

Environmental Risk Assessment of Genetically Modified Organisms: Software for power analysis and analysis of data from field studies

Authors:

Johannes W. Kruisselbrink, Paul W. Goedhart, Hilko van der Voet

Organisation name of lead contractor for this deliverable:

DLO: Wageningen University and Research centre,
Plant Research International, Biometris, Wageningen, Netherlands
http://www.biometris.nl

April 2016

Dissemination Level: Public

Contents

This documents contain the software user manual. The software will be provided to the Commission as an annex to the final report

Summary

This Deliverable describes the second version of the AMIGA Power Analysis software tool. This tool can be used to calculate the power of difference and equivalence tests for the comparison of test and comparator varieties in proposed field trials. The tool will also generate data templates and R scripts that can be used to analyse the data obtained from such trials.

Version 1 of the tool was developed for single-environment field trials. In version 2 the software has been adapted to consider also multi-environment trials. A conclusion of the research has been that taxonomical endpoints with sufficient abundance may be very different between environments. Therefore it is advised to define a relevant set of endpoints for each environment separately. Consequently, also the statistical analysis of data will be performed for each environment. To address multi-environment risk-assessment it has been proposed to standardize observed changes against limits of concern. These so-called concern quotients (*CQ*s) can then be combined over multiple endpoints in the same environment, but also over multiple environments. This approach is available in the software, for two methods of combination, i.e. using maximum or mean levels of *CQ*. The power analysis can then be performed to check if field trials with a proposed size lead to *CQ*s higher than 1 (changes larger than the Limit of Concern) within and over environments. A full description of methods can be found in Deliverable 9.4.

Version 2 of the software (Deliverable 9.5) contains all functionality of version 1 (Deliverable 9.3).

Amiga Power Analysis Tool - User Manual

19 April 2016 – Amiga Power Analysis Version 2.0

1 Introduction

An important task in the field of environmental risk assessment (ERA) is to test whether new varieties have a similar effect on the environment as appropriate, conventional counterparts (EFSA 2010). To address this issue, field trials are designed to compare new varieties with their conventional counterparts (comparators) with respect to the effect on abundance of non-target organisms (NTOs). Using statistical testing, for each NTO measurement unit (or endpoint) it can be determined whether both varieties have a similar effect on the abundance. With the Amiga Power Analysis tool, you can calculate the necessary replication for assessing differences and equivalences between a test and a comparator plant variety under different data models for count and continuous data.

This tool builds on EFSA recommendations (Perry et al. 2009, EFSA 2010) and work in the AMIGA project (Goedhart et al. 2013, 2014, van der Voet et al. 2015) on the amount of replication needed in field trials for GMO safety assessment. It allows to specify the experimental design, additional factors in the experiment, and the method of statistical analysis that will be used. The power of difference tests and equivalence tests (Schuirmann et al. 1987, Perry et al. 2009) is calculated. Difference tests are classical tests where the null hypothesis states equality of mean values. For equivalence tests Limits of Concern (LoCs) have to be specified. The null hypothesis of the equivalence test is that the ratio of test and comparator means is at or outside the LoC(s), against the alternative hypothesis that the ratio is within the LoC boundaries.

This program was developed in the EU project AMIGA (Assessing and monitoring the impacts of genetically modified plants on agro-ecosystems, Arpaia et al. 2014, http://www.amigaproject.eu/).

The software was developed by the Biometris department of Wageningen University and Research centre (http://www.biometris.nl/). Software developers: Johannes Kruisselbrink, Paul Goedhart, Hilko van der Voet.

2 Installation instructions

2.1 Prerequisites

The software is developed for Windows 7 and requires .NET 4.5 client framework. It has not been tested on earlier or later releases of MS Windows.

This software requires the installation of the statistical software R, version 3.0.0 or higher. If not already installed, it is best to install R before the installation of this software.

Follow the steps below to install R:

Step 1: Go to the R website for downloading the Windows version on http://cran.rstudio.org.

Step 2: Click on the link "Download *R.x.x.x* for Windows". This starts downloading R.*x.x.x*-win.exe file for both 32 and 64 bit.

Step 3: After downloading, double click this file to install R. **Important:** Make sure that you keep the default setting under Additional Tasks: "Save version number in registry" checked.

Step 4: Start R and install the packages lsmeans, MASS, reshape, which are required by the software. This can be done by typing:

```
install.packages("lsmeans")
install.packages("MASS")
install.packages("reshape")
```

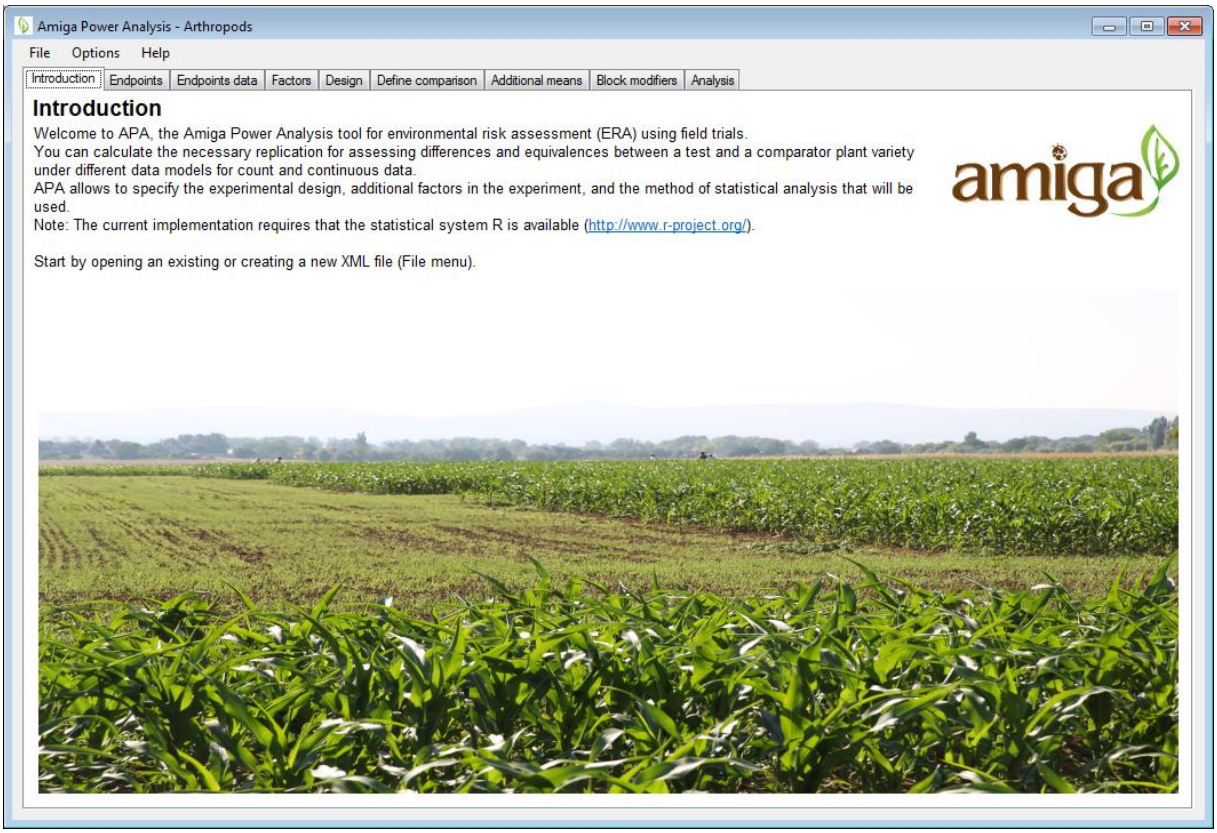
2.2 Installation Steps

Step 1: Double click the appropriate installation file depending on whether your operating system is 32 or 64 bit. (AmigaPowerAnalysis.Installer.Win32.msi or AmigaPowerAnalysis.Installer.Win64.msi). This will run a standard installation. Follow the instructions on the screen – the suggested default settings should apply in most situations.

Step 2: Start Amiga Power Analysis using the desktop shortcut, from the start menu, or from the installation directory.

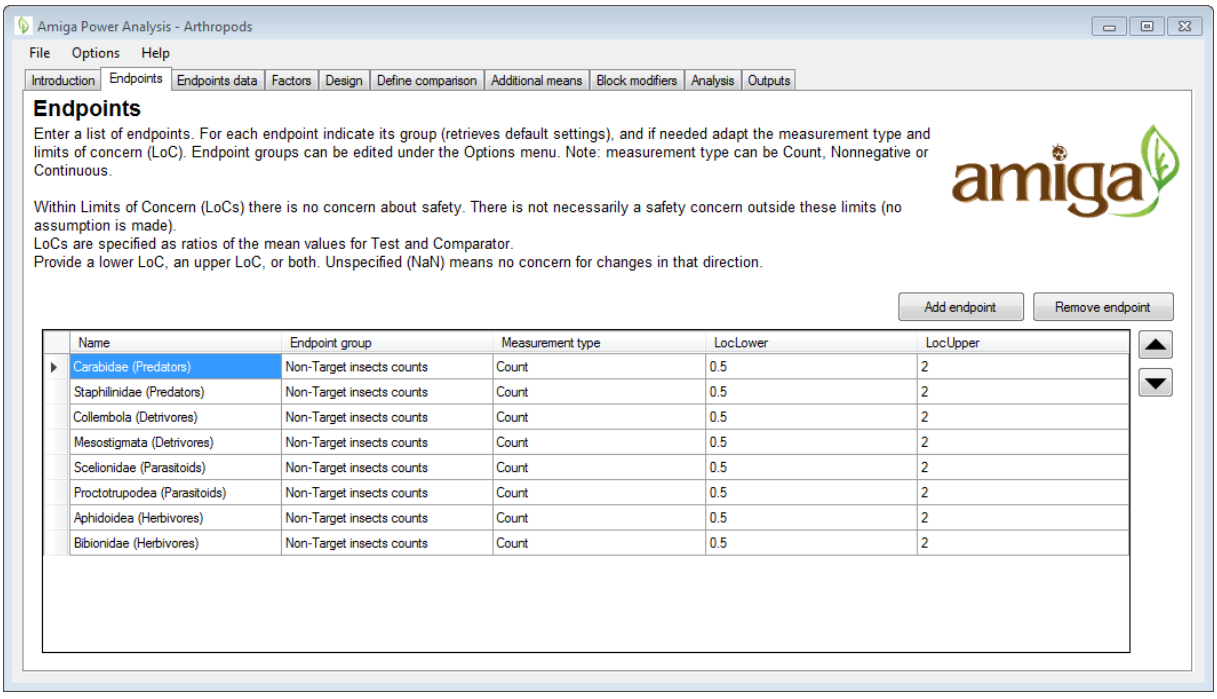
3 Getting Started

Start by opening an existing file or creating a new file. The user interface of Amiga Power Analysis is divided into tabs. In the sections below, the functionality of each tab will be explained separately.



3.1 Endpoints

In the *endpoints* tab, the endpoints that are of interest in the field trial are to be specified. For each endpoint indicate its group (retrieves default settings) and if needed adapt the measurement type and limits of concern (LoC). Endpoint groups can be edited under the Options menu.



Endpoints can be of different measurement types:

- **Count data:** e.g., the number of organisms found on each experimental unit.
- **Non-negative data:** all measurement values are zero or positive (occurs when the measuring time trend curves).
- **Continuous data:** there is no limit on the measurement values.

An essential part of ERA is that for each endpoint, it should be decided beforehand which levels of change between the test and the comparator are still acceptable, and at what level, a change becomes too high to be ignored. In this software, these limits are defined in terms of the limit of concern (LoC) (EFSA 2010). For counts and non-negative data, Limits of Concern are expressed as ratios of the expected values for the test variety (μ_T) and the comparator variety (μ_C), i.e.,

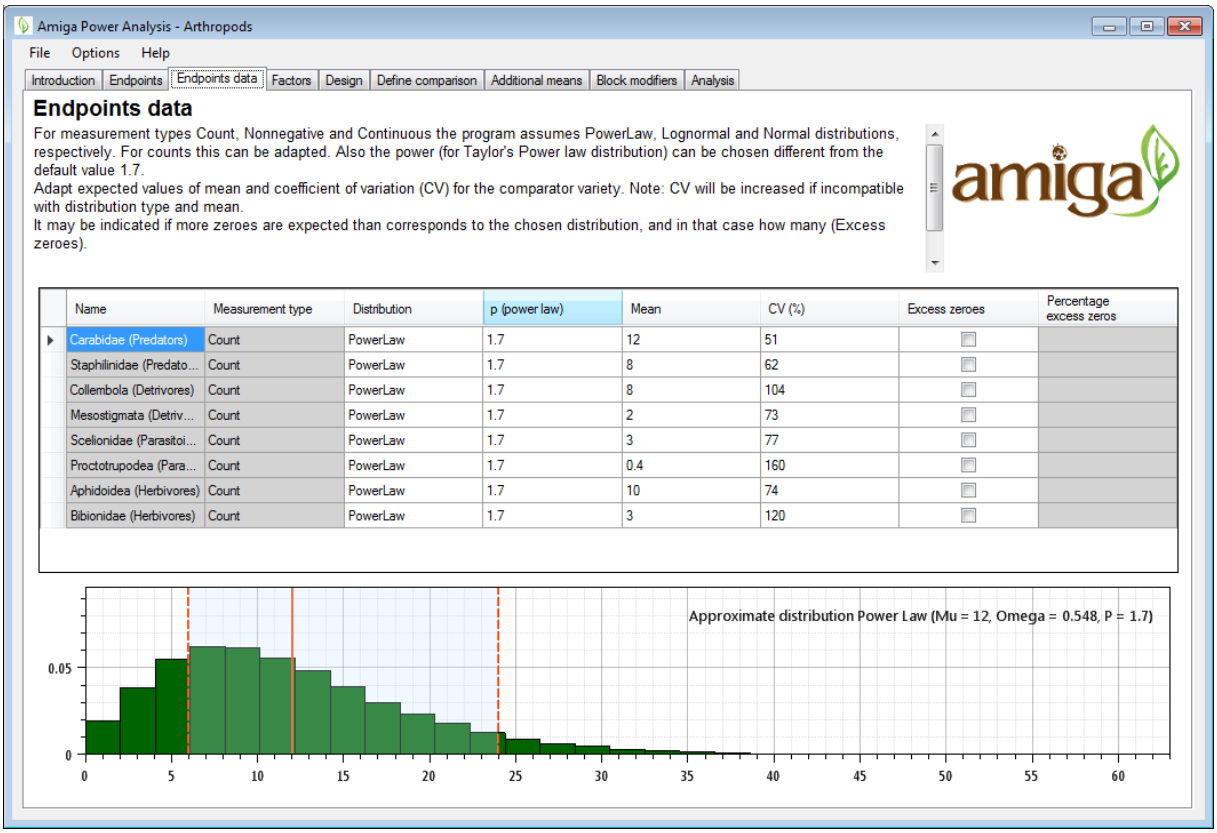
$$\text{LoC} = \mu_T / \mu_C \ .$$

Given this measure, a twofold (or -50%) decrease in abundance is, for example, represented by LoC = 0.5, a twofold (or +100%) increase in abundance is represented by LoC = 2, and LoC = 1 refers to equality. Within these limits there is no concern about safety. Provide a lower LoC, an upper LoC, or both. Unspecified (NaN) means no concern for changes in that direction. For continuous data, Limits of Concern are specified as differences instead of ratios.

Measurement types	**Constraint loc lower**	**Constraint loc upper**	**No difference**	**Remarks**
Counts	> 0	NA	LoC = 1	Suitable when the endpoint data is described in terms of the number of organisms found on each experimental unit. LoC refers to the ratio R of the test mean and the comparator mean, i.e., R = μ_T / μ_C .
Nonnegative	> 0	NA	LoC = 1	For parameters of time trend curves. LoC refers to a difference between the parameters, i.e., D = $\vartheta_T - \vartheta_C$.
Continuous	NA	NA	LoC = 0	

3.2 Endpoints data

The software requires a specification (i.e., a prior estimate) of the data model/distribution of the comparator. This can be specified in the *endpoints data* tab. The data models/distributions of the endpoints can be edited in the table and the graph shows the distribution of the selected endpoint (the red lines indicate the mean and the LoCs). Excess zeroes are not shown.

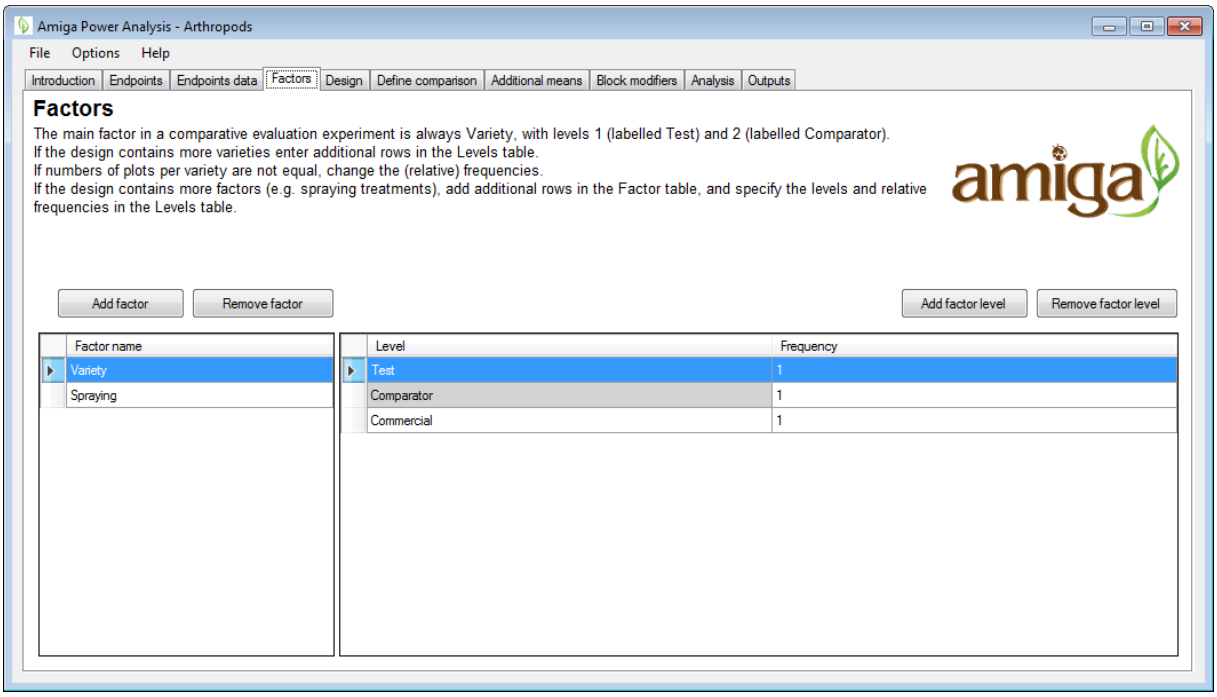


In the software, the specification of the data model is by means of a distribution type, a mean, a CV, and in case of the power model, an additional distribution specific parameter *p*. Additionally, if more zeroes are expected than corresponds to the chosen distribution, the percentage of excess zeroes can be specified using the excess zeroes option. Note that for different measurement types, different distribution types are appropriate. The table below shows the distribution models that are available per measurement type.

Measurement type	**Model**	**Distribution parameters**	**Restrictions**	**Recommended**
Counts	Poisson	$\lambda = \mu$	$\mu > 0$	
	Overdispersed Poisson	$\lambda = \mu$ $\omega = cv^2 \cdot \mu$	$\mu > 0$ $cv > \sqrt{(1/\mu)}$	*
	Negative Binomial	$\omega = cv^2 - 1/\mu$ shape = $1 / \omega$ scale = $\omega \cdot \mu$	$\mu > 0$ $cv > \sqrt{(1/\mu)}$	
	Poisson-Lognormal	$\mu = \mu$ $\omega = cv^2 - 1/\mu$	$\mu > 0$ $cv > \sqrt{(1/\mu)}$	
	Power model	$\mu = \mu$ $\omega = cv^2 - \mu^{2-p}$	$\mu > 0$ $cv > 1 / \sqrt{\mu}$	
Nonnegative	Log-normal	$\mu = \mu$ $\sigma = \lvert\mu \cdot cv\rvert$	$\mu > 0$	*
Continuous	Normal	$\mu = \mu$ $\sigma = \lvert\mu \cdot cv\rvert$		*

3.3 Factors

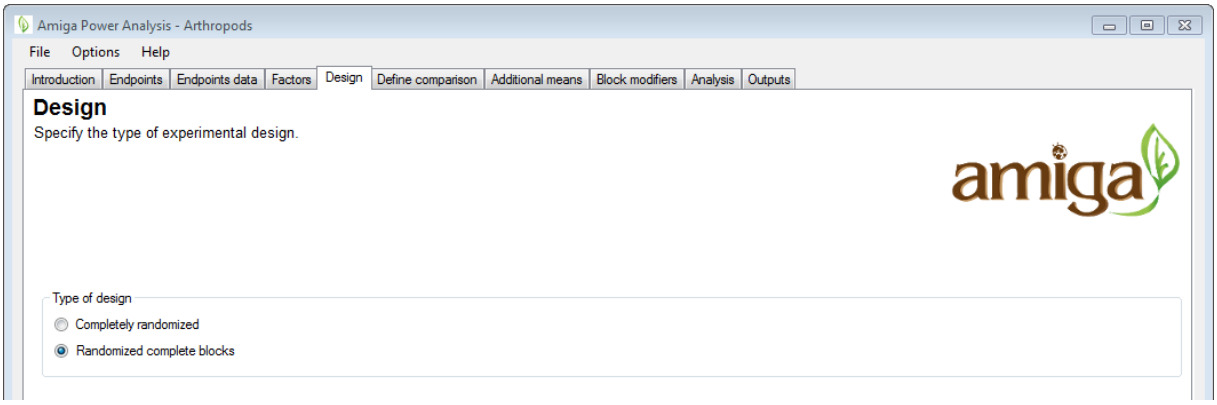
In the *factors* tab, additional varieties and factors of the design can be specified. The main factor in variety-comparative evaluation experiments is always variety, with at least the levels test variety and comparator. However, it may be that the design contains more varieties. These can be expressed as additional variety levels. Also, it may be that the design contains more factors (e.g. spraying treatments). These can be specified by adding additional rows in the factors table and specifying the levels and relative frequencies in the levels table.



Note that unequal numbers of plots per variety or for specific other factor levels can be specified by using (relative) frequencies. If numbers of plots per variety are not equal, change the (relative) frequencies.

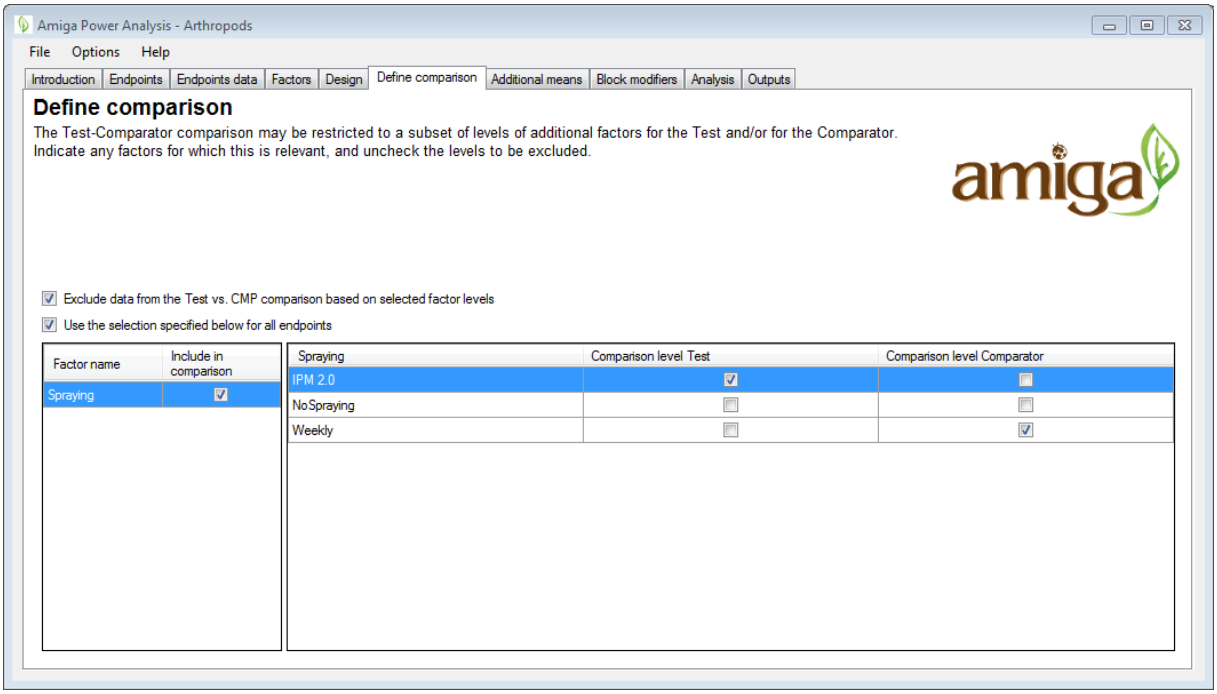
3.4 Design

The *design* tab allows you to specify the type of experimental design. At present, two design types are supported: completely randomized and randomized complete blocks.



3.5 Define comparisons

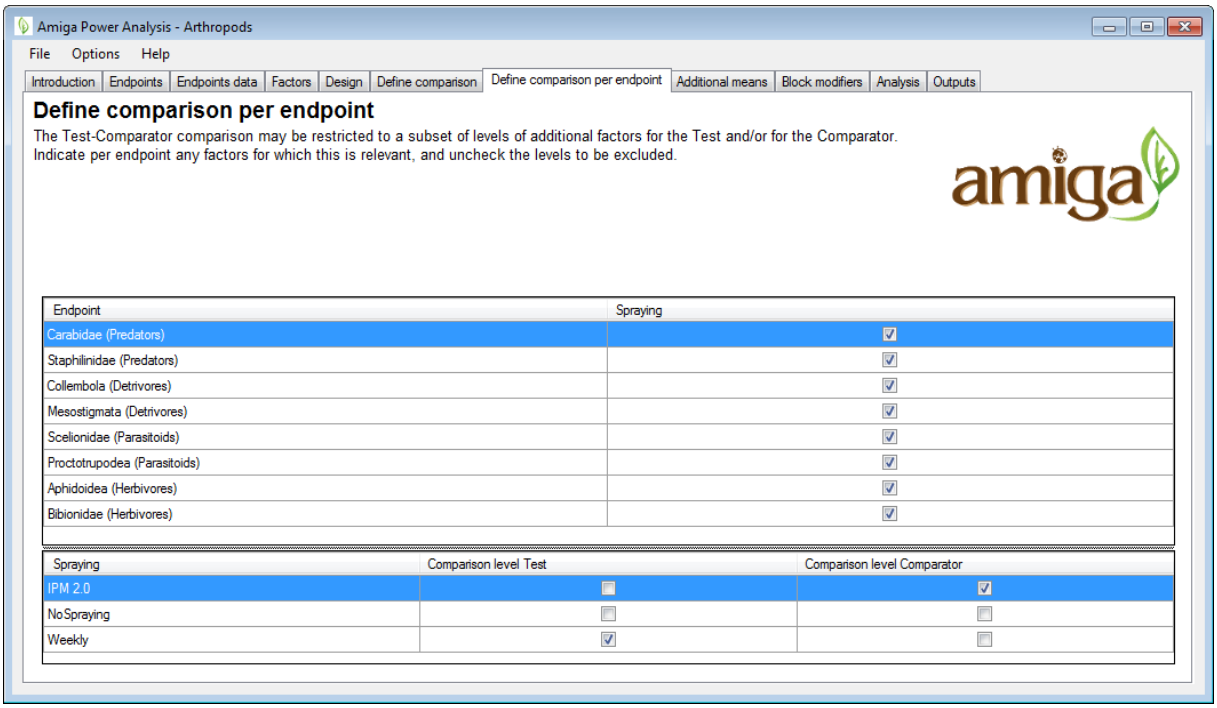
When other factors have been specified, the comparisons between test variety and the comparator can be expected to be either the same for all levels of such a factor (no interaction) or different (interaction). If there are interactions it is necessary to specify which levels of other factors should be looked at when defining the test versus comparator comparison. The *define comparison* tab allows you to specify such interactions. If such interactions are expected, then check the checkbox "*Exclude data from the Test vs. CMP comparison based on selected factor levels*", select the factors for which this is the case, and select the levels that should be included in the test versus comparator comparison. If the comparisons are different for all/some endpoints, uncheck the checkbox "*Use interactions for all endpoints*" will allow you to specify specific endpoints in the next screen.



In this example the comparison of interest is between the Test variety with IPM2.0 spaying and the Comparator variety with weekly spraying . Note that interactions with variety will lower the effective replication, because data from only a subset of levels of the other factor are used in the comparison.

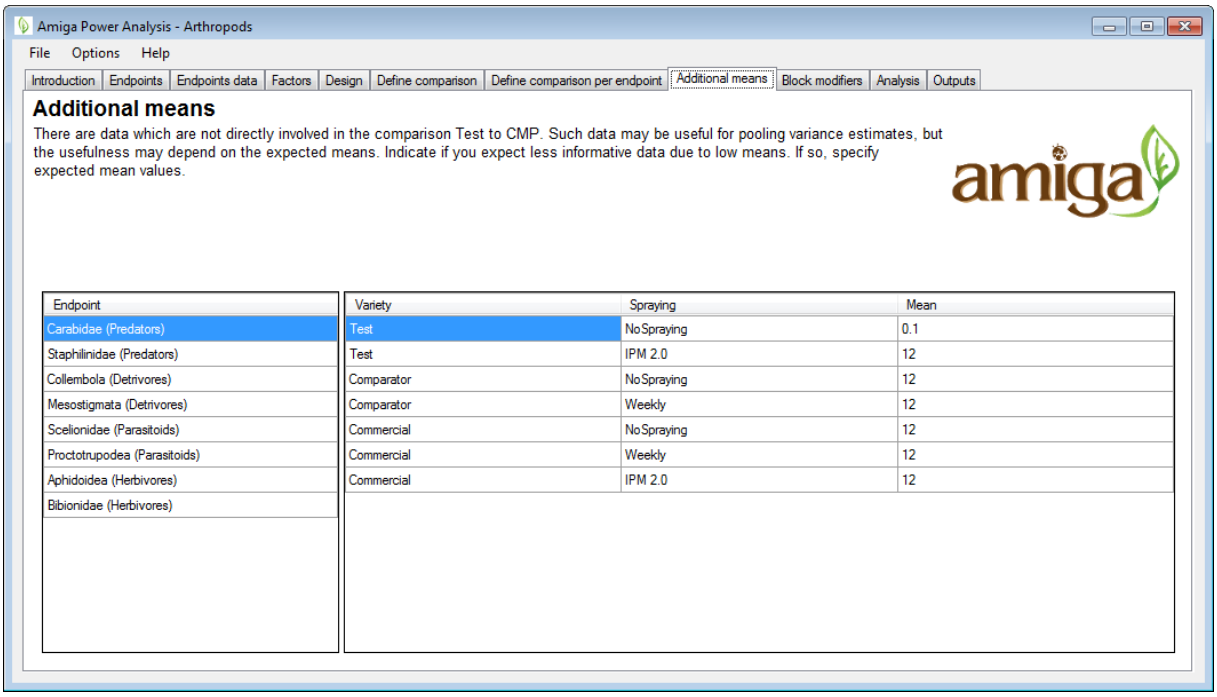
3.6 Define comparisons per endpoint

This tab allows you to specify/modify the comparisons per endpoint. This tab is available only when the checkbox "*Use interactions for all endpoints*" is unchecked in the *define comparison* tab. The top-table allows you to select the endpoint, and to specify for which of the factors, an interaction with variety is expected. The bottom-table allows you to include or exclude specific factor levels.



3.7 Additional means

If factor levels were excluded from the comparison in the *define comparison* tab, then there are data which are not directly involved in the comparison test to comparator. However, such data may still be useful for pooling variance estimates. The usefulness depends on the expected means. In the *additional means* tab, differing means can be specified for factor levels that were excluded from analysis.



Note that the power of tests will be lower if data are uninformative or less informative, e.g., if counts are very low (<5). In principle, the already specified comparator means and CVs are

sufficient to perform the power analysis. However, it should be specified if other factors in the design are expected to make part of the data less informative.

For fixed factors, provide multiplication factors for factor levels where data may become less informative (e.g., counts less than 5).

A restriction for the modifiers is that the joint effect of the modifiers should be neutral:

$$\frac{\sum_{i=1}^{n} \mu_i \cdot w_i}{\sum_{i=1}^{n} w_i} = \mu.$$

Here, μ_i denotes the modified mean for level i and w_i denotes the frequency of this level.

For counts and non-negative measurement types, the modifier effect for level i with modifier Δ_i is

$$\mu_i = \Delta_i \cdot \mu \, .$$

Following the restriction that the joint effect should be neutral, the modifier Δ_i for level i is computed from the other levels as

$$\Delta_i = \frac{\sum_{j=1}^{n} w_j - \sum_{j=1, j \neq i}^{n} \Delta_j \cdot w_j}{w_i} \, .$$

A lower bound for the modifier is $\Delta_i \geq \Delta_l > 0.1$ and from this follows an upper bound the following upper bound

$$\Delta_i \leq \frac{\sum_{j=1}^{n} w_j - \Delta_l \sum_{j=1, j \neq i}^{n} w_j}{w_i} \, .$$

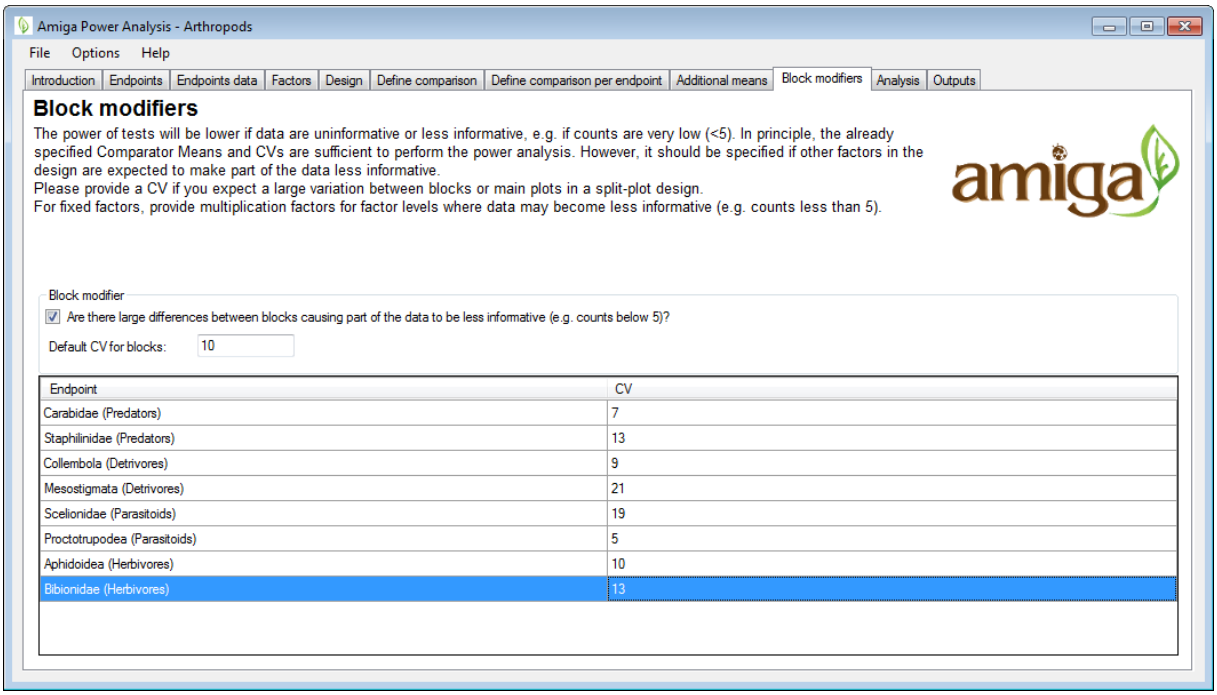
For continuous measurement types, the modifier effect for level i with modifier Δ_i is, in theory, defined as

$$\mu_M = \Delta + \mu \, .$$

However, for this measurement type, the modifier will have no effect on the power analysis.

3.8 Block modifiers

For randomized complete block designs, it may be that there large differences between blocks, causing part of the data to be less informative. If this is the case, then use the *block modifiers* tab to specify the variation between blocks in terms of a CV (%).



Note that within the software, block effects are modelled according to the description of Goedhart et al. (2014).

3.9 Analysis

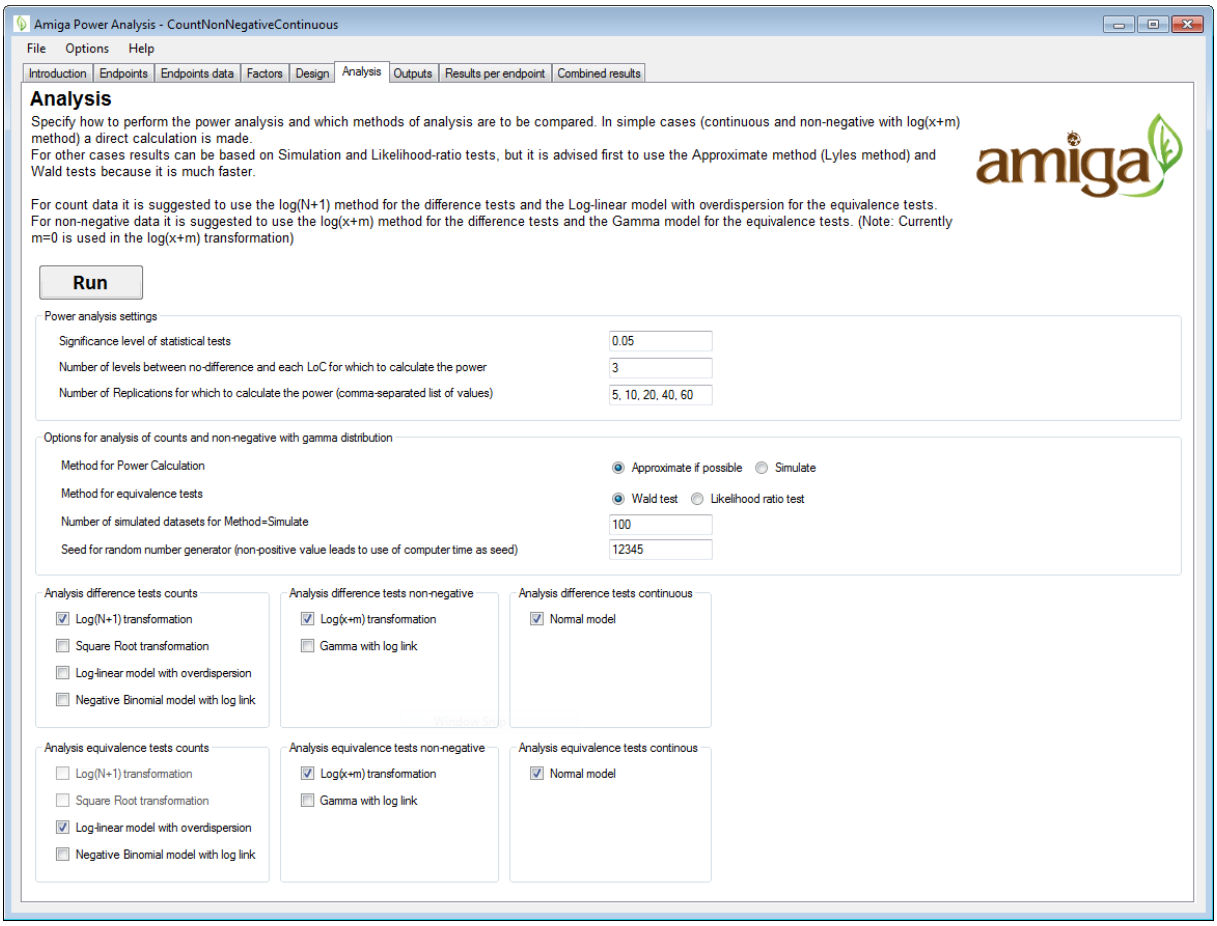
In the *analysis* tab, analysis- and power analysis-specific settings can be specified.

The power analysis settings comprise choosing the significance level, the replication levels, and the number of levels between no-difference and each LoC for which to compute the power.

In simple cases (continuous and non-negative with log(x+m) method) a direct power calculation is made. For counts and non-negative measurement types with a gamma distribution, exact power calculation is not possible. For these endpoints, results can be obtained by means of Monte-Carlo simulation or in some cases it is possible to use the approximate method of Lyles et al. (2007). The latter is recommended, because it is much faster. When the option *approximate if possible* is selected, the method of Lyles will be used when possible.

Two types of statistical tests are considered; the difference test (H0: $\mu_1 = \mu_2$ against HA: $\mu_1 \neq \mu_2$) and the equivalence test (H0: $\mu_1 \neq \mu_2$ against HA: $\mu_1 = \mu_2$, see Schuirmann et al. 1987, Perry et al. 2009). For each test type, the method(s) of analysis method is/are to be specified. These may differ per test type. Different methods of analysis are available/suitable for different measurement types.

When the settings are specified as desired, the pressing the *Run* button will start the power analysis for all endpoints. The analysis may take a while, depending on the number of endpoints, the design, and the specified settings. A progress bar will provide an indication of the progress and the time remaining.



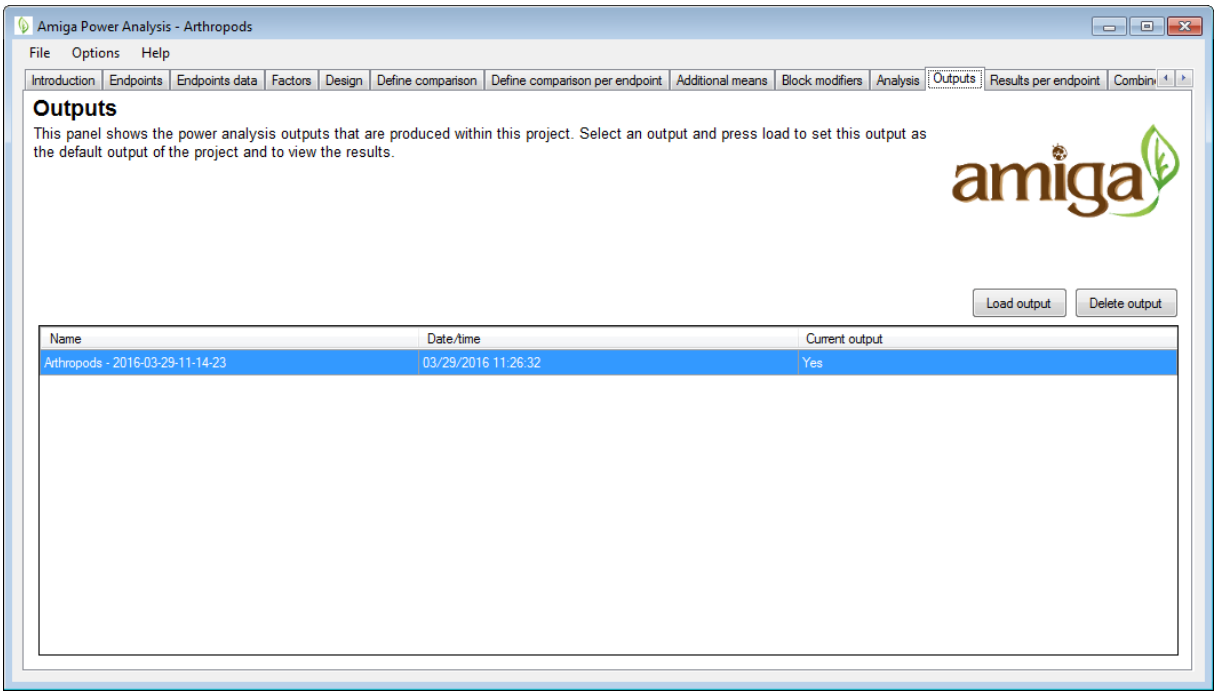
The following methods of analysis are available for the different measurement types:

Measurement type	Model	Recommended[1] for difference test	Recommended[1] for equivalence test
Counts	Log(N+1) transformation		*
	Square Root transformation		
	Log-linear model with overdispersion	*	
	Negative Binomial model with log link		
Nonnegative	Log-normal	*	
	Gamma with log link		*
Continuous	Normal model	*	*

3.10 Output

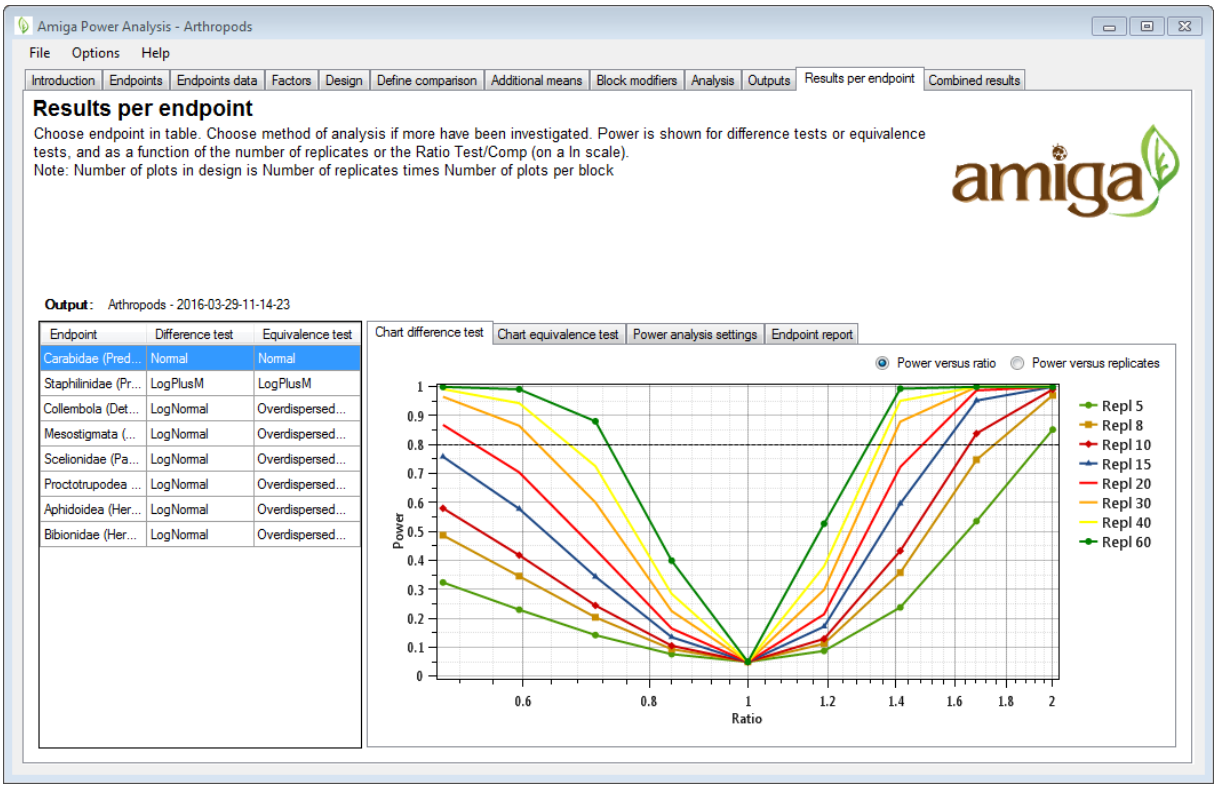
This panel shows the power analysis outputs that are produced. Select an output and press load to set this output as the default output of the project and to view the results.

[1] Recommendations according to AMIGA protocol, see van der Voet & Goedhart (2014).



3.11 Results per comparison

In the *results per comparison tab*, the results of the power analysis are shown per endpoint. Choose endpoint in table. Choose method of analysis if more have been investigated. The tab-panel on the right allows you to switch between the charts for the difference test, charts for the equivalence test, a report on the power analysis settings, and a full analysis report for the selected endpoint.

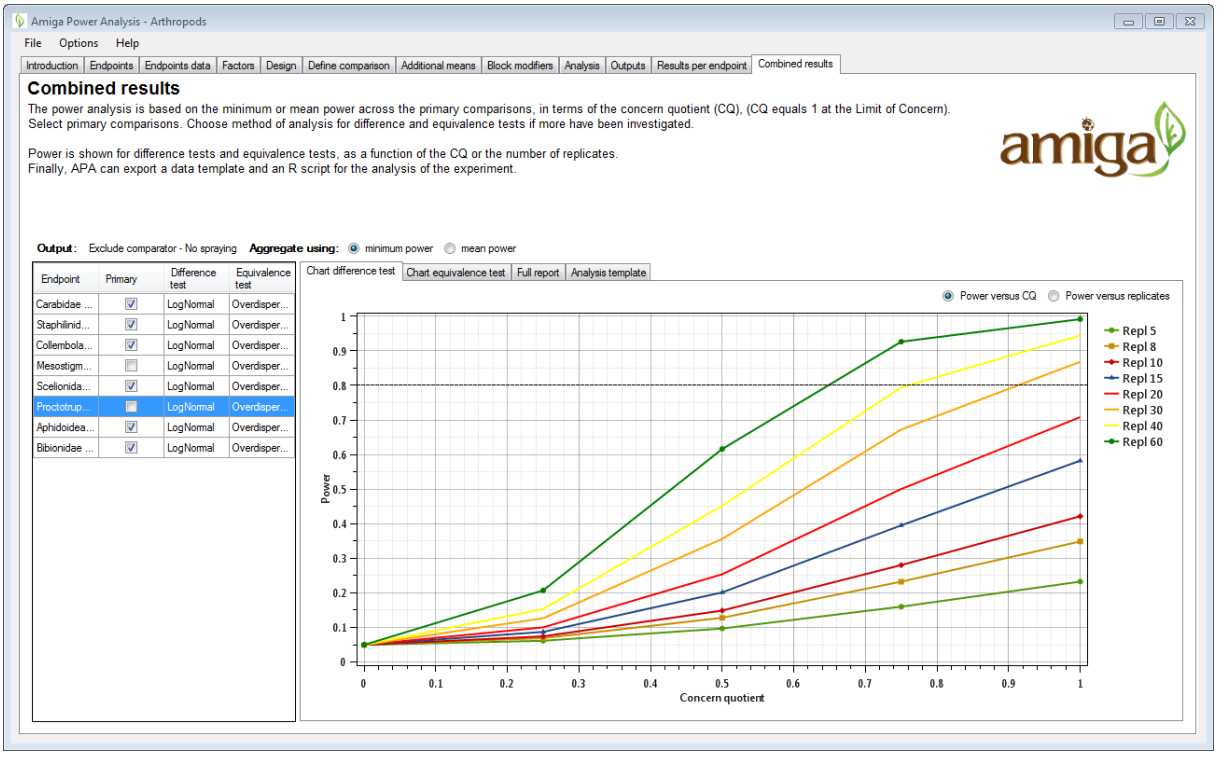


3.12 Combined results

The *combined results* tab provides an combined view of the results of the power analysis for all endpoints. In the left panel, endpoints may be included or excluded for being part of the combined analysis by checking/unchecking the *primary* checkbox. The tab panel on the right provides the combined graph of the difference test and equivalence, as well as a full analysis report for all primary endpoints. The combined power analysis is based on the minimum or mean power across the primary comparisons.

The results per endpoint can be combined by standardizing differences by scaling to a no-concern yardstick representing the minimum limit of potential biological relevance, i.e. the Limit of Concern (LoC). This yields the Concern Quotient (CQ, which equals 0 in case of no difference, and 1 at the Limit of Concern).

Additionally, it is possible to export an analysis template for a specified number of replicates based on the specified design. This will export a data template that can be used for specifying the actual observations, an additional csv file that specifies the comparison contrasts (used by the analysis scripts), and one main analysis R script file and some additional R script that can be used for running the analysis.



References

Arpaia S, Messéan A, Birch NA, Hokannen H, Härtel S, van Loon J, Lovei G, Park J, Spreafico H, Squire GR, Steffan-Dewenter I, Tebbe C, van der Voet H (2014). Assessing and monitoring impacts of genetically modified plants on agro-ecosystems: the approach of AMIGA. Entomologia, 2: 154, 79-86.

EFSA (2010). EFSA Panel on Genetically Modified Organisms (GMO). Guidance on the environmental risk assessment of genetically modified plants. EFSA Journal, 8(11): 1879. [111 pp.], doi:10.2903/j.efsa.2010.1879.

Goedhart PW, van der Voet H, Baldacchino F & Arpaia S (2013). Environmental Risk Assessment of Genetically Modified Organisms: Overview of field studies, examples of datasets, statistical models and a simulation tool. Deliverable 9.1, AMIGA project, http://www.amigaproject.eu/?attachment_id=1801.

Goedhart PW, van der Voet H, Baldacchino F & Arpaia S (2014). A statistical simulation model for field testing of non-target organisms in environmental risk assessment of genetically modified plants. Ecology and Evolution. 4: 1267–1283.

Lyles RH, Lin H-M & Williamson JM (2007). A practical approach to computing power for generalized linear models with nominal, count, or ordinal responses. Statistics In Medicine, 26(7): 1632-1648.

Perry JN, ter Braak CJF, Dixon PM, Duan JJ, Hails RS, Huesken A, Lavielle M, Marvier M, Scardi M, Schmidt K, Tothmeresz B, Schaarschmidt F & van der Voet H (2009). Statistical aspects of environmental risk assessment of GM plants for effects on non-target organisms. Environmental Biosafety Research, 8: 65-78.

Schuirmann DJ (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics, 15(6): 657-680.

van der Voet H & Goedhart PW (2015). Environmental Risk Assessment of Genetically Modified Organisms: Statistical aspects of a protocol for single environment GMO field studies. Deliverable 9.2a, AMIGA project, http://www.amigaproject.eu/?attachment_id=2117.

van der Voet H & Goedhart PW (2015). The power of statistical tests using field trial count data of nontarget organisms in environmental risk assessment of genetically modified plants. Agricultural and Forest Entomology, 17: 164-172.