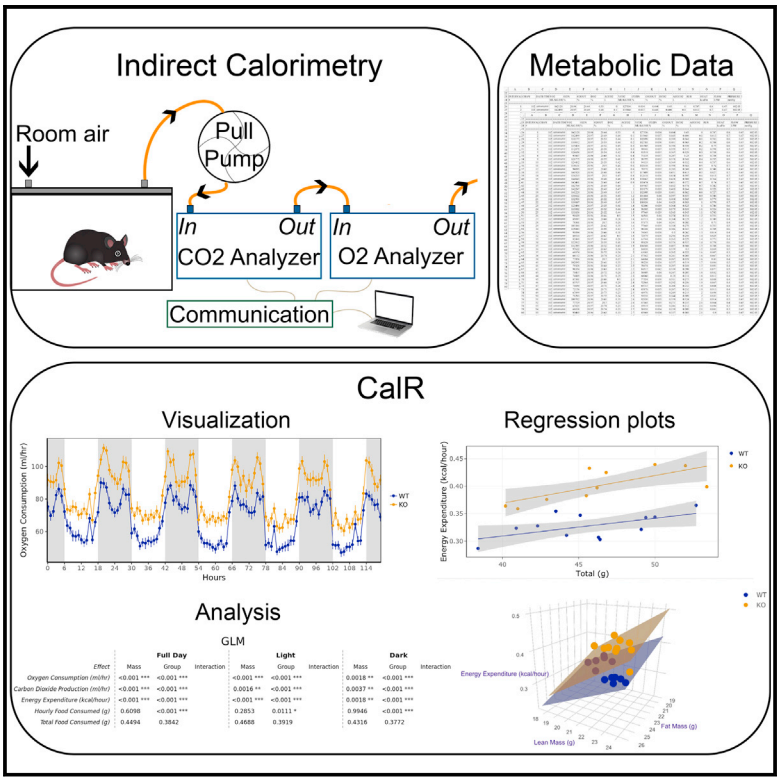
Resource

Cell Metabolism

CalR: A Web-Based Analysis Tool for Indirect Calorimetry Experiments

Graphical Abstract



Authors

Amir I. Mina, Raymond A. LeClair, Katherine B. LeClair, David E. Cohen, Louise Lantier, Alexander S. Banks

Correspondence

abanks@bwh.harvard.edu

In Brief

Indirect calorimetry is a powerful tool for studying energy balance. These experiments produce rich datasets but are difficult to analyze properly and lack transparency. Mina and colleagues created a rigorous tool, *CalR*, that takes raw data from indirect calorimeters, allows comprehensive data exploration and reproducible workflows, and provides standardized analyses.

Highlights

- CalR is a free web tool for analysis of experiments using indirect calorimetry
- Imports data, generates plots, and determines the best-fit statistical model
- Outputs a standardized CalR file that can be shared, deposited, and re-read by CalR
- Increases speed, transparency, and reproducibility of energy balance experiments

Mina et al., 2018, Cell Metabolism *28*, 656–666
October 2, 2018
https://doi.org/10.1016/j.cmet.2018.06.019

CellPress

CellPress

Cell Metabolism
Resource

CalR: A Web-Based Analysis Tool for Indirect Calorimetry Experiments

Amir I. Mina,[1] Raymond A. LeClair,[2] Katherine B. LeClair,[1] David E. Cohen,[3] Louise Lantier,[4] and Alexander S. Banks[1,5,*]
[1]Division of Endocrinology, Diabetes and Hypertension, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA 02115, USA
[2]Springbok LLC, Boston, MA 02109, USA
[3]Division of Gastroenterology & Hepatology, Weill Cornell Medical College, New York, NY 10065, USA
[4]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37212, USA
[5]Lead Contact
*Correspondence: abanks@bwh.harvard.edu
https://doi.org/10.1016/j.cmet.2018.06.019

SUMMARY

We report a web-based tool for analysis of experiments using indirect calorimetry to measure physiological energy balance. *CalR* simplifies the process to import raw data files, generate plots, and determine the most appropriate statistical tests for interpretation. Analysis using the generalized linear model (which includes ANOVA and ANCOVA) allows for flexibility in interpreting diverse experimental designs, including those of obesity and thermogenesis. Users also may produce standardized output files for an experiment that can be shared and subsequently re-evaluated using *CalR*. This framework will provide the transparency necessary to enhance consistency, rigor, and reproducibility. The *CalR* analysis software will greatly increase the speed and efficiency with which metabolic experiments can be organized, analyzed per accepted norms, and reproduced and will likely become a standard tool for the field. *CalR* is accessible at https://CalRapp.org/.

INTRODUCTION

The increased prevalence of obesity, which arises with an imbalance in food intake and energy expenditure (EE), is driving increased morbidity and mortality worldwide (Calle et al., 2003). While much focus has been placed on elevated caloric consumption as the primary force driving the rise in obesity, increasing attention is being directed toward the therapeutic potential of increasing EE (Betz and Enerback, 2018; Guyenet and Schwartz, 2012). In addition, decreased EE following weight loss contributes to the persistence of obesity (Stanford et al., 2013). As such, indirect calorimetry measurements of EE have proven invaluable in furthering our understanding of obesity's pathogenesis (Kaiyala and Schwartz, 2011). Indirect calorimetry is a non-invasive method of EE determination based on gas exchange. Alternatives such as direct bomb calorimetry require sacrificing the animals and harvesting organs, eliminating the possibility for serial measurements. In contrast, indirect calorimetry allows for more flexible and sophisticated experiments that can be repeated in the same animals over time. Although the use of indirect calorimetry has become widespread, controversies have emerged on the appropriate treatment of the data generated by these experiments, fundamentally challenging some published conclusions (Butler and Kozak, 2010). Because analysis of these large datasets is somewhat onerous, there is a need for a tool to assist with appropriate analysis and interpretation of results in a comprehensive and standardized manner. The absence of such tools has led to conflicting analyses of experimental data (Himms-Hagen, 1997; Pelleymounter et al., 1995). A recent effort to promote transparency and increase the rigor of the scientific process, especially regarding biostatistical analysis and improving reproducibility, has created an atmosphere receptive toward novel tools that assist in achieving these goals (Drucker, 2016; Flier, 2017; Jarvis and Williams, 2016).

Following decades of debate, analysis of covariance (ANCOVA) has become the consensus method for the analysis of indirect calorimetry EE data when comparing animals of different body composition (e.g., obesity) (Arch et al., 2006; Kaiyala and Schwartz, 2011; Speakman et al., 2013; Tschöp et al., 2012). ANCOVA statistically detaches (adjusts for) the influence of a continuous variable (e.g., total body mass) from group comparisons of a dependent variable (e.g., EE). ANCOVA was originally developed to extend the precision of ANOVA by adjusting for a continuous variable (called a covariate) that correlates with group means and/or variances, thus increasing the power of the study (Fisher, 1947). ANCOVA is included as a special case of the generalized linear model (GLM), which encompasses a great many classical statistical analysis techniques (e.g., ANOVA, linear regression, and logistic regression) (McCullagh and Nelder, 1998). Because neither lean body mass (LBM) nor fat mass (FM) is metabolically inert, ANCOVA may include LBM or FM as a covariate in the analysis of EE.

In many cases, constraints exist for widespread implementation of ANCOVA in EE analysis. These barriers include "wrangling" large datasets to prepare the raw data for analysis, unfamiliarity with statistical software packages, and the lack of a commercial software package to perform statistical analysis of indirect calorimetry experiments. As a consequence, regression-based analysis, such as ANCOVA, is not consistently being implemented in the analysis of energy balance in mice of

different body composition despite the apparent need for a solution.

A key assumption of ANCOVA in EE studies is that the effect of the body mass covariate on EE is the same for all groups (i.e., parallel slopes in an EE versus mass regression plot). While ANCOVA brings many benefits to the interpretation of energy balance, this assumption can be overly restrictive under non-standard experimental conditions, including those of non-shivering thermogenesis. Activation of brown adipose tissue (BAT) increases EE through heat generation, which is strongly dependent on BAT mass (Stanford et al., 2013). Experiments of thermogenesis in mice with greater BAT mass may violate assumptions of the ANCOVA with a differential interaction of mass and EE (i.e., non-parallel slopes of an EE versus mass plot). A GLM can be used to analyze experiments where a significant interaction effect exists between body mass and EE. In GLM models, the magnitude of the adjusted difference between groups depends on the value of the body mass covariate, whereas in ANCOVA, the adjusted difference between groups is the same across the range of the mass covariate.

Here we describe our *CalR* software project, which is an easy-to-use web-based software tool freely available to the scientific community. *CalR* is developed to enable investigators to thoroughly and reproducibly perform statistical analyses of indirect calorimetry data. *CalR* allows users to import large data files, evaluate the experiment's validity, examine data for outliers from experimental artifacts, and compare statistical differences between groups. The results and workflow are exportable as files that can be shared in a centralized repository or as supplementary data accompanying publications. In this article, we focus on describing the structure, and the rigorous statistical methodology of this software tool and provide its application in analyzing data from examples. *CalR* has the potential to become a standard resource for examination of energy balance experiments in laboratory animals.

RESULTS

Software Architecture

Overview of Design

This software package, designated *CalR* (an abbreviated form of *calor*, the Latin word for heat), primarily functions to include reading and visualizing raw calorimetry data and performing statistical analysis. The user-friendly *CalR* web pages allow the user to specify body mass data and assign subjects into groups. Navigating through the tabs of *CalR*, users will find their data for metabolic variables plotted either as group averages or as individual tracings. Once the time region of interest is selected, *CalR* conducts the appropriate data analyses depending on the selected metabolic parameter and incorporates mass as a covariate. The abundance of input options gives users the flexibility to explore data from a variety of experimental designs.

Framework and Availability

CalR is written in R (R Core Team, 2017) using a Shiny graphical user interface (GUI) to capitalize on robust statistical analysis routines, free availability, and intuitive user interface. The *CalR* R package provides the essential functions for importing, curating, viewing, and analyzing the data. These functions are designed to be flexible with respect to the distinctions between templates. The web application is hosted on a Linux cloud computing server (Ubuntu 16.04) through Amazon Web Services' (AWS) Elastic Compute Cloud (EC2). We generated an Elastic IP address in AWS and, in partnership with Partners Healthcare, hosted the *CalR* web application on a secure, publicly available domain. Each analysis template is set up as a distinct Shiny application individually hosted on the server and routed to a web page within https://CalRapp.org/.

Distribution

The *CalR* graphical front-end of this software operates in a browser window and can be executed by navigating to a web site hosted at https://CalRapp.org/.

System Compatibility

The data generated by any of the three widely used, high-quality manufacturers of indirect calorimeter systems for small animals (Sable Systems, TSE, and Columbus Instruments) can be imported directly using *CalR*'s GUI. The "Input" tab within any *CalR* template contains a section in which a user may import one or more Comma Separated Value (CSV) files; this will depend on the manufacturer's system. Below are specific steps for selecting the preferences to allow data import into *CalR* from each of these systems.

User Workflow

Data Preparation

Columbus Instruments' Comprehensive Lab Animal Monitoring System. A critical issue for the use of data from this system software is the "automatic normalization," which divides metabolic parameters by body weight (e.g., VO_2 mL/kg/hr). When analyzing Comprehensive Lab Animal Monitoring System (CLAMS) data, *CalR* will automatically reverse this normalization (e.g., VO_2 mL/hr) before any further calculations. For this reason, when setting up the experiment within Oxymax by navigating to Experiment > Setup, users may enter any value for the subject mass and maintain the default "Volume Rate Units" setting to "mL/kg/hr" under Experiment > Properties. Here the user should also make sure the "Heat Calculation" setting is "Standard, kcal." After an experiment has been completed and stopped, open the Oxymax program and select "Run Oxymax as Data Viewer." When prompted, choose the hardware configuration file (.INI) used for setting up your experiment. Next, navigate to File > Open experiment data and open the .CDTA file from the CLAMS run. Once opened, navigate to File > Export > Export all subject CSVs. Each cage run by the CLAMS system generates a separate output file, all of which are needed for analysis with *CalR*.

Sable Systems' Promethion. The high-density data collected by Promethion systems necessitates pre-processing steps to reduce file sizes and processing times. The Expedata software system allows for macro functions that will produce standardized output formats. Macro 13 provides users with the metabolic variables of interest at each reading for each cage. *CalR* can import data generated by Macro 13 processing. Files must first be saved as CSV formatted files.

TSE's PhenoMaster/LabMaster. The TSE system produces three outputs for each mass-dependent variable: values normalized to total body weight, e.g., $VO_2(1)$, allometric scaling to approximate normalization to LBM $VO_2(2)$, or uncorrected values $VO_2(3)$. *CalR* uses this latter set, the uncorrected values for

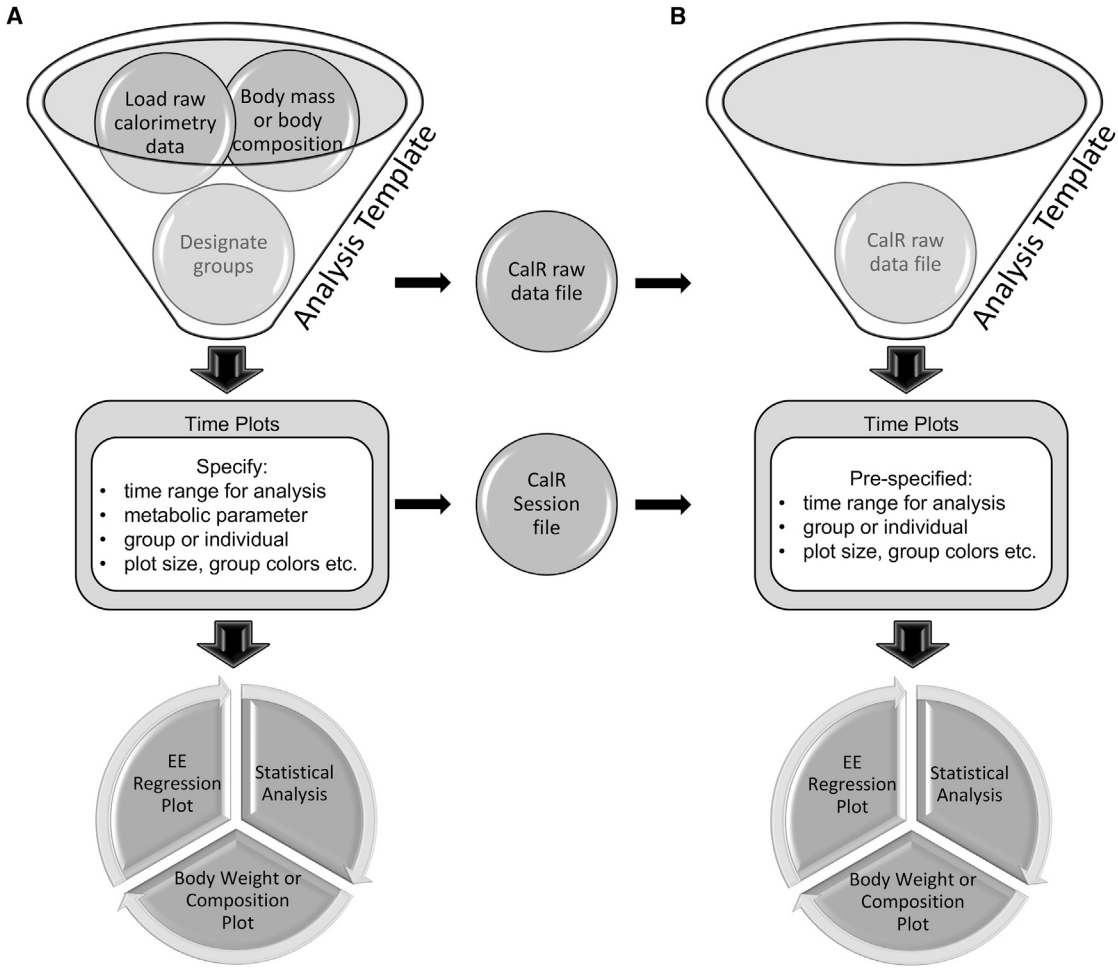


Figure 1. CalR Data Analysis Workflow

Users will select an analysis template that best matches their experimental design.

(A) First-time analysis of an experiment includes loading raw indirect calorimetry data, optionally loading body composition data, and assigning animals into groups. This raw data can be exported into a standardized CalR data file for fast loading in subsequent sessions. Plotting parameters, including the time range for analysis and other aesthetic preferences, are set and visualized. These settings are saved as a CalR Session file. Once parameters are defined, statistical analysis and additional plotting results are available.

(B) Exported CalR raw data files and CalR Session files allow fast, transparent analysis and a record for reproducible research. Simply stated, the CalR data file is an unaltered experimental record. The CalR Session file contains what was done to produce the analysis.

VO_2(3), VCO_2(3), and EE, denoted as H(3). To select the variables suitable for *CalR*, go to the "View" menu and select the following parameters: XT+YT, XA, YA, H(3), VO_2(3), VCO_2(3), respiratory exchange ratio (RER), Drink, Feed, and Weight. Also, make sure that the "Export table" setting is "Format 1." When ready to export, enter the "Export" menu and navigate to Export > Table and set "Save as type" to be ".CSV."

Typical CalR *Workflow*

The information required to plot and analyze an experiment includes the raw calorimetry data file, body weight, or body composition information and to which group each subject belongs. The raw data files are parsed and read into memory. Group names are specified, and the animal identifiers are moved into the corresponding column (Figure 1A; Methods S1). Users can then explore the data under the "Time Plots" tab. Further preferences are subject to users' selections of inputs, including which metabolic variable to examine, plotting by group or individual, the time range of interest, the inclusion of error bars (±SEM), removal of outliers (±3 SDs), and aesthetic features. The raw data are reformatted into a *CalR* raw data file, and all aesthetic features are included in a *CalR* Session file. These files obviate the need for repeated data entry in subsequent sessions (Figure 1B). Once these are specified, the tabs producing analyses, including weight plots, average value plots, regression plots, and analysis results, are populated.

Defining an Experiment

Each indirect calorimetry run may contain more than one experimental intervention. We present an example in which mice are maintained at thermoneutrality (30°C) for 3 days followed by a transition to 4°C (Figure 2). The time corresponding to the experimental period of interest can be selected with a slider bar under the "Time Plots" tab (see Supplemental Information).

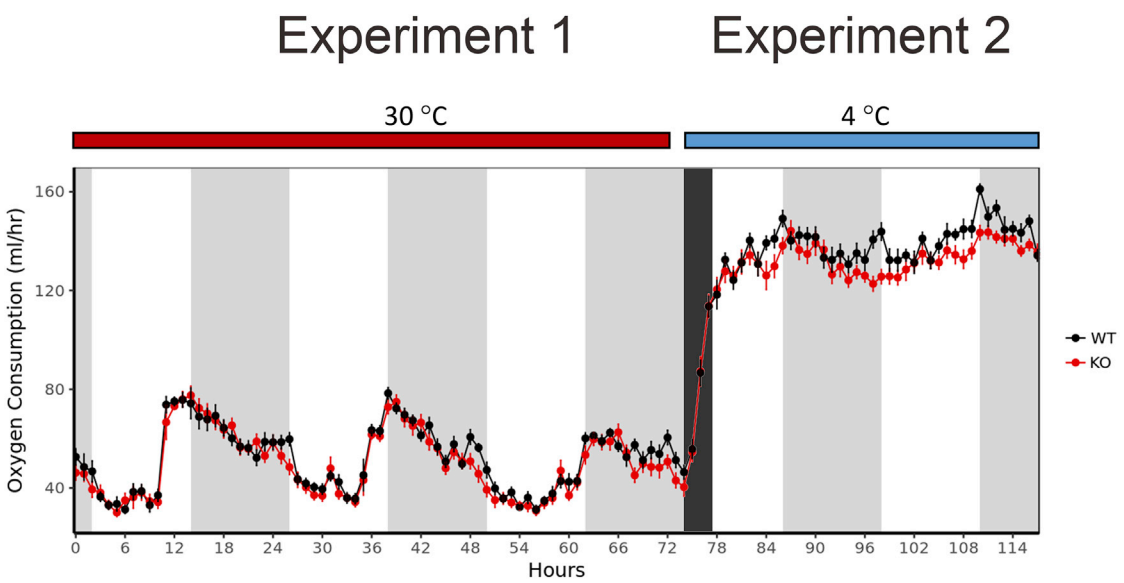


Figure 2. Defining an Experiment

This one calorimetry run included two experiments in which two groups of mice were maintained at thermoneutrality (30°C) (Experiment 1) followed by a transition to a cold challenge and maintenance at 4°C (Experiment 2). Users of CalR would sequentially analyze these two experiments by selecting the corresponding time regions.

A "Notes" section is included to allow the user to describe the criteria used in selecting the interval used for analysis. In providing a generalized framework for analysis, this single-experiment analysis offers the greatest flexibility for a range of experimental designs.

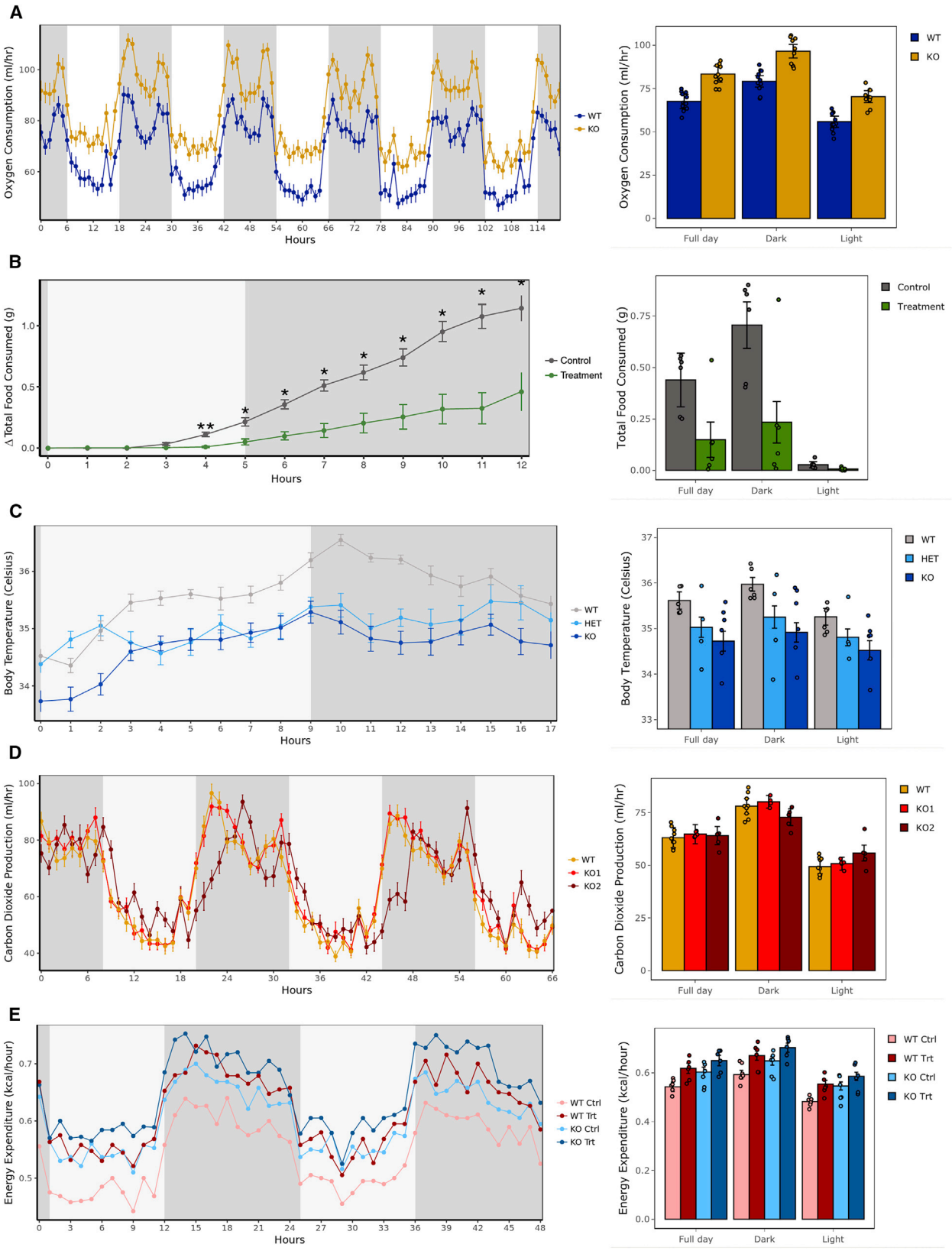
Analysis and Experiment Templates

CalR provides the flexibility to interpret many common experimental designs. Each indirect calorimetry experiment is unique, but we have created standardized templates for many common practices (Figure 3). Users are directed to select a template that will apply the most suitable statistical approach. For this reason, we have prepared templates to analyze the following six commonly encountered experimental designs and one template to combine experimental runs from different times on different subjects. A description of each template, including a sample dataset and step-by-step instructions, are included (Supplemental Information):

1. Two groups (e.g., two distinct genotypes). A common paradigm where two groups are studied simultaneously for more than 12 hr (Methods S1).
2. Two groups with acute treatment (e.g., administration of a β-adrenergic receptor agonist to stimulate metabolism). This template is ideal for targeting analysis over a region of 12 hr or less. It includes time plots of metabolic differences from a designated start hour. Analysis is performed every hour (Methods S1).
3. Three ordered groups (e.g., dose-response or wild-type [WT]/heterozygous/knockout). This template is for observing dose effects, either allelic, pharmacologic, or conditioning. Groups are ordered into a hierarchy for analysis, which includes post hoc tests (Methods S1).
4. Three factored groups (e.g., Vehicle versus two independent treatments or WT versus two independent knockouts). Contrary to the ordered template, this does not assume a hierarchical ordering of the group variable. This makes the analysis, including post hoc tests, distinct from the previous template (Methods S1).
5. Four groups (e.g., two genotypes with two treatments or four independent genotypes). This template allows for any combination of four independent comparators and includes post hoc tests (Methods S1).
6. Crossover experiment (for two groups and one intervention). The intervention regimen is alternating between two groups (e.g., vehicle followed by drug and drug followed by vehicle). Animals are examined for treatment effect, period effect, and interaction effect (treatment × period) (Methods S1).
7. Experimental run combination Tool. *CalR* also provides a template that generates a graphical interface to facilitate combining multiple experimental runs into one *CalR* data file (Methods S1).

Data Visualization

There are several tabs available in the *CalR* web pages for observing the data with distinct perspectives. Navigating the tabs, a user can see body mass and body composition bar plots, time plots, average value plots, and regression plots. The data presented are not normalized by total body mass, LBM, or any allometric scaling factor due to the significant distortions that these can introduce (Himms-Hagen, 1997). The values presented under the "Time Plots" tab are the mean values for each group per hour. In this tab, a caution button labeled "Abnormal Readings" may appear if any of the raw, non-excluded data points in the selected time are physiologically impossible (e.g., RER at 0.4). Clicking this button will display the qualifying criteria and the ID of the corresponding subject(s). Under the "Average Plots" tab, the mean value for each 24-hr,



(legend on next page)

light, or dark photoperiods are presented as well as overall means for the selected region of analysis. The experimental period of interest can be specified with a slider bar to perform analysis on one experiment at a time. Many of the features of the time plots can be customized using the dialogue box to specify colors, sizes, or inclusion of error bars. Weight plots represent the mean body mass or body composition if these data are supplied. Regression plots are often informative for understanding the relationship between EE and mass but require appropriate sample size for useful interpretation.

Transparency, Portability, and Reproducibility

The CalR *Data File*

The data files produced by indirect calorimeters of different manufacturers are formatted differently. Nonetheless, data loaded into *CalR* from each of the three supported manufacturers are first transformed into a standardized format that can be exported as a *CalR* data file. The *CalR* data file contains the unmodified raw data for each animal as initially collected. As a complete standardized record of the experiment, a *CalR* data file can be shared or included as Supplemental Information.

The CalR *Session File*

Outside of the raw calorimetry data, additional information is required to complete the analyses. This includes importing body weights or body compositions, designating groups, specifying which cages are to be included in the analysis, selecting the time of the experiment, and choosing aesthetic preferences. The session file allows for specific and reproducible analysis of either raw data or data from a *CalR* file. Multiple *CalR* Session files should be produced for calorimetry runs with numerous distinct experiments. This modular format will facilitate the sharing of information and the creation of repositories of metabolic datasets. We have included examples of generating and reading the *CalR* file in the vignettes included as Supplemental Information.

The Excluded Data File

As described above, *CalR* contains optional features to automatically remove outliers and manually exclude cages from the analyses. These specified data are not plotted, included in the group averages, or included in statistical analyses. However, these data are retained within the *CalR* data file, which is a complete record of the raw experimental data. Furthermore, manually selected cages and automatically identified outlier data are included together in an ''excluded data'' file accessible through the Time plots tab. When the ''remove outlier'' feature is toggled on, the investigator is strongly encouraged to review the downloadable excluded data file to confirm the validity of these exclusions.

Statistical Approach

Overview

CalR implements GLMs to describe the group effect under investigation while properly accounting for body mass effects on mass-dependent metabolic variables, such as EE. For a selected time range, metabolic variable, and photoperiod, all of the data points that are not excluded are averaged into a single value per animal. This value will be used in the ANCOVA/GLM for each animal using the selected mass variable as the covariate. Alternatively, we tested the use of all measurements as individual data points with the use of a random effect variable to account for within-mouse correlation with no notable differences yielded between this and our current model. For measurements not associated with mass (e.g., RER), the difference between groups is analyzed by a one-way ANOVA (Figure 4A). To model mass-dependent variables, the user specifies the body mass variable and includes it as a covariate. This action is required independently of whether or not the masses are significantly different between groups in order to better fit the data (Allison et al., 1995; Arch et al., 2006; Kaiyala, 2014; Kaiyala et al., 2010; Kaiyala and Schwartz, 2011; Katch, 1972a, 1972b; Kronmal, 1993; Speakman et al., 2013; Tanner, 1949; Tschöp et al., 2012). However, one essential requirement for ANCOVA is that there is not a significant group by mass interaction. This means that it is reasonable to assume that the slope of EE on mass is the same for each group (i.e., the slopes of the groups must be parallel; Figure 4B). How do we interpret an experiment in which the slopes are not parallel, groups have different associations between EE and mass, and the assumptions of ANCOVA have been violated? The GLM can adequately account for this scenario in what might unconventionally be called ''ANOVA with interaction'' or ''ANCOVA for non-parallel slopes'' (Kowalski et al., 1994) but is here referenced only as GLM for lack of more specific statistical nomenclature (Figure 4C). *CalR* will first perform a GLM; if a significant interaction effect is observed, the significance of the group, mass, and interaction effects are reported. If there is no significant interaction effect, *CalR* returns an ANCOVA with group and mass effects only. This automated algorithm will prove applicable to most indirect calorimetry experiments. The two assumptions of this analysis are that (1) a study animal's measurements are independent of those from others and (2) the true distribution of the measurements within the population (e.g., mice of the same strain and genotype as the study animals) is approximately normally distributed.

ANOVA with Interaction

The GLM is performed on metabolic parameters that are anticipated to depend on body mass based on core physiological principles. From the group names listed in the "Input" tab, the first will be the reference group when this categorical variable is coded into the model. The GLM model contains ''group'' as

Figure 3. Example Data from Five Analysis Templates

Left column, Time Plot; Right column, Overall Experiment Summary.

(A) Two-group template. VO2 for two groups of mice monitored for 4 days at room temperature.

(B) Two-group acute response template. Food intake for two groups over 10 hr following treatment.

(C) Three-group template: ordered. Body temperature in WT, heterozygote, and knockout animals maintained at 4°C.

(D) Three-group template: non-ordered. EE of WT and two independent knockout strains.

(E) Four-group template. EE analysis of two genotypes of mice on two different diets. Not shown: crossover template or run combination tool. Note: CalR plots are not normalized or adjusted to body weight, lean mass, or another allometric scaling.

*$p < 0.05$; **$p < 0.01$ by ANCOVA.

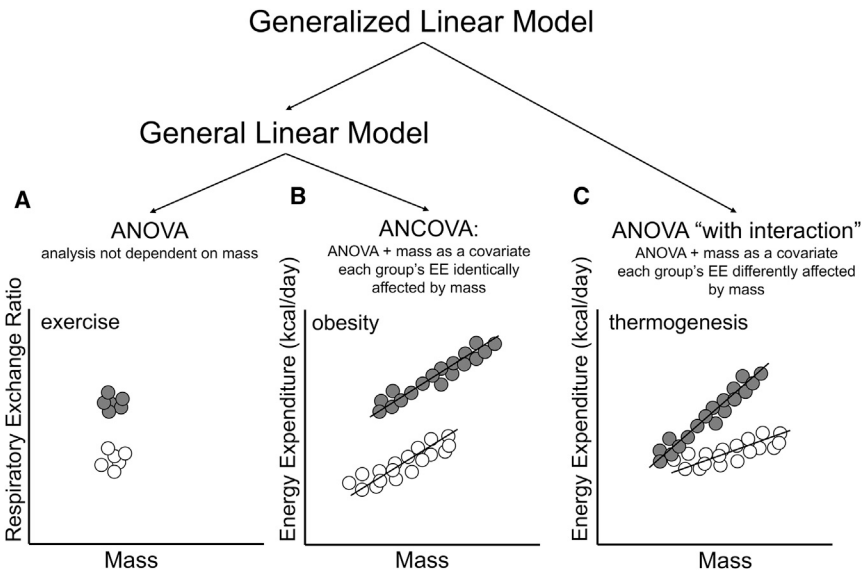


Figure 4. Analysis Models of EE Based on the General Linear Model

CalR determines the appropriate statistical model from the experimental data.

(A) ANOVA is applied where mass or body composition is not expected to affect the metabolic parameter. A sedentary group (gray) compared with an exercised group (white) with similar mass would use the ANOVA to examine group differences in RER.

(B) The ANCOVA (ANOVA with the addition of a covariate) when mass is significantly different, but slopes are parallel. This model could interpret an obese group of mice (gray) compared with lean controls (white) in which greater mass and EE are observed.

(C) The GLM can interpret the different effect of mass on EE between groups by including an interaction effect. Mice with greater BAT mass (gray) with a more pronounced thermogenic response could be interpreted by this model.

the main predictor variable, "mass" as the covariate, and the interaction between group and mass is included to test whether it should be kept. The GLM performed in *R* is generalizable to:

➢ glm(y ~ mass + group + mass:group, family = gaussian (link = "identity")),

where *y* is a metabolic variable in the set of VO_2, VCO_2, EE, food intake, and water intake. For the "family" parameter, we specify a Gaussian error structure, and for the "link" parameter, we specify an identity link, which indicates that no transformations are to be made. The user selects mass to be total body mass, LBM, or FM. A type III model first computes the sum of squares, but if the interaction effect of group and mass is found to be insignificant, then it is dropped from the model.

ANCOVA

When a mass-dependent parameter is estimated using the GLM and the interaction between group and mass is found to be insignificant, then the reduced GLM model resembles an ANCOVA model with the generalizable form:

➢ glm(y ~ mass + group, family = gaussian (link = "identity")).

ANOVA

The ANOVA is performed on parameters measured that are not strictly linked to body mass. The GLM model is reduced to an ANOVA with "group" as the sole predictor variable and mass is not included in this model.

The ANOVA performed in *R* is generalizable to:

➢ glm(z ~ group, family = gaussian (link = "identity")),

where *z* is a metabolic variable in the set of RER, locomotor activity, ambulatory activity, body temperature, and wheel running. These variables are independent of mass.

The analysis table of *CalR* is populated with the results of the ANCOVA or GLM computed with the covariate selected by the user (total mass, lean mass, or FM.) For each metabolic variable measured, *CalR* reports the effects of the group, mass, and interaction of mass and group (if significant). For this model to work, all of the included subjects must have their masses specified.

Post Hoc

For experiments with analysis of more than two groups, Tukey's honest significant difference post hoc test is performed and graphed to display confidence intervals. For the selected metabolic variable, the "Analysis" tab presents the mean difference with a 95% confidence interval for all pairwise group comparisons.

➢ glht(glm.model, mcp (group = "Tukey")),

where *glm.model* is one of the aforementioned models depending on the metabolic parameter and whether or not the group by mass interaction effect is significant for mass-dependent parameters (i.e., ANOVA, ANCOVA, and ANOVA with interaction).

Automatic Outlier Detection

Within the time range selected, the group means and SDs are calculated and stratified into light and dark photoperiods. If "Yes" is selected for the "remove outlier" radio button, the values that fall beyond 3 SDs from the group mean for the respective light/dark period will be excluded from the analysis. Since VO_2, VCO_2, EE, and RER are interdependent, then the removal of data for one of these variables will lead to the removal of the data for all of them at the corresponding time point. Additionally, the indirect calorimetry apparatus is both complex and error-prone. Exclusion of data from malfunctioning feeders (e.g., momentary readings of negative food intake) is also justified. By default, automatic outlier detection is set to off, and users must opt in to remove these values from the analysis.

Manual Subject Exclusion

After beginning a calorimetry run, animals may require veterinary intervention or have reached a prespecified endpoint for humane euthanasia according to institutional guidelines. When one animal is physically removed from a cage, the data collected from this empty cage should cease to be included in the group analysis for the relevant experiment. Manual

data exclusion is designed to remove all data from an empty cage starting at a designated time until the conclusion of the experiment. The "Subject Exclusion" tab is automatically populated with all subject names and the exclusion set to the hour at which the experiment concludes (i.e., no data excluded). While data from excluded cages will be omitted from the analysis, no data are removed or excluded from *CalR* data files. Next to each exclusion, there is an additional text box for the user to enter their reasoning for the corresponding exclusion. All excluded data, including user annotations from manual exclusions, are saved in an exportable data file and combined with automatic outliers when in use.

Metabolic Variables Versus Time

For visualization and analysis, the data read into *CalR* is cropped to the hour range selected by the user. To generate hourly time plots, *CalR* subsets the data by either group or subject depending on user input and computes averages and standard errors of the measurements at each hour for the metabolic variable being plotted. The values for the daily bar plots are calculated for each group at each photoperiod (light, dark, or 24-hour) and stratified by day, while the overall bar plot does not stratify by day. The structure of the data used for analysis consists of the average value of the metabolic variables for each subject for any one of the photoperiods.

Body Masses and Compositions

CalR will automatically conduct unpaired two-sample t tests on all two-group combinations to compute p values that indicate if there are significant differences in the body mass averages of the groups.

```
➢ t.test(mass ~ group)
```

If, in addition to total body mass, lean and fat masses are included, then *CalR* will also conduct similar statistical analyses of the body compositions. Bar plots display the average masses (and composition), and stars above the bars represent statistical significance. When more than two groups are involved, the significance tests are pairwise with the comparisons being indicated by the start and end location of the horizontal lines drawn above the bars.

Interactive 2D and 3D Regression Plots

By default, *CalR* Regression plots display the average EE against total mass for the subjects included in the data. The default variable shown is EE because it is less prone to contain error than VO_2 and VCO_2 in open circuit indirect calorimeters (Arch et al., 2006). However, *CalR* provides the option to plot any of the following metabolic variables against body mass: VO_2, VCO_2, EE, RER, cumulative food intake, and body temperature. To examine the association between the metabolic variable and mass, we generate a plot with each subject's mass against the average value of the selected metabolic variable of the respective subject over the experimental time frame chosen by the user (light, dark, 24-hr). Lines of best fit are produced for each group, with optional shading of the range of SE. Slopes are then computed and compared by linear regression analysis. For the selected metabolic variable, time of day, and mass variable, *CalR* will automatically calculate the p values of GLM-based coefficients to indicate if a significant group, mass, or interaction effect exists. We also include the ability to plot EE versus locomotor activity, in which the activity effect on EE is reported in place of the mass effect. When lean mass and FM have been provided, the option to produce a 3D plot of EE (or other metabolic variables) versus lean mass and FM becomes available to the user. A plane of best fit, computed from a GLM that includes both lean mass and FM as covariates, is produced for each group. These 3D plots can be rotated, zoomed in, or saved as image files.

Data Analysis Example

CalR is foremost a tool to facilitate analysis of organismal energy balance experiments. Here we briefly describe one example of data analysis for an experiment of two groups with 12 WT and 10 knockout mice. For a 2-min video of this example, see Video S1. Additional step-by-step instructions for this and other examples can be found in the Supplemental Information. The user's first step is to select and launch the two-group analysis template by navigating from the *CalR* homepage to select "Templates," and from the drop-down, selecting "Two groups." Using the "Input" tab of the GUI (Figure 5A) in the first column, raw data files or *CalR* data files are selected and uploaded. In the second column, we also wish to add body composition data by uploading a spreadsheet with animal identifiers, lean mass, FM, and total mass. In the third column, we assign group names and subject identifiers to the appropriate group. When the "Go to Plots" button is activated, the "Time Plots" tab is populated (Figure 5B). The default settings will plot oxygen consumption for the total duration of the experiment, averaged by group, with outliers removed. These and other settings can be adjusted by the user. A recommended practice is to exclude the initial 12- to 24-hr acclimation period of an experiment from the analysis (Tschöp et al., 2012). The "Generate Plot" button will populate hourly mean values in a time plot. In addition, daily mean values or overall averages will appear under the "Average Plots" tab. The bar plots are divided by photoperiod specified in the input tab (default light at 0700 and default dark at 1900 hr). Error bars are SEM. These data are plotted without normalization or adjustment to mass. In the "Regression Plots" tab (Figure 5C), we plot metabolic variables versus either one or two mass parameters (lean, fat, or total). At the bottom of the plot are values from the analysis of whether the selected variable is affected by mass, group, or interaction. Under standard room temperature housing and experimental conditions, we expect a significant mass effect ($p < 0.05$), where greater body mass will correlate to greater values for metabolic variables, including VO_2, VCO_2, EE, and Food Intake. This information can be valuable for understanding the quality of the experimental conditions. As such, an alert will be displayed if either the mass effect is not significant or a mass by group interaction effect is present. While these results can be visualized with each metabolic parameter in the regression plots tab, a full table of results is available under the Analysis Tab (Figure 5D). See Vignettes in Supplemental Information for additional examples. Another feature of note, the data being visualized (time range selected, groups specified, outliers removed) can be exported for further analysis as CSV files.

DISCUSSION

CalR provides users with much-needed comprehensive data analysis tools for indirect calorimetry experiments. Using *CalR*

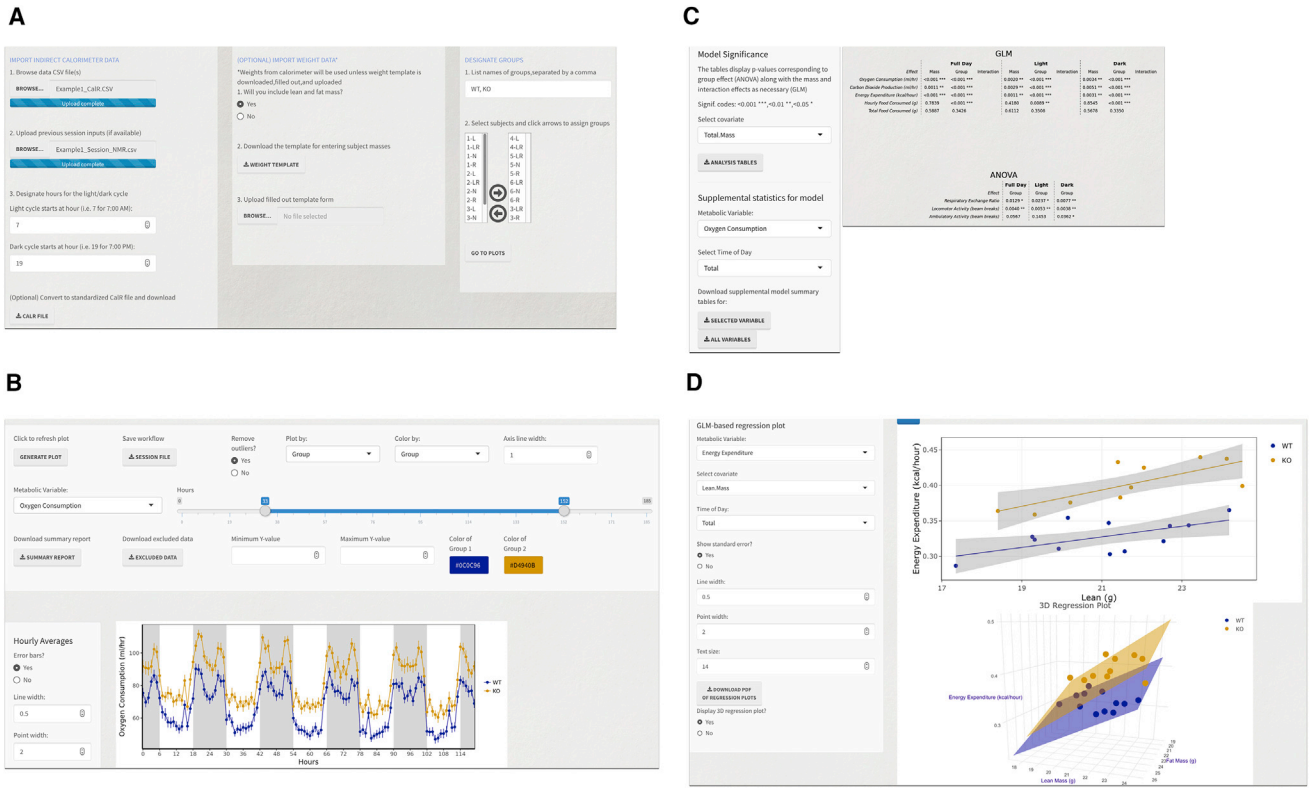


Figure 5. Example of the CalR GUI for the Two-Group Analysis

(A) Input tab for data upload and group assignments.

(B) Time Plots tab for data visualization of individual metabolic parameters for the selected period. Selected plots for oxygen consumption versus time. Shading denotes dark and light photoperiod.

(C) Results of statistical tests are shown in the "Analysis" tab.

(D) Regression Plots tab allows for analysis of metabolic data versus mass in two or three dimensions EE versus LBM and/or FM.

will enable easy access to analyze data using the GLM, which should reduce the "recurring problem" in which EE is inappropriately normalized by allometric scaling or divided by LBM in mice of different body compositions (Arch et al., 2006; Butler and Kozak, 2010). Indeed, *CalR* removes much of the guesswork facing investigators who are understandably more interested in biological interpretation and less focused on learning specialized statistical methods. Rigorous hypothesis testing requires formally defining the model and parameters prior to analysis. We implement a defined computational and statistical procedure, tackling and obviating the pressing issues that accompany the time-demanding, highly variable, and otherwise virtually untraceable steps typically taken toward generating complete analyses.

ANCOVA has been the method of choice for indirect calorimetry experiments, as it efficiently models the effect of mass on multiple metabolic variables. However, by definition, the ANCOVA cannot analyze a differential interaction between mass and group on EE (Figure 4). One approach to circumvent these limitations is using the Johnson-Neyman procedure (D'Alonzo, 2004) to find and analyze regions where there is no significant interaction between groups. While the Johnson-Neyman procedure permits the ANCOVA to be performed, it is accompanied by the dual drawbacks of (1) excluding data that decreases already limited statistical power and (2) possibly compromising interpretations in cases where the interaction effect may be biologically relevant. Although ANCOVA has been widely recognized as a suitable model for indirect calorimetry data analysis, it is essential to have the data drive the decision on which models to use. By transitioning from classical linear regression (ANOVA or ANCOVA) to GLM, assumptions of normality and constancy of variance within the samples being tested are no longer required (McCullagh and Nelder, 1998). In the GLMs, interaction effects are included when they are statistically significant. Since the interaction effect could be an essential component of an experimental metabolic story, it is added in *CalR* to provide a more complete and applicable analysis.

CalR will allow for the sharing of raw data files of experiments across calorimetry platforms. This will enable whole-body physiology to join the broader trend in biomedical research led by genomics and transcriptomics. The ability to efficiently share files as supplemental data will foster increased transparency and reproducibility with the cooperation of interested investigators. We propose a centralized repository of *CalR* indirect calorimetry data files to accelerate global research into metabolism and whole-body physiology.

Despite its many existing features, *CalR* is designed to support further innovation. Specifically, newer calorimetry systems

provide more frequent sampling times and higher-resolution understanding of metabolic parameters. However, this nuance is lost in ANOVA/ANCOVA/GLM analyses. Regardless of the high-resolution time data, the consensus approach of ANCOVA-like analysis depends on a single mean value per metabolic variable per mouse (Speakman et al., 2013; Tschöp et al., 2012; MMPC energy expenditure analysis page, https://www.mmpc.org/shared/regression.aspx). Future advances may be possible by implementing time series and body composition into a statistical framework for indirect calorimetry. *CalR* opens the door for a dialogue on how indirect calorimetry data should best be analyzed under different experimental conditions. What *CalR* offers investigators is the architecture for ensuring that their indirect calorimetry data are properly handled and analyses are conducted in a transparent and reproducible manner.

Limitations

There are notable caveats to consider with implementing *CalR* for analysis. Good experimental design is critical for reproducible analysis. Small sample size or short run times may compromise experimental interpretation due to large individual animal variation. We recommend guidelines for experimental design for indirect calorimetry experiments (Tschöp et al., 2012). *CalR* cannot detect whether a calorimeter is out of calibration and may, therefore, return results that are predicated on faulty data. As with any experimental system, quality control is dependent on the rigorous upkeep of the instrument and vigilance of the operator. Even under optimal conditions, animals may become sick or equipment failure can spoil the appearance of an experiment. *CalR* provides tools that will allow for the exclusion of data from cages where animals have been removed from an experiment for humane reasons. *CalR* also provides the ability to combine multiple experimental runs to help overcome the difficulties in generating sufficient numbers of sex- and age-matched littermates for the study of genetically modified mouse lines. The investigator is responsible for justifying the appropriateness of two groups being joined for analysis, as *CalR* cannot. Another important consideration is to look beyond the p values for interpretation of the results. Just as, if not more, important is the effect size between the groups and physiological relevance of the result. Since sample size is highly influential on the ability to detect a true difference, and it is often challenging to have a large sample size for these types of experiments (Kaiyala, 2014), it would be unreasonable to discard (or fail to publish) results that are not statistically significant but show potentially biologically important effect sizes (Gelman and Stern, 2006).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes one method file and one video and can be found with this article online at https://doi.org/10.1016/j.cmet.2018.06.019.

ACKNOWLEDGMENTS

Testing of this program was supported by help from a Boston-centered user group: Beth Israel Deaconess Medical Center: Marie Mather, Bhavna Desai, and Terry Maratos-Flier; Brigham and Women's Hospital: Dimitrije Cabarkapa, C.J. Bare, and Barbara Calderone; Boston University Medical Center: Tom Balon; Massachusetts General Hospital: Joe Brancale, Agostina Santoro, Lee Kaplan, Ben Zhou, Lianfeng Wu, and Alex Soukas; University of Massachusetts Medical Center: Jason Kim and Hye Lim Moh. We thank Owen McGuiness and Karl Kaiyala for helpful discussions. We are grateful for financial support from the NIDDK Mouse Metabolic Phenotyping Centers (MMPC, www.mmpc.org) under the MICROMouse Program, grants DK076169 and DK115255, and from the Harvard Digestive Disease Center, DK034854. NIH Instrumentation Award OD020100 (A.S.B.), ARRA supplement DK048873-14S2 (D.E.C.), DK048873 (D.E.C.), DK056626 (D.E.C.), DK103046 (D.E.C.), and DK107717 (A.S.B.).

AUTHOR CONTRIBUTIONS

Conceptualization, A.I.M., A.S.B., and D.E.C.; Methodology, A.I.M., A.S.B., and R.A.L.; Software, A.I.M. and R.A.L.; Validation, A.I.M., A.S.B., L.L., and K.B.L.; Writing, A.I.M., A.S.B., D.E.C., and L.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 9, 2018
Revised: April 22, 2018
Accepted: June 19, 2018
Published: July 12, 2018

REFERENCES

Allison, D., Paultre, F., Goran, M., Poehlman, E., and Heymsfield, S. (1995). Statistical considerations regarding the use of ratios to adjust data. Int. J. Obes. Relat. Metab. Disord. *19*, 644–652.

Arch, J., Hislop, D., Wang, S., and Speakman, J. (2006). Some mathematical and technical issues in the measurement and interpretation of open-circuit indirect calorimetry in small animals. Int. J. Obes. (Lond.) *30*, 1322–1331.

Betz, M.J., and Enerback, S. (2018). Targeting thermogenesis in brown fat and muscle to treat obesity and metabolic disease. Nat. Rev. Endocrinol. *14*, 77–87.

Butler, A.A., and Kozak, L.P. (2010). A recurring problem with the analysis of energy expenditure in genetic models expressing lean and obese phenotypes. Diabetes *59*, 323–329.

Calle, E.E., Rodriguez, C., Walker-Thurmond, K., and Thun, M.J. (2003). Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N. Engl. J. Med. *348*, 1625–1638.

D'Alonzo, K.T. (2004). The Johnson-Neyman procedure as an alternative to ANCOVA. West J. Nurs. Res. *26*, 804–812.

Drucker, D.J. (2016). Never waste a good crisis: confronting reproducibility in translational research. Cell Metab. *24*, 348–360.

Fisher, R.A. (1947). The analysis of covariance method for the relation between a part and the whole. Biometrics *3*, 65–68.

Flier, J.S. (2017). Irreproducibility of published bioscience research: diagnosis, pathogenesis and therapy. Mol. Metab. *6*, 2–9.

Gelman, A., and Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. Am. Stat. *60*, 328–331.

Guyenet, S.J., and Schwartz, M.W. (2012). Regulation of food intake, energy balance, and body fat mass: implications for the pathogenesis and treatment of obesity. J. Clin. Endocrinol. Metab. *97*, 745–755.

Himms-Hagen, J. (1997). On raising energy expenditure in ob/ob mice. Science *276*, 1132.

Jarvis, M.F., and Williams, M. (2016). Irreproducibility in preclinical biomedical research: perceptions, uncertainties, and knowledge gaps. Trends Pharmacol. Sci. *37*, 290–302.

Kaiyala, K.J. (2014). Mathematical model for the contribution of individual organs to non-zero y-intercepts in single and multi-compartment linear models of whole-body energy expenditure. PLoS One *9*, e103301.

Kaiyala, K.J., and Schwartz, M.W. (2011). Toward a more complete (and less controversial) understanding of energy expenditure and its role in obesity pathogenesis. Diabetes *60*, 17–23.

Kaiyala, K.J., Morton, G.J., Leroux, B.G., Ogimoto, K., Wisse, B., and Schwartz, M.W. (2010). Identification of body fat mass as a major determinant of metabolic rate in mice. Diabetes *59*, 1657–1666.

Katch, V. (1972a). Correlational υ ratio adjustments of body weight in exercise-oxygen studies. Ergonomics *15*, 671–680.

Katch, V.L. (1972b). Use of the oxygen-body weight ratio in correlational analyses: spurious correlations and statistical considerations. Med. Sci. Sports *5*, 253–257.

Kowalski, C.J., Schneiderman, E.D., and Willis, S.M. (1994). ANCOVA for nonparallel slopes: the Johnson-Neyman technique. Int. J. Biomed. Comput. *37*, 273–286.

Kronmal, R.A. (1993). Spurious correlation and the fallacy of the ratio standard revisited. J. Roy. Stat. Soc. Stat. Soc. *156*, 379–392.

McCullagh, P., and Nelder, J.A. (1998). Generalized Linear Models (Chapman & Hall/CRC).

Pelleymounter, M., Cullen, M., Baker, M., Hecht, R., Winters, D., Boone, T., and Collins, F. (1995). Effects of the obese gene product on body weight regulation in ob/ob mice. Science *269*, 540–543.

R Core Team. (2017). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Speakman, J.R., Fletcher, Q., and Vaanholt, L. (2013). The '39 steps': an algorithm for performing statistical analysis of data on energy intake and expenditure. Dis. Models Mech. *6*, 293–301.

Stanford, K.I., Middelbeek, R.J., Townsend, K.L., An, D., Nygaard, E.B., Hitchcox, K.M., Markan, K.R., Nakano, K., Hirshman, M.F., Tseng, Y.H., et al. (2013). Brown adipose tissue regulates glucose homeostasis and insulin sensitivity. J. Clin. Invest. *123*, 215–223.

Tanner, J. (1949). Fallacy of per-weight and per-surface area standards, and their relation to spurious correlation. J. Appl. Physiol. *2*, 1–15.

Tschöp, M.H., Speakman, J.R., Arch, J.R., Auwerx, J., Brüning, J.C., Chan, L., Eckel, R.H., Farese, R.V., Jr., Galgani, J.E., and Hambly, C. (2012). A guide to analysis of mouse energy metabolism. Nat. Methods *9*, 57–63.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
CalR	This paper	OMICtools: OMICS_23756 RRID: SCR_015849 https://CalRapp.org

CONTACT FOR RESOURCE SHARING

Further information and should be directed to and will be fulfilled by the Lead Contact, Alexander Banks (abanks@bwh.harvard.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Eight examples are included in the Methods S1 File. The subject details for the mice in these studies are described below. All animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of the Harvard Center for Comparative Medicine (HMS), the Brigham and Women's Hospital (BWH), or Vanderbilt University. Some of the data sets were artificially contrived to illustrate a specific aspect of CalR, while others are presented as collected. These are indicated below.

Examples 1 and 2 are idealized synthetic datasets meant to illustrate two groups of mice with and without an interaction effect between mass and all mass-dependent metabolic variables.

Example # and Description	Genetic Background	Sex	Age (wk, Approx.)	Diet	IACUC Approval
1. Two-group ANCOVA					
2. Two-group interaction					
3. Two-group acute response	C57Bl/6	M	20	60% HFD	BWH
4. Three-group (ordered)	C57Bl/6 x 129S/v mixed	M	16	Standard Chow Picolab 5053	HMS
5. Three-group (un-ordered/factorial)	C57Bl/6	M	16	Standard Chow Picolab 5053	BWH
6. Four-group	C57Bl/6	M	20	Standard Chow and 60% HFD	BWH
7. Combining Runs	C57Bl/6	M	14	Standard Chow Picolab 5053	HMS
8. Two-group crossover design	C57Bl/6	M	15	Standard Chow Picolab 5053	Vanderbilt

METHOD DETAILS

The experimental parameters for each of the eight example data sets are indicated below.

Example # and Description	Equipment	Ambient Temperatures	Data Described in Methods S1
1. Two-group ANCOVA		Room temp	After acclimation
2. Two-group interaction		Room temp	After acclimation
3. Two-group acute response	CLAMS	22C→30C→CL	CL injection
4. Three-group (ordered)	CLAMS	30C→22C→ 4C→22C	4C
5. Three-group (un-ordered/factorial)	CLAMS	30C→4C	30C→ 4C
6. Four-group	CLAMS	22C→30C→CL	22C
7. Combining Runs	CLAMS	23C	23C
8. Two-group crossover design	Prometheon	23C	23C

*CL, beta-3 adrenergic agonist CL-316,243.

DATA AND SOFTWARE AVAILABILITY

Example data sets 1-8 can be found in Methods S1. CalR can be accessed at https://CalRapp.org.