Received: 10 March 2023 | Revised: 14 June 2023 | Accepted: 28 July 2023

DOI: 10.1111/phen.12416

ORIGINAL ARTICLE

Physiological Entomology: How Insects Work - Linking Genotype to Phenotype

Royal Entomological Society

# Sample size assessments for thermal physiology studies: An R package and R Shiny application

Clarke J. M. van Steenderen[1] | Guy F. Sutton[1] | Candice A. Owen[2] | Grant D. Martin[1,3] | Julie A. Coetzee[1]

[1]Centre for Biological Control, Department of Zoology and Entomology, Rhodes University, Makhanda/Grahamstown, South Africa

[2]Department of Biological Sciences, University of Chester, Chester, UK

[3]Afromontane Research Unit, Department of Zoology and Entomology, University of the Free State, Phuthaditjhaba, South Africa

**Correspondence**
Clarke J. M. van Steenderen, Centre for Biological Control, Department of Zoology and Entomology, Rhodes University, Makhanda/Grahamstown, Eastern Cape 6139, South Africa.
Email: vsteenderen@gmail.com; clarke.vansteenderen@ru.ac.za

**Funding information**
National Research Foundation; South African Research Chairs Initiative of the Department of Science and Technology; Working for Water (WfW) programme of the Department of Environmental Affairs: Natural Resource Management programme (DEA: NRM)

## Abstract

Required sample sizes for a study need to be carefully assessed to account for logistics, cost, ethics and statistical rigour. For example, many studies have shown that methodological variations can impact the critical thermal limits (CTLs) recorded for a species, although studies on the impact of sample size on these measures are lacking. Here, we present `ThermalSampleR`; an R CRAN package and Shiny application that can assist researchers in determining when adequate sample sizes have been reached for their data. The method is particularly useful because it is not taxon specific. The Shiny application offers a user-friendly interface equivalent to the package for users not familiar with R programming. `ThermalSampleR` is accompanied by an in-built example dataset, which we use to guide the user through the workflow with a fully worked tutorial.

**KEYWORDS**
CTL, R Shiny, sample size, thermal tolerance

## INTRODUCTION

Insufficient sample sizes in a study represent a waste of resources by not having the power to reliably detect patterns in the data, which can lead to incorrect inferences and inappropriate management interventions (Duffy et al., 2021). Oversized studies consume more resources than is necessary, which imposes unnecessary costs and provides little improvement in the ability to answer particular ecological research questions (Forcino et al., 2015). For studies that involve animals, and particularly threatened species, sample size determination is important for ethical reasons too (Duffy et al., 2021). Indeed, many journals, institutions and ethics committees require that researchers justify the number of samples used during the study (Hampton et al., 2019), which should be determined as the minimum sample size necessary to achieve the goals of the study (Fitts, 2011).

Recently, several studies have shown that the results and inferences obtained from thermal tolerance studies can be significantly affected by methodological choices when designing and performing the experiments, such as the use of a pre-experimental acclimation period, the temperature ramping rate and ramping intervals (e.g., Chown et al., 2009; Nyamukondiwa & Terblanche, 2009; Rezende et al., 2014). Similarly, Duffy et al. (2021) demonstrated that the number of individuals tested during thermal tolerance studies (sample size) can significantly bias the results obtained and any inferences drawn from these studies. Determining the sample size requirements for a study is an essential component of study design, which

can have serious consequences for the logistics, cost, ethics and statistical rigour of the study (Arnold et al., 2011; Gerrodette, 1987).

Insect thermal limit studies are plentiful and therefore offer an ideal source of data for exploring sample size requirements. Most insects are poikilothermic ectotherms (Sinclair et al., 2015), and so their bodily functions and life history characteristics are strongly correlated to the ambient microclimate (Neven, 2000; Nguyen et al., 2014; Sinclair et al., 2015). To survive and reproduce, insect body temperatures need to be maintained within the limits of their thermal tolerance range (Koštál et al., 2011; Nguyen et al., 2014). As such, thermal tolerances can be used to explain the geographical distributions (Rezende et al., 2014; Sinclair et al., 2015) and the performance of insects under different environmental conditions (Nguyen et al., 2014; Nyamukondiwa & Terblanche, 2009; Sinclair et al., 2015). In this vein, thermal tolerance investigations have been used to determine the establishment and spread of insect pests (Wang et al., 2019) and biological control agents (Coetzee et al., 2007). The applicability of these studies has increased recently as researchers aim to forecast changes in faunal and floral assemblages under current and future climate change scenarios (Bennett et al., 2018; Duffy et al., 2015; Rezende et al., 2014).

In this paper, we present `ThermalSampleR`—an R package and R Shiny graphical user interface (GUI) application that allows users to easily assess the sample sizes required to obtain reliable and accurate thermal physiology parameters (e.g., critical thermal limits [$CT_{min}$/$CT_{max}$]). `ThermalSampleR` is designed to make analysing sample size requirements simple and provide easily interpretable summary statistics. The Shiny GUI provides the functionality of the full R package to researchers with little to no experience in R.

## PACKAGE BACKGROUND

Several tools and analyses have been developed to aid in sample size planning for biological studies, primarily focusing on the use of power calculations (Peterman, 1990; Toft & Shea, 1983). The power of a statistical test refers to the probability that the test correctly rejects the null hypothesis. However, power calculations are centred on assessing whether sample sizes are large enough to detect a statistically significant difference between groups (i.e., correctly rejecting the null hypothesis using a *p*-value). They are, therefore, of little use for estimating the critical thermal limit (CTL) of a single population or assessing the accuracy and precision of between-group differences in thermal tolerance parameters. To remedy this, many researchers have adopted the practice of calculating the effect of sizes (e.g., difference in means/medians) and 95% confidence intervals (CIs) for a more rigorous and intuitive method to make comparisons amongst groups, rather than by simply relying on a *p*-value (Gardner & Altman, 1986; Halsey, 2019; Nakagawa & Cuthill, 2007). Practitioners need to consider sample size planning for both power and accuracy in parameter estimation (AIPE), which both require different statistical approaches (Maxwell et al., 2008).

To account for sample size planning for both power and AIPE, the `ThermalSampleR` package uses simulation and bootstrap resampling procedures to calculate population parameters and CIs (Maxwell et al., 2008). The CI approach to power planning has the added benefit (as compared to obtaining a *p*-value) of indicating a direction of effect. Moreover, CIs can be used to assess sample size planning for AIPE by computing and controlling the CI of the parameter of interest (Maxwell et al., 2008). This contains two distinct components: (1) planning for accuracy, whereby researchers assess the probability that the CI contains the true population parameter of interest (e.g., $CT_{min}$/$CT_{max}$), and (2) planning for precision, where precision is measured by the width of the CI (i.e., a smaller CI width indicates a more precise estimate of the population parameter; Maxwell et al., 2008).

CTL studies can be divided into two broad categories: single-sample studies and multiple-group comparison studies. Single-sample studies use an estimate of a population parameter of interest, such as the $CT_{min}$/$CT_{max}$ of a single population of a species. These kinds of studies are usually descriptive, or may be of interest to predict where the best release sites could be in the country of introduction for a new biocontrol agent, or how insects could be expected to respond to climate change (e.g., Coetzee et al., 2007). Two- or multiple-group comparison studies use an estimate of the possible difference in CTLs between different groups. Examples of these kinds of studies include, amongst others, those where multiple species or populations of a biological control agent need to be compared to determine which would be better suited for release at a specific site, or where the CTLs of groups exposed to different environmental conditions are compared to determine whether acclimation is possible (e.g., Porter et al., 2019). The functions provided within the `ThermalSampleR` package are distinguished by whether the experimental data originates from a single-sample (**boot_one()** and **plot_one_group()**) or multiple-group comparisons study (**boot_two()** and **boot_two_groups()**).

## TUTORIAL

The following tutorial illustrates the core functions available within the `ThermalSampleR` package. Our goal is to provide an easy-to-follow and fully reproducible analysis of both a sample size assessment for (a) a single-sample study and (b) a multiple-groups comparison study.

### Package installation

`ThermalSampleR` can be accessed by running one of the options below in R:

1. Via the CRAN repository

```
install.packages("ThermalSampleR")
```

2. GitHub

```
devtools::install_github
("clarkevansteenderen/ThermalSampleR")
```

3. The R Shiny GUI can be accessed directly on the R console by running

```
library(shiny)
shiny::runUrl(
"https://github.com/clarkevansteenderen/
ThermalSampleR_Shiny/
archive/main.tar.gz")
```

or via the link to the R Shiny application server:
https://clarkevansteenderen.shinyapps.io/ThermalSampleR_Shiny/

## Data structure

This tutorial uses the **coreid_data** dataset as an example, which is a data frame/tibble included in the package. This dataset represents the $CT_{min}$ data for the twig-wilting bug *Catorintha schaffneri* (Hemiptera: Coreidae), a biological control agent introduced into South Africa from Brazil to control the invasive cactus *Pereskia aculeata* Miller (Cactaceae; Muskett et al., 2020). The dataset contains two columns, the first being **col**, which contains a unique identifier label (e.g., a species/taxon/population name), distinguishing data obtained from adults (**Catorhintha_schaffneri_APM**) or nymphs (**Catorhintha_schaffneri_NPM**). The second column, **response**, contains a numeric vector containing our response variable, the $CT_{min}$ value (in °C). Each row represents a unique individual that was tested during the experiment. Before starting any analyses, we can inspect the raw data:

```
head(ThermalSampleR::coreid_data)
## col response
## 1 Catorhintha schaffneri_APM 5
## 2 Catorhintha schaffneri_APM 5
## 3 Catorhintha schaffneri_APM 5
## 4 Catorhintha schaffneri_APM 4
## 5 Catorhintha schaffneri_APM 4
## 6 Catorhintha schaffneri_APM 4
```

## Sample size assessment—single sample

The simplest application of `ThermalSampleR` is to evaluate whether a study has used a sufficient sample size to accurately estimate a parameter of interest for a single taxon. Below, we illustrate this by performing these calculations to estimate sample sizes required to accurately estimate the $CT_{min}$ of adults of *C. schaffneri* (denoted by **Catorintha_schaffneri_APM** in **coreid_data**; Muskett et al., 2020). This simulation uses a bootstrap resampling procedure to estimate the width of the 95% CI of the parameter of interest estimate across a range of sample sizes, which defaults to starting at $n = 3$ individuals tested, and which can be extrapolated to sample sizes greater than the sample size of the existing data by specifying a value to **n_max**:

```
# Set a seed to make the results
reproducible, for illustrative purposes.
  set.seed(2012)
# Perform simulations
ThermalSampleR::bt_one = boot_one(
   # Which dataframe does the data come from?
   data = coreid_data,
   # Provide the column name containing the
   taxon ID
   groups_col = col,
   # Provide the name of the taxon to be
   tested
   groups_which = "Catorhintha_schaffneri_
   APM",
   # Provide the name of the column
   containing the response variable (e.g
   CTmin data)
   response = response,
   # Maximum sample to extrapolate to
   n_max = 49,
   # How many bootstrap resamples should be
   drawn?
   iter = 299)
```

The variable containing the bootstrap resamples should then be passed to the **plot_one_group()** function to visualise the simulation results. A number of optional parameters can be passed to the function to alter the aesthetics of the graphs:

```
ThermalSampleR::plot_one_group(
   # Variable containing the output from
   running boot_one() function
   x = bt_one,
   # Minimum sample size to plot
   n_min = 3,
   # Actual size of your existing dataset
   n_max = 15,
   # Colour for your experimental data
   colour_exp = "forestgreen",
   # Colour for the extrapolated predictions
   colour_extrap = "orange",
   # Position of the legend
   legend.position = "right")
```

Inspecting Figure 1a, we visualise the precision of our $CT_{min}$ estimate for adult *C. schaffneri*, whereby precision is measured as the width of a 95% CI. For example, in the context of CTLs, a CI width of 1 indicates that practitioners can be 95% confident that their CTL estimate is within 1°C of the true $CT_{min}$ value. The smaller the CI width, the greater the precision of the CTL estimate. In this example, the precision of our $CT_{min}$ estimate was high and was not predicted to improve substantially by increasing sample size once approximately $n = 20$ individuals were tested, as the 95% CI reached a plateau at $n = 20$. The plateau is in the extrapolation section of the graph

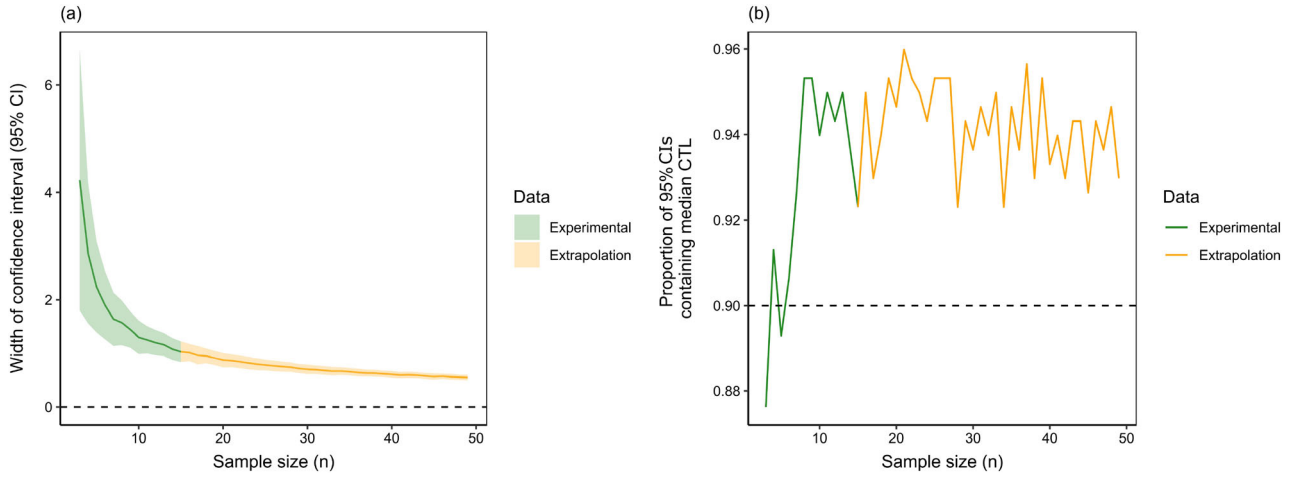


**FIGURE 1** Results of the single-sample sample size assessment for *Catorintha schaffneri* adults. Panel (a) shows the precision of the $CT_{min}$ estimate for *C. schaffneri*, where precision is measured as the width of a 95% confidence interval. Panel (b) shows the sampling distribution (i.e., the range of plausible $CT_{min}$ values) for *C. schaffneri*.

indicating that more individuals would need to be tested for the 95% CI to become approximately stable. However, at the existing sample size of $n = 15$, the researchers could be relatively confident that the $CT_{min}$ estimate they have obtained is precise to within approximately 1.2–1.5°C. Researchers will need to decide for themselves what an acceptable degree of precision is for their own datasets.

Inspecting Figure 1b, we visualise the sampling distribution (i.e., the range of plausible $CT_{min}$ values) for the taxa under study. This assessment can produce biased results at small sample sizes because the population parameter (e.g., the taxon's $CT_{min}$) is unknown and must therefore be estimated from the experimental data. Figure 1b gives an indication of parameter estimation accuracy by plotting the proportion of bootstrap resamples across each sample size for which the 95% CI included the estimated population parameter. An accurate parameter estimate should produce CIs that, on ~95% of occasions, contain the estimated population parameter. In this example, the accuracy of our $CT_{min}$ estimate was high once $n > 10$ individuals were tested. As noted above, because the true population parameter is estimated from the raw data, this analysis of parameter accuracy may be biased, and thus, should be interpreted with caution.

*Take-home message*: As long as the researchers were content with obtaining a $CT_{min}$ estimate for adult *C. schaffneri* with a precision of approximately 1.2–1.5°C, the experiment could be concluded at $n = 15$ individuals tested. Adding additional samples above $n = 15$ would likely improve the precision of the $CT_{min}$ estimate; however, the gain in precision must be considered in light of the logistics, costs and ethics of testing additional specimens.

## Sample size assessment—Multiple-group comparisons

`ThermalSampleR` also allows the user to estimate sample size adequacy for studies comparing the CTLs across multiple groups (e.g., testing for differences in $CT_{min}$ between different taxa, populations, treatments applied and sexes). For example, the built-in example data (**coreid_data**) in `ThermalSampleR` contains $CT_{min}$ data for 30 adults and 30 nymphs of *C. schaffneri*. Researchers may be interested in determining whether releasing adults or nymphs would lead to better establishment rates in the field. As such, the researchers could assess the $CT_{min}$ of each life stage and use these data to release the life stage with the lower $CT_{min}$ value as they would be assumed to better tolerate low temperatures. To do this, we apply a similar workflow as per the 'single sample' assessments in the previous section. We use a bootstrap resampling procedure to estimate the width of the 95% CI of the difference in $CT_{min}$ estimates between our two groups of interest (*C. schaffneri* adults vs. nymphs) across a range of sample sizes:

```
# Set a seed to make the results
reproducible, for illustrative purposes.
  set.seed(2012)
# Perform simulations
ThermalSampleR::bt_two = boot_two(
   # Which dataframe does the data come from?
   data = coreid_data,
   # Provide the column name containing the
   taxon ID
   groups_col = col,
   # Provide the name of the column
   containing the response variable (e.g
   CTmin data)
   response = response,
   # Provide the name of the first taxon to be
   compared
   group1 = "Catorhintha_schaffneri_APM",
   # Provide the name of the second taxon to
   be compared
   group2 = "Catorhintha_schaffneri_NPM",
   # Maximum sample to extrapolate to
   n_max = 49,
   # How many bootstrap resamples should be
   drawn?
   iter = 299)
```

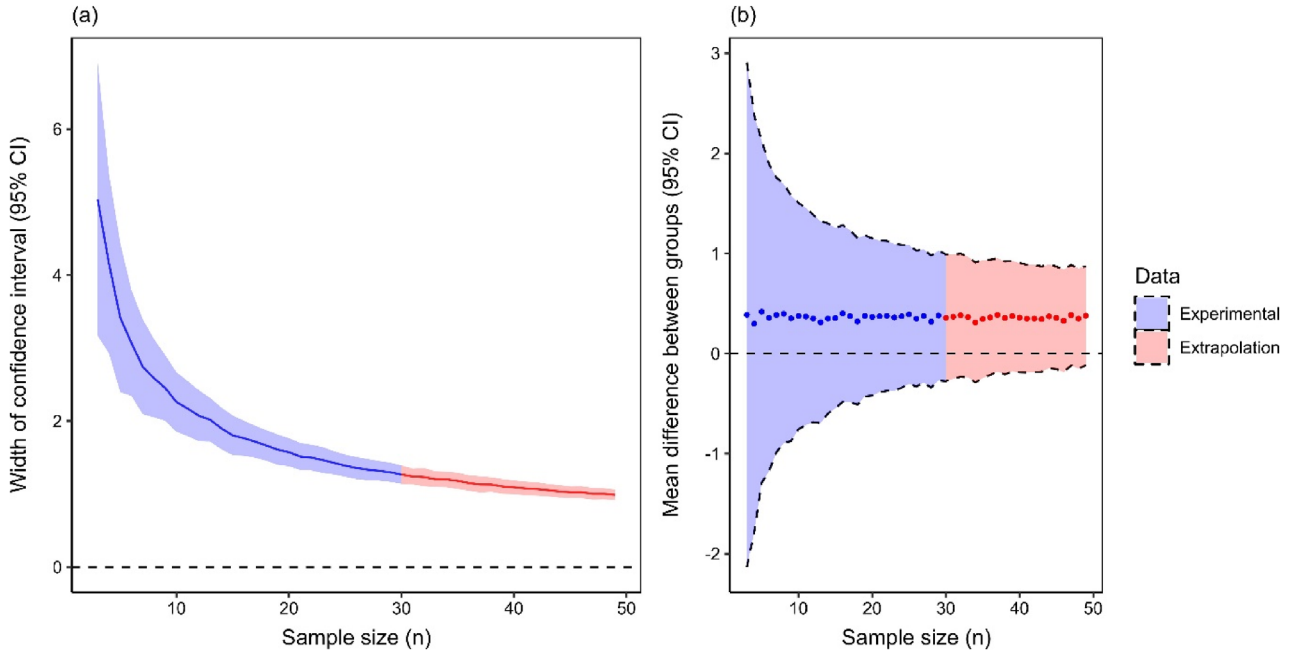


**FIGURE 2** Results of the sample size assessments for the comparison of *Catorintha schaffneri* adults and nymphs. Panel (a) shows the precision of the estimate for the difference in $CT_{min}$ for *C. schaffneri* adults versus nymphs across sample sizes. Panel (b) shows the 95% confidence interval of the mean difference in $CT_{min}$ between adults and nymphs.

The variable containing the bootstrap resamples should then be passed to the **plot_two_groups()** function to visualise the simulation results. A number of optional parameters can be passed to the function to alter the aesthetics of the graphs:

```
ThermalSampleR::plot_two_groups(
   # Variable containing the output from
   running the boot_two() function
   x = bt_two,
   # Minimum sample size to plot
   n_min = 3,
   # Actual size of your existing dataset
   n_max = 30,
   # Colour for your experimental data
   colour_exp = "blue",
   # Colour for the extrapolated predictions
   colour_extrap = "red",
   # Position of the legend
   legend.position = "right",
   # Change the degree of shading on the
   graph
   alpha_val = 0.25)
```

Figure 2a can be interpreted analogously to Figure 1a produced during the 'single sample' assessments in the previous section. Here, we are visualising the precision of our estimate for the difference in $CT_{min}$ of *C. schaffneri* adults versus nymphs across sample sizes. In this example, where $n = 30$ individuals were tested for both adults and nymphs of *C. schaffneri*, the precision of our estimated difference between the groups was high and is not predicted to improve substantially by increasing sample size, as the 95% CI reached a plateau at approximately $n = 25$. At $n = 30$, the researchers could be relatively confident that the difference in $CT_{min}$ between adults and nymphs could be estimated to within 1.5°C precision. The researchers will need to decide for themselves what an acceptable degree of precision is for their own datasets. In Figure 2b, we visualise the 95% CI of the mean difference in $CT_{min}$ between adults and nymphs. At $n = 30$ individuals tested, it appears that the $CT_{min}$ of one group (*C. schaffneri* adults) may be slightly higher than for nymphs. However, the 95% CI overlaps 0, indicating that the $CT_{min}$ of adults and nymphs are unlikely to be significantly different. Moreover, limits of the 95% CI are relatively stable, indicating that adding additional samples is unlikely to change the results obtained.

*Take-home message*: As long as the researchers were content with obtaining an estimate for the difference in $CT_{min}$ between *C. schaffneri* adults and nymphs with a precision of approximately 1.5°C, the experiment could be concluded at $n = 30$ individuals tested. Adding additional samples above $n = 30$ would likely improve the precision of estimate; however, the gain in precision appears minimal and must be considered in light of the logistics, costs and ethics of testing additional specimens.

## Sample size assessment—Test of Total Equivalency

Duffy et al. (2021) adopted a slightly different approach for assessing sample size requirements for CTL studies by using an equivalency

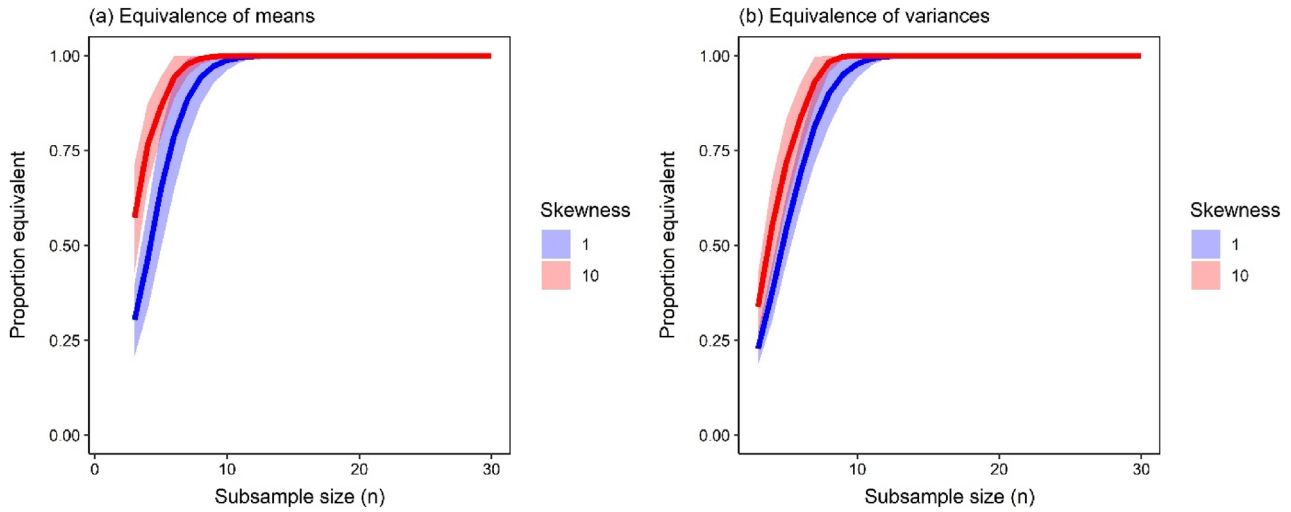


**FIGURE 3** Test of Total Equivalency output for *Catorintha schaffneri*. Panel (a) shows the equivalence of means, and panel (b) shows the equivalence of variances. Both graphs are simulated for low (1) and high (10) skewness in the data and show a plateau in the curves.

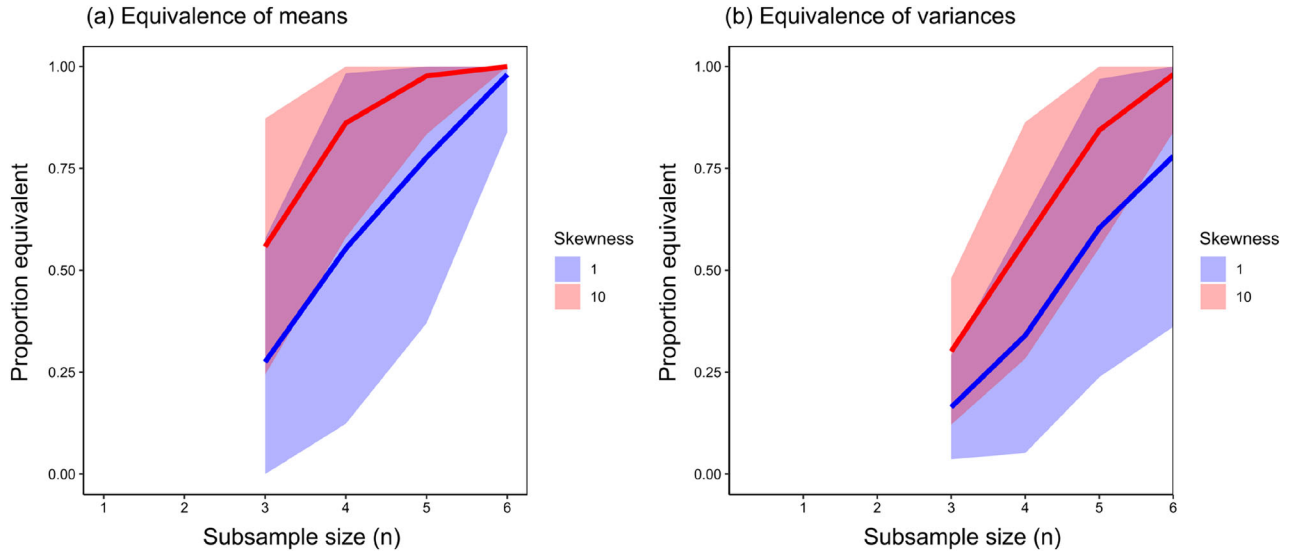


**FIGURE 4** Test of Total Equivalency using only six *Catorintha schaffneri* individuals. Panel (a) shows the equivalence of means, and panel (b) shows the equivalence of variances. Both graphs are simulated for low (1) and high (10) skewness. Neither panel has reached a plateau.

testing approach. Their approach differed from ours by randomly resampling simulated datasets with varying skewness characteristics rather than resampling the raw data. Thereafter, the authors compare the mean and variance of smaller subsets of the full dataset to the full dataset using a 'two one-sided t-test' approach (Duffy et al., 2021). Tests applied were either standard one-sided *t*-tests (for normally distributed datasets) or Chen's modified one-sided *t*-test (Chen, 1995). The user can specify an equivalence margin indicating the acceptable degree of error between the data subsets and the full dataset (e.g., an equivalence margin of 1°C indicates whether the mean or variance of the thermal limit for each subsample was within 1°C of the full dataset). The value of the approach adopted by Duffy et al. (2021) is that it accounts for the often-skewed distribution of thermal limits datasets (Janion-Scheepers et al., 2018). `ThermalSampleR` allows users to calculate sample size requirements using this Test of Total Equivalency (TOTE) as developed by Duffy et al. (2021), using the **equiv_tost()** function. Using the same `coreid` dataset from the previous sections, we illustrate below how to assess sample size requirements to precisely estimate the $CT_{min}$ parameter for adult *C. schaffneri* across a range of sample sizes (i.e., in a single-sample study design):

```
tte = ThermalSampleR::equiv_tost(
  # Which dataframe does the data come from?
  data = coreid_data,
  # Provide the column name containing the
  taxon ID
  groups_col = col,
  # Provide the name of the taxon to be
  tested
```

```
  groups_which = "Catorhintha
  schaffneri_APM",
# Provide the name of the column containing
the response variable (e.g CTmin data)
  response = response,
  # Define the skewness parameters
  skews = c(1,10),
# Define the equivalence of subsets to full
population CT estimate (unit = degree
Celcius)
  equiv_margin = 1,
# Size of the population to sample (will
test subsamples of size pop_n - x against
pop_n for equivalence). Defaults to
population size = 30
  pop_n = 30
  )
```

Inspecting both panels in Figure 3 indicates that the researchers would have been able to obtain $CT_{min}$ estimates (in terms of both the mean and variance) equivalent to within 1°C of the estimates derived from the full dataset ($n = 30$) if they had tested approximately 10–12 individuals, irrespective of the skewness in the underlying data.

The more important application of the TOTE approach is to iteratively assess sample sizes during the course of the experiment. Duffy et al. (2021) recommend collecting some pilot data and then assessing the sample size requirements to estimate CT traits. For example, had we tested six insects in a pilot study and assessed the sample size requirements, we would obtain the graphs in Figure 4. It is evident that testing six individuals was not sufficient to obtain a reliable estimate of the CT trait in this example. The researchers would then add additional samples to their study (e.g., add another 10 individuals) and then retest the sample size requirements, repeating the process until the TOTE curves plateau.

## CONCLUDING REMARKS

Statistical tools that aid researchers in gaining a clearer understanding of the strengths and limitations of their analyses are essential (Dushoff et al., 2019). Duffy et al. (2021) showed that sample size is an important consideration in thermal tolerance experiments. Here, we have provided the tool for researchers to determine whether their sample sizes are appropriate or not. The tutorial workflow presented here illustrates how assessing sample size improves the understanding of results obtained from CTL studies, and we advocate for its inclusion in future insect thermal tolerance studies.

Sample size planning should be performed in the framework of sequential sampling (Kelley et al., 2018), whereby researchers iteratively perform experimental trials, collect data, complete simulations, calculate the sample size metric of choice (e.g., CIs, type S/M errors) and critically evaluate whether the required degree of certainty has been achieved to warrant terminating or continuing the study. By iteratively assessing sample size requirements, practitioners can determine whether their study is sufficiently powered, accurate and precise for their research goals, whilst simultaneously ensuring that resources are allocated efficiently and no unnecessary testing is performed. Ideally, only once accurate CTL estimates have been obtained that are sufficiently precise, would testing cease. Practically, there are circumstances where this would not be possible (e.g., such as when working with endangered species). We fully acknowledge that researchers often need to balance sample sizes with other constraining factors.

Although our simulations were run on CTL data obtained from insects, these simulations are not taxon specific. The `ThermalSampleR` package and Shiny application are free and open-source, and we encourage feedback from users via pull requests on the associated GitHub repository. The package is an ongoing project, and future updates will focus on adapting our methods to incorporate a variety of other measures in addition to CTL data.

## AUTHOR CONTRIBUTIONS

**Clarke J. M. van Steenderen:** Methodology; software; validation; formal analysis; data curation; visualization; writing – original draft; writing – review and editing. **Guy F. Sutton:** Conceptualization; methodology; validation; software; formal analysis; visualization; writing – review and editing; writing – original draft; data curation. **Candice A. Owen:** Investigation; conceptualization; resources; writing – original draft; writing – review and editing; project administration. **Grant D. Martin:** Conceptualization; resources; writing – review and editing; supervision; project administration; funding acquisition. **Julie A. Coetzee:** Conceptualization; investigation; writing – review and editing; supervision; project administration; funding acquisition.

## ACKNOWLEDGEMENTS

Sisanda Mvandaba, Cara Trivella, Dan Rogers, Pippa Muskett, Jordan Porter, Tamzin Griffith, Rosali Smith and Khalipha Dambuza all provided their data for inclusion in the CBC CTL database. We would like to thank Grant Duffy (University of Otago, Dundedin, New Zealand) for granting us permission to use the source code for the TOTE function. Part of the funding for this work was provided by the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation (NRF) of South Africa. Any opinion, finding, conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard. Funding was also provided by the Working for Water (WfW) programme of the Department of Environmental Affairs: Natural Resource Management programme (DEA: NRM). Rhodes University is thanked for logistical support.

## FUNDING INFORMATION

Part of the funding for this work was provided by the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation of South Africa. Any opinion, finding, conclusion or recommendation expressed in this

material is that of the authors and the NRF does not accept any liability in this regard.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Github at https://github.com/clarkevansteenderen/ThermalSampleR. The package is also available on the R CRAN repository as "ThermalSampleR".

## ETHICS STATEMENT

The ethical clearance number for the Centre for Biological Control's 2023 projects is 2022-3842-7223.

## ORCID

*Clarke J. M. van Steenderen* https://orcid.org/0000-0002-4219-446X

## REFERENCES

Arnold, B.F., Hogan, D.R., Colford, J.M. & Hubbard, A.E. (2011) Simulation methods to estimate design power: an overview for applied research. *BMC Medical Research Methodology*, 11, 94. Available from: https://doi.org/10.1186/1471-2288-11-94

Bennett, J.M., Calosi, P., Clusella-Trullas, S., Martínez, B., Sunday, J., Algar, A.C. et al. (2018) GlobTherm, a global database on thermal tolerances for aquatic and terrestrial organisms. *Scientific Data*, 5, 180022. Available from: https://doi.org/10.1038/sdata.2018.22

Chen, L. (1995) Testing the mean of skewed distributions. *Journal of the American Statistical Association*, 90, 767–772. Available from: https://doi.org/10.2307/2291090

Chown, S.L., Jumbam, K.R., Sørensen, J.G. & Terblanche, J.S. (2009) Phenotypic variance, plasticity and heritability estimates of critical thermal limits depend on methodological context. *Functional Ecology*, 23, 133–140.

Coetzee, J.A., Byrne, M.J. & Hill, M.P. (2007) Predicting the distribution of *Eccritotarsus catarinensis*, a natural enemy released on water hyacinth in South Africa. *Entomologia Experimentalis et Applicata*, 125, 237–247. https://doi.org/10.1111/j.1570-7458.2007.00622.x

Duffy, G.A., Coetzee, B.W., Janion-Scheepers, C. & Chown, S.L. (2015) Microclimate-based macrophysiology: implications for insects in a warming world. *Current Opinion in Insect Science*, 11, 84–89. Available from: https://doi.org/10.1016/j.cois.2015.09.013

Duffy, G.A., Kuyucu, A.C., Hoskins, J.L., Hay, E.M. & Chown, S.L. (2021) Adequate sample sizes for improved accuracy of thermal trait estimates. *Functional Ecology*, 35, 2647–2662. Available from: https://doi.org/10.1111/1365-2435.13928

Dushoff, J., Kain, M.P. & Bolker, B.M. (2019) I can see clearly now: reinterpreting statistical significance. *Methods in Ecology and Evolution*, 10, 756–759. Available from: https://doi.org/10.1111/2041-210X.13159

Fitts, D.A. (2011) Ethics and animal numbers: informal analyses, uncertain sample sizes, inefficient replications, and type I errors. *Journal of the American Association for Laboratory Animal Science*, 50, 445–453.

Forcino, F.L., Leighton, L.R., Twerdy, P. & Cahill, J.F. (2015) Reexamining sample size requirements for multivariate, abundance-based community research: when resources are limited, the research does not have to be. *PLoS One*, 10, e0128379. Available from: https://doi.org/10.1371/journal.pone.0128379

Gardner, M.J. & Altman, D.G. (1986) Confidence intervals rather than P values: estimation rather than hypothesis testing. *British Medical Journal (Clinical Research Ed.)*, 292, 746–750. Available from: https://doi.org/10.1136/bmj.292.6522.746

Gerrodette, T. (1987) A power analysis for detecting trends. *Ecology*, 68, 1364–1372. Available from: https://doi.org/10.2307/1939220

Halsey, L.G. (2019) The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology Letters*, 15, 20190174. Available from: https://doi.org/10.1098/rsbl.2019.0174

Hampton, J.O., MacKenzie, D.I. & Forsyth, D.M. (2019) How many to sample? Statistical guidelines for monitoring animal welfare outcomes. *PLoS One*, 14, e0211417. Available from: https://doi.org/10.1371/journal.pone.0211417

Janion-Scheepers, C., Phillips, L., Sgrò, C.M., Duffy, G.A., Hallas, R. & Chown, S.L. (2018) Basal resistance enhances warming tolerance of alien over indigenous species across latitude. *Proceedings of the National Academy of Sciences the United States of America*, 115, 145–150. Available from: https://doi.org/10.1073/pnas.1715598115

Kelley, K., Darku, F.B. & Chattopadhyay, B. (2018) Accuracy in parameter estimation for a general class of effect sizes: a sequential approach. *Psychological Methods*, 23, 226–243. Available from: https://doi.org/10.1037/met0000127

Koštál, V., Korbelov, J., Rozsypal, J., Zahradníčková, H., Cimlová, J., Tomčala, A. et al. (2011) Long-term cold acclimation extends survival time at 0°C and modifies the metabolomic profiles of the larvae of the fruit Fly *Drosophila melanogaster*. *PLoS One*, 6, e25025. Available from: https://doi.org/10.1371/journal.pone.0025025

Maxwell, S.E., Kelley, K. & Rausch, J.R. (2008) Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537–563. Available from: https://doi.org/10.1146/annurev.psych.59.103006.093735

Muskett, P.C., Paterson, I.D. & Coetzee, J.A. (2020) Ground-truthing climate-matching predictions in a post-release evaluation. *Biological Control*, 144, 104217. Available from: https://doi.org/10.1016/j.biocontrol.2020.104217

Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, 82, 591–605. Available from: https://doi.org/10.1111/j.1469-185X.2007.00027.x

Neven, L.G. (2000) Physiological responses of insects to heat. *Postharvest Biology and Technology*, 21, 103–111. Available from: https://doi.org/10.1016/S0925-5214(00)00169-1

Nguyen, C., Bahar, M.H., Baker, G. & Andrew, N.R. (2014) Thermal tolerance limits of diamondback moth in ramping and plunging assays. *PLoS One*, 9, e87535. Available from: https://doi.org/10.1371/journal.pone.0087535

Nyamukondiwa, C. & Terblanche, J.S. (2009) Thermal tolerance in adult Mediterranean and Natal fruit flies (*Ceratitis capitata* and *Ceratitis rosa*): effects of age, gender and feeding status. *Journal of Thermal Biology*, 34, 406–414. Available from: https://doi.org/10.1016/j.jtherbio.2009.09.002

Peterman, R.M. (1990) The importance of reporting statistical power: the forest decline and acidic deposition example. *Ecology*, 71, 2024–2027. Available from: https://doi.org/10.2307/1937612

Porter, J.D., Owen, C.A., Compton, S.G. & Coetzee, J.A. (2019) Testing the thermal limits of *Eccritotarsus catarinensis*: a case of thermal plasticity. *Biocontrol Science and Technology*, 29, 565–577. Available from: https://doi.org/10.1080/09583157.2019.1572712

Rezende, E.L., Castañeda, L.E. & Santos, M. (2014) Tolerance landscapes in thermal ecology. *Functional Ecology*, 28, 799–809. Available from: https://doi.org/10.1111/1365-2435.12268

Sinclair, B.J., Coello Alvarado, L.E. & Ferguson, L.V. (2015) An invitation to measure insect cold tolerance: methods, approaches, and workflow. *Journal of Thermal Biology*, 53, 180–197. Available from: https://doi.org/10.1016/j.jtherbio.2015.11.003

Toft, C.A. & Shea, P.J. (1983) Detecting community-wide patterns: estimating power strengthens statistical inference. *The American Naturalist*, 122, 618–625. Available from: https://doi.org/10.1086/284162

Wang, Z., Chen, Y. & Diaz, R. (2019) Thermal tolerance and prediction of northern distribution of the crapemyrtle bark scale (Hemiptera: Eriococcidae). *Environmental Entomology*, 48, 641–648. Available from: https://doi.org/10.1093/ee/nvz019

**How to cite this article:** van Steenderen, C.J.M., Sutton, G.F., Owen, C.A., Martin, G.D. & Coetzee, J.A. (2023) Sample size assessments for thermal physiology studies: An R package and R Shiny application. *Physiological Entomology*, 1–9. Available from: https://doi.org/10.1111/phen.12416