BioOne COMPLETE

A Unified Framework for the Analysis of Germination, Emergence, and Other Time-To-Event Data in Weed science

Authors: Onofri, Andrea, Mesgaran, Mohsen B., and Ritz, Christian

Source: Weed Science, 70(3) : 259-271

Published By: Weed Science Society of America

URL: https://doi.org/10.1017/wsc.2022.8

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Weed Science

www.cambridge.org/wsc

A unified framework for the analysis of germination, emergence, and other time-to-event data in weed science

Andrea Onofri[1], Mohsen B. Mesgaran[2] and Christian Ritz[3]

[1]Associate Professor, Department of Agricultural, Food and Environmental Sciences, University of Perugia, Perugia, Italy; [2]Assistant Professor, Department of Plant Sciences, University of California, Davis, CA, USA and [3]Professor, National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark

Review

Cite this article: Onofri A, Mesgaran MB, Ritz C (2022) A unified framework for the analysis of germination, emergence, and other time-to-event data in weed science. Weed Sci. **70**: 259–271. doi: 10.1017/wsc.2022.8

Received: 10 December 2021
Accepted: 22 January 2022
First published online: 7 February 2022

Associate Editor:
William Vencill, University of Georgia

Keywords:
Censored data; experimental design; kernel density estimators; model fitting; nonparametric models; parametric models

Author for correspondence:
Andrea Onofri, Department of Agricultural, Food and Environmental Sciences, University of Perugia, Borgo XX Giugno 74, 06121, Perugia, Italy. Email: andrea.onofri@unipg.it

Abstract

Germination and emergence assays represent the most notable examples of time-to-event data in agriculture and related disciplines. In spite of the peculiar characteristics of this type of data, there has been little effort to establish a specific and comprehensive framework for their analyses. Indeed, a brief survey of the literature shows that germination and emergence data, along with other phenological measurements such as flowering time, have been analyzed through myriad approaches, giving rise to confusion and uncertainty among scientists and practitioners as to what may represent the best statistical practice. This lack of coherence in statistical approach may reduce the efficiency of research, while making the communication of results and the cross-study comparisons extremely challenging. Here, we attempt to provide a coherent framework and protocol for the analyses of germination/emergence and other time-to-event data in weed science and related disciplines, together with a software implementation in the form of a new R package. We propose a similar approach to biological assays in ecotoxicology, based on: (1) fitting a time-to-event model to describe the whole time course of events; (2) comparing time-to-event curves across experimental treatments, and (3) deriving further information from the fitted model to better focus on some traits of interest. The most appropriate methods to accomplish this procedure were carefully selected from the framework of survival analysis and related sources and were modified to comply with the specific needs of weed, seed, and plant sciences. Finally, they were implemented in the new R package DRCTE. In this article, we describe the procedure and its limitations by way of providing examples of several types of germination/emergence assays. We highlight that our proposed procedure can also serve as the first step of data analyses, with its output subsequently submitted to traditional or meta-analytic approaches.

Introduction

Time-to-event data are very common in agriculture, and they arise in experiments where the key variable of interest is the timing of a certain event, such as seed germination, seedling emergence, or flowering. The common feature of these experiments is that the measurements are taken periodically by counting individuals that have experienced the event. For example, let us consider a simple germination assay with a sample of 30 weed seeds in a petri dish in which the investigator inspects the dish four times on the 5th, 10th, 15th, and 20th days of experiments and counts 10, 15, 2, and 1 new germinations at each assessment time, respectively, with 2 seeds remaining ungerminated at the end of the assay. Given this monitoring scheme, each count should be associated with a time interval during which germinations must have taken place: (0, 5], (5, 10], (10, 15], (15, 20], and (20, ∞). These intervals are usually open to the left, in the sense that, for example, for the 15 germination counts occurring in the second time interval, (5, 10], we only know that the time-to-germination (t) must be $5 < t \leq 10$. Furthermore, for the latest interval, the upper limit remains unknown and is usually indicated as "infinity."

This type of data seems peculiar and marks the difference between time-to-event data and other types of data with which we are familiar, such as the yield, height, or weight of a plant, for which we record real numbers with varying levels of precision. For the time-to-event measurement, we can only record an interval of possible values, and this lack of full information may pertain to all individuals (e.g., seeds) or a subset of them, a phenomenon often referred to as "censoring" (left, interval, and right censoring; see, e.g., Onofri et al. 2010). This phenomenon is typical of time-to-event data, but it can also occur in several other instances (Onofri et al. 2019).

Given the commonality of time-to-event data (e.g., germination and flowering time) in weed research studies and, more generally, agricultural studies, one may expect, reasonably, that there should be a unified and well-established framework for reliable data analyses. Such a framework

© The Author(s), 2022. Published by Cambridge University Press on behalf of the Weed Science Society of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

exists, for example, in medicine, where time-to-event data are, mostly, analyzed by using "survival analysis." In contrast, a brief survey of literature in the agricultural sciences shows that an impressive plethora of general-purpose methods of data analysis have been used or suggested, which is rather confusing to scientists and practitioners. For example, in germination assays, the time course of events has been often described by using single-value indices (see, e.g., Ranal and Santana 2006). Such an approach has the advantage of simplicity: the indices can be derived from the observed data by simple "hand calculations" and then submitted to ANOVA, regression, or other traditional methods of data analysis. However, the reliability of some indices has been questioned (Soltani et al. 2015), and more importantly, a single value is too few to unambiguously describe the whole time course of events (Brown and Mayer 1988b).

Fitting nonlinear regression models, another approach that has been widely used in germination and emergence studies, is intuitively appealing, as it is not subject to most of the ambiguities associated with germination/emergence indices (Brown and Mayer 1988a). However, to fit a nonlinear regression model, we need to make a preliminary transformation of the observed counts into the cumulative percentages of germinated seeds or emerged seedlings. For the previous example, the transformation would result in the following cumulative percentages: 33% germination by day 5, 83% at day 10, 90% at day 15, and 93% at day 20. However, several statistical and conceptual issues can arise when we use these percentages as the response variable in nonlinear regression (e.g., see Gianinetti 2020). From a conceptual point of view, implicit in a nonlinear regression is the assumption that a certain germination percentage is attained at the exact moment when the inspection is made, while this is not generally true. In other words, nonlinear regression ignores the uncertainty associated with of the observed data and assumes that we know the exact timing of the event.

Owing to the aforementioned problems with traditional methods, several proposals have been made that should be better suited to time-to-event data. For example, the use of generalized linear models (GLM) or generalized linear mixed models (GLMM) with binomial errors has been proposed (Gianinetti 2020; Hay et al. 2014), which is appropriate for studying the germination/emergence at a given time point (e.g., the final germinated/emerged proportion), but it is not suitable for analyzing the time course of events, because the distribution of counts across the inspection intervals is not binomial, but is intrinsically multinomial (McCullagh and Nelder 1998: 164–184). Nonetheless, the use of multinomial models with censored data, although possible, may be inefficient, as it does not recognize that the underlying classification variable (time) is continuous (Onofri et al. 2019). Furthermore, multinomial models cannot be used when the time intervals overlap.

Recently, the use of "survival analysis" for time-to-event data has been suggested, more or less explicitly, in agricultural literature (Hay et al. 2014; Humplik et al. 2020; Manso et al. 2013; McNair et al. 2012; Onofri et al. 2011, 2018; Ritz et al. 2010; Romano and Stevanato 2020; Scott et al. 1984). In principle, this is a sound suggestion, but in practice, the framework of survival analysis has been developed to address very different needs and does not appear to be immediately transferable to the specific needs of, for example, seed germination/emergence, as outlined below:

1. The term "survival analysis" makes little sense in, for example, germination and emergence studies, because we are not dealing with a survival process.
2. Survival models consider the cumulative probability of avoiding the event (death), which is a decreasing function of time (e.g., see Romano and Stevanato 2020). This is rather counterintuitive in agriculture, where we are mainly interested in the cumulative probability of observing the event, which is an increasing function of time (see Figure 1).
3. In survival analysis, it is most often assumed that all individuals will experience the event at some future time, while in agriculture, we know that, very often, some individuals will not experience the event, for example, seeds do not germinate due to dormancy or some plants may not reach flowering stage because of predation or natural death. Although the so-called cure model has been proposed in the medical sciences to account for the fraction that will not experience the death event following a certain disease (Lambert et al. 2007), this modeling approach is not always included in most widespread computer software for survival analysis.
4. The concept of hazard rate is central to survival analysis and forms the basis of several modeling approaches, for example, for semiparametric and random effects (frailty) models. Although such a concept has already been introduced in agriculture (Ritz et al. 2010), its real usefulness has yet to be elucidated.
5. The methods used in survival analysis are not necessarily correct in agriculture. For example, the Kaplan-Meyer estimators, the log-rank test, and Cox regression, in their most widespread formulations, are not tailored to deal with interval censoring (Kleinbaum and Klein 2005), and therefore, their usage with time-to-event data in agriculture may not necessarily be warranted.

In the wake of the earlier criticism against the traditional methods of data analysis for time-to-event assays, several relatively new techniques have been proposed, but the information on these new methods is still scattered, and no clear and consistent guidelines exist for their implementation. This lack of coherence is likely to confuse scientists and practitioners in selecting the most appropriate method, cast doubt on the reliability of results, and reduce the efficiency of communication, while making cross-study comparisons a challenging task.

To fill such gaps, we provide here a coherent framework and protocol for the analyses of time-to-event data in weed and plant science. To this end, we (1) preliminarily reviewed the unique aspects of time-to-event experiments in agriculture; (2) defined a step-by-step procedure for data analysis; (3) reviewed the existing literature and selected a body of suitable methods to accomplish this procedure; and (4) implemented those methods in a new software facility, that is, an R package. In this article, we demonstrate the whole framework by giving several "real-life" examples, mainly referring to germination and emergence assays, which are the most widespread types of time-to-event assays in agriculture. Nonetheless, the techniques we propose will also be useful for other time-to-event data and, in general, for all types of censored data. For the sake of readability, we will refrain from giving much statistical detail and will refer readers to the available literature for such detail.

Data Analyses: Preliminary Issues

Aims of Time-to-Event Studies

The main aims of time-to-event studies are to estimate and compare the time to an event of interest for a group of individuals (e.g.,

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

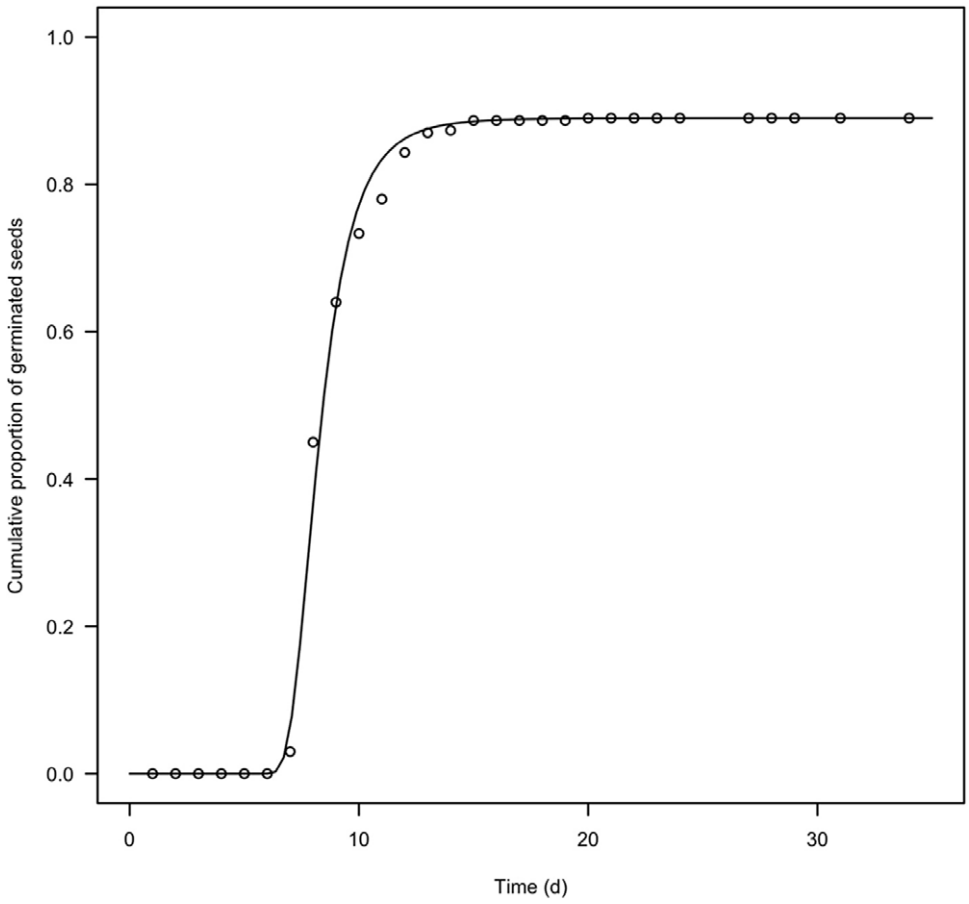


Figure 1. Parametric time-to-event curve for a germination assay with alfalfa (Example 1). Symbols show the observed data, and the solid line represents the maximum likelihood fit, according to a Weibull cumulative distribution function (CDF).

seeds or seedlings) that might have been submitted to different environmental conditions, experimental treatments, or agronomic practices. Apart from the time to event for each individual, other population-based information is also sought, such as the average speed, the variability (or uniformity) across individuals, the prevalence (proportion of individuals that experience the event within a certain time), or the incidence (the number of individuals that experience the event in a certain time instant, relative to the total number of individuals that have already experienced the event before that instant).

In this review, we will focus on the most appropriate techniques to achieve these aims, implying that there is the fundamental interest in the whole time course of events. However, we will not consider studies that have simply assessed the proportion of individuals (e.g., seeds or seedlings) that experienced the event at some specific time point (e.g., total germination/emergence at the end of an assay); these studies are usually based on a single inspection date and, according to the abovementioned definition, cannot to be regarded as time-to-event studies.

Experimental Design

The observational units for time-to-event experiments can be seeds, seedlings, or plants at a specific developmental stage. These units are inspected several times, until they commit to the event and, subsequently, are removed from the study or no longer followed.

A unique feature of most experiments in agriculture is that the observational units can be grouped within randomization units, such as petri dishes, containers, boxes, or field plots, to which the experimental treatments are allocated. Therefore, those randomization units represent the true replicates, which are repeatedly inspected over time (repeated-measures design).

The allocation of treatments to the randomization units can be performed according to several types of experimental designs. Very frequently, the design is simple: for example, we may have independently treated 24 petri dishes with three replicates of eight increasing concentrations of polyethylene glycol (PEG) and then distributed these dishes randomly between the shelves of a growth chamber. In other cases, we may consider each shelf within a chamber as a different block, so that the experimental design is in randomized complete blocks.

More complex designs are also possible. For example, a germination assay may be designed by using two chambers set at two different temperatures and seven petri dishes per chamber, with seven different water potential levels. If the assay is repeated four times, the experiment includes the following blocking factors: the two replicated runs (as blocks), the two chambers in each run (as plots, to which temperatures are randomly allocated), and the seven petri dishes in each chamber (as subplots, to which water potential levels are randomly allocated), so that the result is a split-plot design.

Complex designs are also common with experiments aimed at assessing the time to flowering or other relatively late phenological

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

stages, which are usually conducted in the field; multi-tier experiments (Brien and Bailey 2006) may also be common, for example, with seedbank experiments, in which seed or other material is collected in the field during an experiment and later submitted to further treatments in the lab and assayed.

It is worth emphasizing that the experimental design should be appropriately considered in the data analysis, as the observations may come in groups, so that the independence of residuals is no longer valid, which may invalidate inferences based on methods that make such an assumption.

Data Structures

As mentioned earlier, the observed data are the counts of individuals that have experienced the event before the inspection time. For each count, we need to record at least two values: the date at which the count was made (usually as days after the beginning of the assay) and the previous inspection date, which marks the beginning of the interval during which the event(s) must have happened. Other variables also may need to be recorded for each count, such as the experimental treatment and the randomization unit (e.g., the dish/pot/plot).

To record this time-to-event data, one may adhere to the principles of "tidy data" (Wickham and Grolemund 2016), that is: (1) each variable forms a column, (2) each observation forms a row, and (3) each type of observational units forms a table. For time-to-event assays, it is appropriate to use each count to form a row and the columns to store the information about each count (long grouped format). Table 1 shows a simple example of long grouped data, relating to a germination assay with periodical inspections on a sample of 20 seeds. For right-censored observations, in which the end of the interval is unknown, either "infinity" or "not available" may be used ("Inf" or "NA," in the R language, respectively).

Variations to this basic scheme are possible. In survival analysis, the usual data structure is slightly different, in that each individual patient forms a row, and the interval limits for each patient ("Start" and "End" in Table 1) are stored in two columns, while the column of counts is no longer necessary (long ungrouped format). This is the basic data structure for survival analysis, and it could also be used for time-to-event data in agriculture.

A third scheme is often used for data entry in, for example, germination assays, in which each petri dish forms a row, and each interval forms a column (wide grouped format; Table 2). This storage scheme may be convenient for swift calculations with a spreadsheet (e.g., to calculate the mean germination time), but it should be avoided, as it does not adhere to the principles of tidy data, and it is only feasible when the monitoring schedule is homogeneous for all the rows.

Considering these points, for the definition of our proposed framework, we will assume that the data are prepared according to one of the long formats given, while data in other formats are to be transformed into a long format before analyses.

Definition of a Data Analysis Framework

For analysis of all types of time-to-event data, we suggest the following step-by-step procedure: (1) fit a model to describe the whole time course of events; (2) compare the time course of events across experimental treatments; (3) if necessary, derive information from the fitted model to better focus on some biological traits of interest; and (4) submit the derived information to other types of statistical analyses (second-step analyses). Conceptually, such an approach is not new; it has been commonly used in ecotoxicology for biological assays (Ritz 2010), and it has been put forward also for germination assays (Ritz et al. 2019). However, we emphasize that the procedure must be accomplished by using methods that comply with the specific requirements of time-to-event assays. In particular, the selected methods should adhere to the following mandatory requirements:

1. capability of accounting for interval censoring, which is a very common trait of time-to-event data in agriculture;
2. capability of accounting for a final fraction of individuals that are not going to experience the event ("cured" fraction);
3. high degree of flexibility to describe the time course of events in all possible experimental settings, without making unrealistic assumptions; and
4. capability of accounting for the experimental design, particularly for field experiments.

We will describe the selected methods in the following sections. However, we note that a framework can only be useful if it is made readily available in a user-friendly freeware software. In this respect, we implemented our method as package in the R statistical environment (R Core Team 2021), due to its freeware/open-source nature and the possibility of building publicly available add-on packages. Our package was built upon the backbone of the DRC package, which represents the most notable effort in creating a unified framework for dose–response curve analysis in agriculture (Ritz et al. 2019). The package already includes some relevant time-to-event methods (Ritz et al. 2013) and has become very widespread among agricultural scientists, so that many people are familiar with its syntax. We worked to enhance the functionalities of DRC in two directions: (1) increase the number of available time-to-event methods and (2) optimize the whole structure and terminology for time-to-event studies.

This work resulted in the DRCTE add-on package, which derives from and depends on the DRC package, using the same syntax and approaches. Relevant time-to-event functions of DRC were imported into DRCTE, following some adjustments to comply with

Table 1. Example of long grouped structure for a germination assay.[a]

Petri dish	Start	End	Counts
1	0	1	1
1	2	3	5
1	3	4	2
1	4	Inf	4
2	0	1	2
2	2	3	3
2	3	4	6
2	4	Inf	9

[a]"End" represents the moment when the inspection was made (e.g., as days from the beginning of the assay), "Start" represents the moment of the previous inspection, "Counts" represents the number of seeds that germinated within the interval from Start to End. Inf: infinity.

Table 2. Example of wide grouped structure for the data set in Table 1.[a]

Petri dish	1	3	4	Inf
1	1	5	2	4
2	2	3	6	9

[a]The numeric column names represent the moments when the inspections were made, while the final column represents the final count of ungerminated seeds. Inf: infinity.

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

the conceptual and methodological peculiarities of time-to-event experiments. Furthermore, we borrowed several useful functions from the BINNEDNP (Barreiro-Ures et al. 2019), SURVIVAL (Therneau 2021), and INTERVAL (Fay and Shaw 2010) packages and reimplemented these functions to make their syntax similar to that used in the DRC package. Eventually, we included all these reworked functions in the DRCTE package.

The DRCTE package is publicly available on GitHub (Onofri 2021), while the code to run all the examples in this paper is provided as Supplementary Material.

Methods and Examples

Describing the Time Course of Events

The individual times to event may be highly variable within a population, and we can describe such variability by using a cumulative distribution function (CDF), which depicts the probability P that the time-to-event T is equal to or lower than any given time:

$$G(t) = \mathrm{P}(T \leq t) \quad [1]$$

where $G(t)$ corresponds to the proportion of individuals that experienced the event until time t; within the framework of survival analyses, $G(t)$ is referred to as the "time-to-event curve" and, analogously, for germination/emergence assays, we may call it "time-to-germination/emergence curve."

Relating to the shape of the time-to-event curve $G(t)$, we distinguish: (1) parametric and (2) nonparametric models; the selection between the two is mainly based on the assumptions we are willing to make about the observed data.

Parametric Time-to-Event Models

In many cases, we can reasonably assume that the time-to-event curve within a population can be described by a parametric model Φ, based on a few meaningful parameters:

$$G(t) = d\ \Phi(t, e, b) \quad [2]$$

where e is the location parameter, b is the shape parameter, and d represents the maximum fraction of individuals that experience the event ($0 < d \leq 1$), so that $1 - d$ is the so-called cured fraction that has not experienced the event by the end of experiment (e.g., ungerminated seeds). The term Φ represents any valid CDF, such as lognormal, log-logistic, or Weibull.

To fit the time-to-event curve and fully respect the nature of time-to-event data, we use Equation 2 to build a likelihood, based on the observed counts and related intervals (see, e.g., Onofri et al. 2011); the likelihood function is then maximized to estimate the value of d, b, and e (maximum likelihood estimation [MLE]). As an example, a time-to-event model referring to the data in Table 1 would be specified in R language as:

$$\texttt{Counts} \sim \texttt{Start} + \texttt{End}$$

We see that the fitting process uses the observed counts with no need of any preliminary transformation. Furthermore, for each count, we consider the whole uncertainty interval during which the event(s) must have taken place. In other words, we are using just as much information as is contained in the data, which results in more reliable estimates of parameters and related variances (Onofri et al. 2014; Ritz et al 2013).

In our framework, parametric time-to-event models can be fit using the *drmte()* function in the DRCTE package, which works similarly to the *drm()* function in the DRC package. Assuming that the data (the variables: 'count', 'start', and 'end') are stored in a 'data frame' named 'dataset' in one of the long formats, the R codes to fit a time-to-event model follow:

```
mod <- drmte(count ~ start + end, data = dataset,
             fct = selectedCDF())
```

where the argument 'fct' can be assigned to one of the available R functions (Ritz 2010), such as: *LL.3()/LL.2()* (log-logistic CDF), *LN.3()/LN.2()* (log-normal CDF), *W1.3/W1.2* (Weibull type 1 CDF), and *W2.3/W2.2* (Weibull type 2 CDF), as provided in DRC. We also provide the *loglogistic(), lognormal(), weibull1(),* and *weibull2()* functions, where the sign of the shape parameter b is reversed and constrained to be positive for increasing curves, which is conceptually more appropriate to CDFs. Those who are experienced with the *drm()* function will note that the 'type = "event"' argument used in *drm()* is not necessary in *drmte()*, as this latter function is specifically designed for time-to-event methods.

Example 1

The germination of alfalfa (*Medicago sativa* L.) was tested at 3 C, by using three replicated petri dishes with 100 seeds each. Germinated seeds were periodically counted and removed from the dishes during a period of 34 d (A Onofri and P Benincasa, unpublished data). A parametric time-to-event model (Weibull CDF, type 1) was fit to the observed counts of germinated seeds; the equation is:

$$G(t) = d\ \exp\{-\exp\{-b[\log(t) - \log(e)]\}\} \quad [3]$$

where e is the abscissa of the inflection point (location parameter), b is the slope at inflection point, and d is the maximum proportion of germinated seeds (upper asymptote). The maximum-likelihood estimates of model parameters were $b = 8.20$, $d = 0.89$, and $e = 7.88$ d (we will discuss SEs later); the time-to-event curve is shown in Figure 1.

Nonparametric Time-to-Event Models

Parametric models make some stringent assumptions on the shape of the time-to-event curve that may not be realistic in some cases. For instance, in emergence studies, events may come in successive flushes, perhaps, due to the existence of several subpopulations with different dormancy cycles. Therefore, the time course of emergences assumes the form of a staircase, which cannot be described properly through parametric models. In other instances, there might be a uniform and abrupt flux of events in a very short time or, conversely, events might be very few and sporadic, so that the iterative estimation process does not converge to reasonable estimates.

In all these (and other) situations, we may prefer (or be forced) to fit a time-to-event model that makes fewer a priori assumptions on the shape of the time-to-event curve, accommodating also unusual (stepwise) shapes, as dictated by the observed data. The most widespread solution in survival analysis for interval-censored data is to use an expectation-maximization algorithm to find the nonparametric maximum likelihood estimator (NPMLE) of the time-to-event curve (Fay and Shaw 2010); we suggest that this same approach to be used for time-to-event data in agriculture.

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

The NPMLE is based on a set of m unique and non-overlapping time intervals (Turnbull's intervals; Fay and Shaw 2010), denoted by the $I_j = (l_j, r_j)$, where j varies from 1 to m with l and r representing, respectively, the left and right interval margins. Each time interval is associated to a probability mass value $w_j = P(l_j < t \leq r_j)$ corresponding to the probability that the events happen in the jth interval. The determination of Turnbull's intervals and related masses is obvious with non-overlapping intervals, while it requires iterative calculations when intervals overlap. The resulting time-to-event curve has a staircase shape and is univocally defined by the Turnbull's intervals and related masses, which can be conceptually regarded as model parameters:

$$G(t) = \begin{cases} 0, & \text{if } t = l_1 \\ \sum^{w_j}, & \text{if } t = r_j, \textit{with } 2 \leq j \leq m-1 \\ 1, & \text{if } t = r_m \end{cases} \quad [4]$$

The important characteristic of such a nonparametric approach is that the time-to-event curve is only defined at the end of each interval, while it is undefined elsewhere; therefore, it is often represented by a shaded area (Figure 2). As will be shown, in some instances, it may be convenient to assume that the events are evenly scattered within each interval, which appears to be a biologically reasonable assumption (Fay and Shaw 2010).

Nonparametric time-to-event models can be fit in DRCTE by simply setting 'fct = NPMLE()' in the *drmte()* function.

Example 2

A pot assay was performed to evaluate the time to emergence for 50 seeds of prickly lettuce (*Lactuca serriola* L.) Emerged seedlings were counted and removed every second day from the beginning of the assay, for 32 d. The assay was repeated with a slightly different monitoring schedule, that is, the number of emerged seedlings was counted 1 d after the beginning of the assay and, afterward, every second day for 33 d (AO, unpublished data).

The observed pattern of emergence indicates two distinct flushes around the 4th and the 24th day after the beginning of the assay. Therefore, we fit an NPMLE of the time-to-event curve; the results show that the overlapping intervals were merged into one set of Turnbull's intervals, and the time-to-event curve follows closely the observed emergence pattern, with no predefined shape (Figure 2). The gray boxes indicate our uncertainty, as we do not know the exact moments that the germination took place within each interval.

This nonparametric approach is very flexible and can be used to describe any time-to-event curve shapes. However, in some instances, such as making a prediction, we might be reluctant to accept a "broken-stick" model to describe a continuous phenomenon and prefer a smooth curve. Another nonparametric option was proposed by Barreiro-Ures et al. (2019), which uses a kernel density estimator (KDE) to provide a smooth CDF with no predefined shape:

$$G(t) = \sum_{j=1}^{m} w_j \mathrm{K}\left(\frac{t - I_j}{h}\right) \quad [5]$$

where $I_j = (l_j + r_j)/2$ is the center of each interval, K is a gaussian kernel function, and h is the so-called bandwidth, that is, the parameter controlling the amount of smoothing, or how closely the CDF should follow the observed data. The selection of the bandwidth is crucial: a small bandwidth value will give to a very close fit, but the resulting shape may be highly unrealistic (overfitting). Alternatively, a large bandwidth will result in a poor fit (underfitting). To optimize the selection of the bandwidth, Barreiro-Ures et al. (2019) proposed the asymptotic mean integrated squared error (AMISE) and the bootstrap method: both methods have been implemented in the DRCTE package, although the first one is less computer intensive and is used as the default method.

KDEs can be fit with 'drcte()' by setting 'fct = KDE()' for AMISE method or 'fct = KDE(bw = "boot")' for the bootstrap method in the abovementioned R code. For Example 2, the AMISE bandwidth was equal to 1.767, while the bootstrap bandwidth was 0.9031; both the estimated curves appear to provide better fits to the observed data than a parametric (log-logistic) curve (Figure 3).

It should be noted that there is no particular preference between parametric and nonparametric methods in our framework, and the selection is left to the user, according to the aims of the experiment. In general, parametric methods are well suited to make inferences about underlying biological mechanisms of the process, while nonparametric methods may give the best fit to the observed data (Cao et al. 2013; Gonzalez-Andujar et al. 2016) and might be preferred for some emergence studies or, more generally, whenever the time course of events cannot be adequately described by a parametric time-to-event model.

Modeling the Effect of Experimental Factors

The effect of treatment factors can be accounted for by coding a different time-to-event curve for each factor level. We may be interested in testing whether, overall, those curves are not significantly different from one another, which would imply that there is no treatment effect. Statistical tests of significance are based on the selection of an appropriate test statistic and depend on the model type (parametric, NPMLE, or KDE). Based on our experience and the literature, we suggest, at least as the default choice, a likelihood ratio statistic for parametric models (Bolker 2008), a Wilcoxon-type statistic for NPMLEs (Gomez et al. 2009), and a Cramér-von Mises–type distance statistic for KDEs (Barreiro-Ures et al. 2019). It is beyond the scope of this paper to give details about these methods: the three statistics are conceptually very different, but they share the property that they are close to 0 when the time-to-event curves are very similar across the treatment levels. Conversely, the values of these statistics progressively increase as the difference between the curves increases.

Example 3

A germination assay was performed to compare three species belonging to the *Verbascum* genus (data are taken from Catara et al. 2016). For each species, four replicated petri dishes with 25 seeds were inspected daily for 15 d and the germinated seeds were counted and removed at each inspection date. For the sake of this exercise, we used this data set to fit an NPMLE of the time-to-event curve, although a parametric time-to-event model would be appropriate as well.

The time-to-event curves were rather different from one another (Figure 4) and the sums of the Wilcoxon scores for the three groups are, respectively, −58.3 for in *Verbascum arcturus* L., 7.95 for moth mullein (*Verbascum blattaria* L.), and 50.36 for Cretan mullein (*Verbascum creticum* Cav.), confirming that the germination of the first species was, on average, the slowest (lowest score sum), while the germination of the third species

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

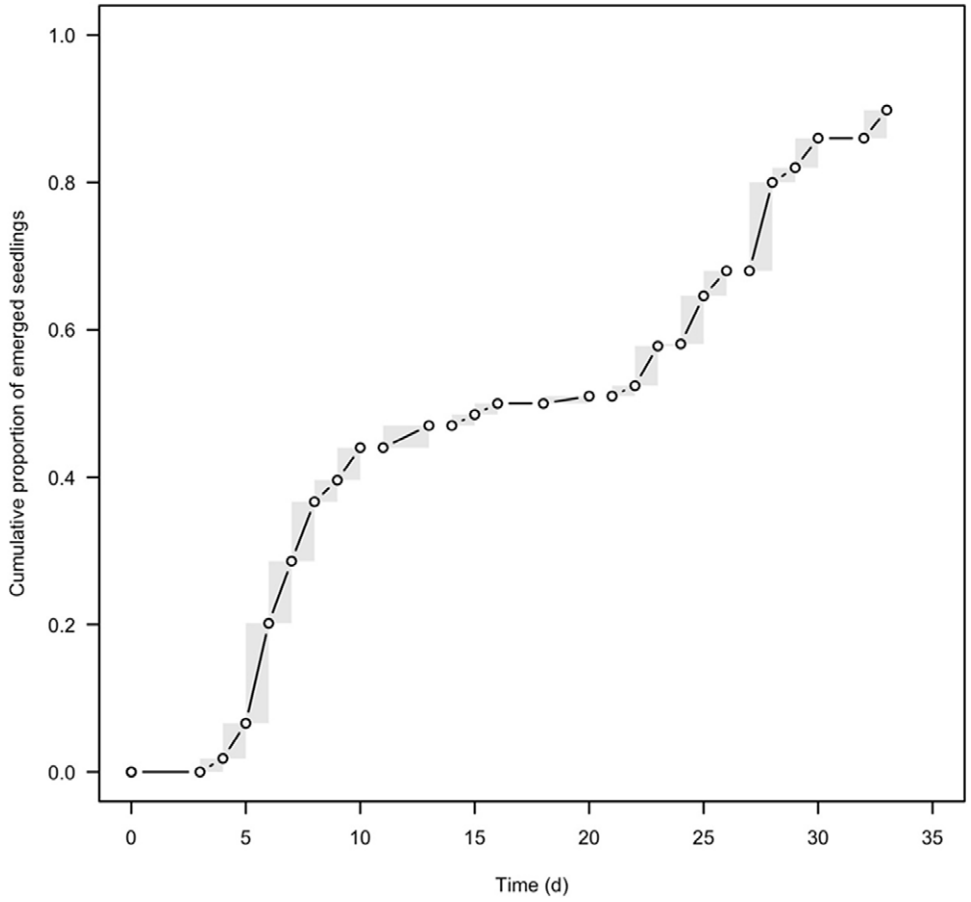


Figure 2. Nonparametric time-to-event curve (nonparametric maximum likelihood estimator [NPMLE]) for a seedling emergence assay with two marked emergence flushes (Example 2). Symbols show the NPMLE for the time to event at the end of Turnbull's intervals, while the gray areas represent the uncertainty due to censoring.

was the fastest (largest score sum). The overall Wilcoxon score is equal to 58.99, which is a rather high value, supporting the idea that the time-to-event curves are different from one another.

In order to calculate a P-value for the null hypothesis that the three curves are not significantly different, we suggest a permutation approach. The basic idea is that, if the null hypothesis is true, the labels for the groups can be freely permuted among observations. Accordingly, we will:

1. randomly permute the labels for the groups across seeds, in order to obtain a new set of observations for each group, while keeping the same size as the original groups;
2. calculate the statistic of interest (in this example the Wilcoxon statistics);
3. repeat this process a large number of times (e.g., 999), so that we can build a permutation distribution for the statistics of interest;
4. compare the observed value of statistics (in this case, 58.99) against the permutation distribution and calculate the proportion of values that are as large as or larger than the observed value. This proportion can be regarded as the P-value for the null hypothesis.

The above process seems intuitive, but there is one issue that we should take into consideration, that is, the observations (the seeds, in this example) are usually clustered within randomization units (the petri dishes, in this example) and, therefore, permuting the labels for the individual observations belonging to the same petri dish is meaningless, as it ignores the original experimental design. It is, thus, more appropriate to permute the group labels across randomization units and not across observations. In this example (Example 3), there are 12 dishes, to which we can randomly allocate the 12 group labels to create new permutation samples (Winkler et al. 2015). The resulting P-value is 0.005; accordingly, the null hypothesis is rejected, and we can conclude that the time courses of germination differ significantly among species.

Dealing with treatment factors, the *drmte()* function in the DRCTE package behaves very much the same as the *drm* function in the DRC package; the argument 'curveid' is used to specify the treatment variable. The *compCDF()* function can be used to compare the different time-to-event curves across treatment levels. The most appropriate statistic is automatically selected, depending on the fitted model, although the user can override the default choice, if necessary. A P-level is calculated by using the permutation approach, and if a variable coding for randomization units is provided by way of the 'units' argument, group labels are permuted across groups and not across individuals.

Modeling the Effects of Quantitative Variables

The progress to relevant events in plant/animal life is usually affected by several environmental covariates (e.g., water, temperature, oxygen availability, and other continuous variables), which we might want to incorporate in the time-to-event function. For example, the effects of water availability and temperature on seed

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

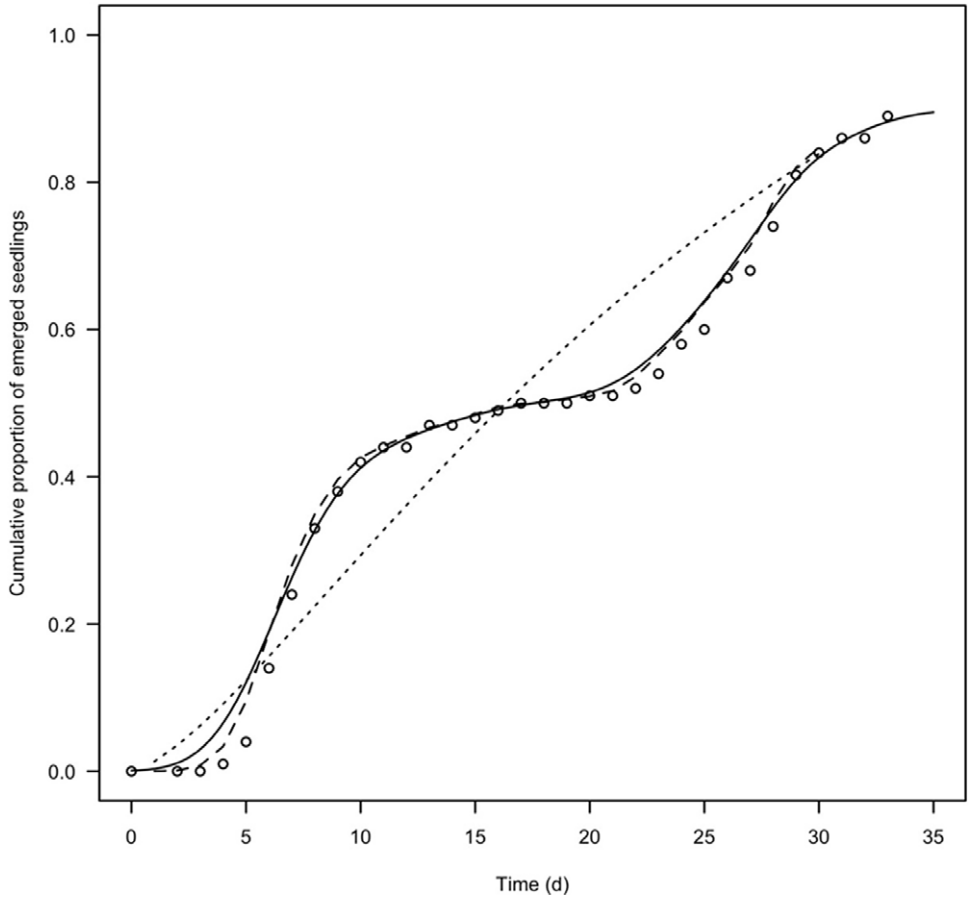


Figure 3. Nonparametric time-to-event curves (kernel density estimator [KDE]) for the same assay as in Figure 2 (Example 2). Symbols show the observed data, the solid line shows the KDE for the cumulative probability with bandwidth = 1.767 (AMISE method), while the segmented line shows the KDE with a bandwidth of 0.9031 (bootstrap method). The dotted line shows a log-logistic (parametric) fit.

germination/emergence have been described by using the so-called hydrotime, thermal time, and hydrothermal time models (Bradford 1995, 2002).

All such models can be easily cast as parametric time-to-event models by defining a CDF for germination time. This CDF should have the following characteristics: (1) it should contain the external covariates together with time as the predictors; (2) it should be defined for time values between 0 and infinity; and (3) the response G should be bound between 0 and 1. Hydro-(thermal)-time-to-event models were proposed in Onofri et al. (2018), with their respective instructions to implement them in R. Other models had been previously proposed by Mesgaran et al. (2013, 2017) as nonlinear regression models, which nonetheless depict the cumulative proportion of germination times within the population, and therefore, they can be easily cast into the time-to-event platform.

The DRCTE package includes functions to fit most of the hydro-thermal time models proposed insofar. These models can be passed through the 'fct' argument, while the names of the data columns corresponding to these covariates need to be added to the model definition after the names of the columns coding for the time interval margins.

Example 4

Here we examine a germination assay with rapeseed (*Brassica napus* L. var. *oleifera* 'Excalibur') that tested the effects of 13 water potentials (−0.03, −0.15, −0.3, −0.4, −0.5, −0.6, −0.7, −0.8, −0.9, −1, −1.1, −1.2, −1.5 MPa) created by using a polyethylene glycol solution (PEG 6000). For each water potential level, three replicated petri dishes with 50 seeds were incubated at 20 C, and germinated seeds were counted and removed every 2 to 3 d for 14 d (Pace et al. 2012).

For this data set, we fit the following hydrotime model (adopted from Mesgaran et al. [2013] and slightly reparametrized):

$$G(t,\Psi) = \frac{1}{1+\exp\left\{\frac{\log\left[\Psi-\frac{\theta_H}{t}+\delta\right]-\log\left(\Psi_{b(50)}+\delta\right)}{\sigma}\right\}} \quad [6]$$

In Equation 6, the predictors are t (time, day) and Ψ (water potential, MPa), and the parameters to be estimated are the hydrotime constant θ_H, the median base water potential $\Psi_{b(50)}$, the threshold parameter δ (Mesgaran et al. 2017), and the scale parameter σ.

This function can be seen as the CDF of event times within the population and hence can be fit within the time-to-event framework by using the observed counts of germinated seeds as the response and the corresponding time interval extremes as the predictors (the two variables 'timeBef' and 'timeAf'), together with water potential in the substrate. We used the *drmte* function in the DRCTE package, together with the *HTLL()* function, which corresponds to Equation 6. The code to fit the model is given in the

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

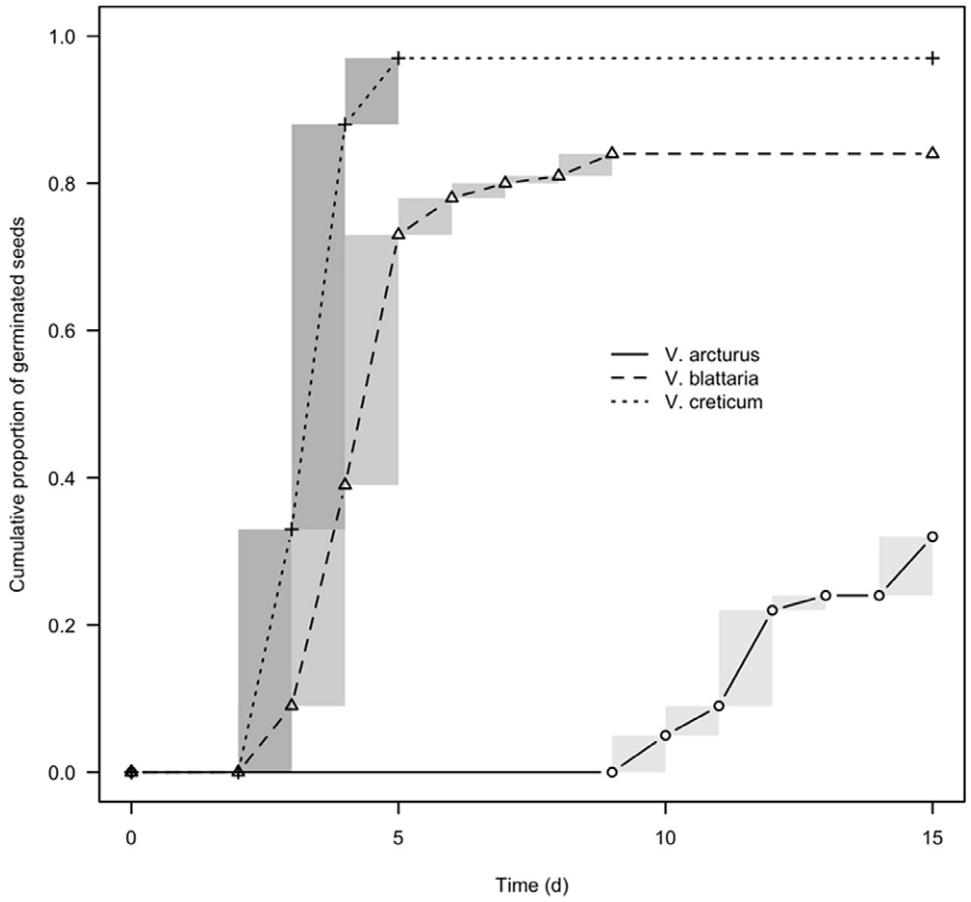


Figure 4. Nonparametric time-to-event curves for a germination assay with three species of the genus Verbascum (Example 3). The gray areas represent the uncertainty due to censoring.

Supplementary Material and parameter estimates were: $\theta_H = 0.677$ MPa d, $\Psi_{b(50)} = -0.948$ MPa, $\delta = 1.137$ MPa, and $\sigma = 0.372$. The observed data and fitted values are shown in Figure 5. It is worth noting that the time-to-event curve is a function of a continuous environmental variable (water potential), and predictions can be also obtained for water potential levels that were not included in the original experiment.

Extracting Information from a Time-to-Event Curve

Determining the whole time-to-event curve gives us a full and unambiguous description of the progress to germination, emergence, or other relevant events, and for this reason, we suggest that this is the first step of data analysis for all types of time-to-event assays. We have also shown that time-to-event curves can be compared to assess whether the overall time course of events is affected by the experimental treatments.

If necessary, the time-to-event curves can be also used to derive further information to better focus on some specific traits of the biological process under investigation.

Model Parameters and SEs

For a parametric time-to-event curve, model parameters (e.g., b, d, and e in Equation 3) are not only useful to fully specify the curve itself, but they also imbed biological meanings, for example, to describe the prevalence (d), the speed (e), and the uniformity (b) of event times within the population of interest. Obviously, these parameters should be reported together with a measure of variability, for example, SE.

As MLE methods are used, "naive" SEs can be obtained from the second derivative of the likelihood function, as implemented in most statistical software. Unfortunately, these naive SEs are only valid when the individual observations (seeds/seedlings/plants) are totally independent of one another. To account for the experimental design and obtain cluster-robust SEs, a possible approach is to use the so-called sandwich estimator (Carroll et al. 1998), which has been shown to be reliable for clustered survival data (Yu and Peng 2008).

In the DRCTE package, as well as in several other packages in R, parameter estimates and SEs are obtained from the model object, by using the 'summary()' method. In DRCTE objects, the 'units' argument can be used to provide variable coding for the randomization units to derive cluster-robust SEs by way of the facilities provided in the SANDWICH package (Zeileis et al. 2020).

Example 5

In a pot emergence assay, seeds of prickly lettuce (*Lactuca serriola* L.) were put in 6 replicated trays filled with soil (50 seeds per tray, at a depth of 10 mm), and emerged seedlings were counted daily for 15 d and removed from the trays. The value of estimated parameters from a lognormal time-to-event model are shown in Table 3. As shown, there is a marked difference between naive and

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

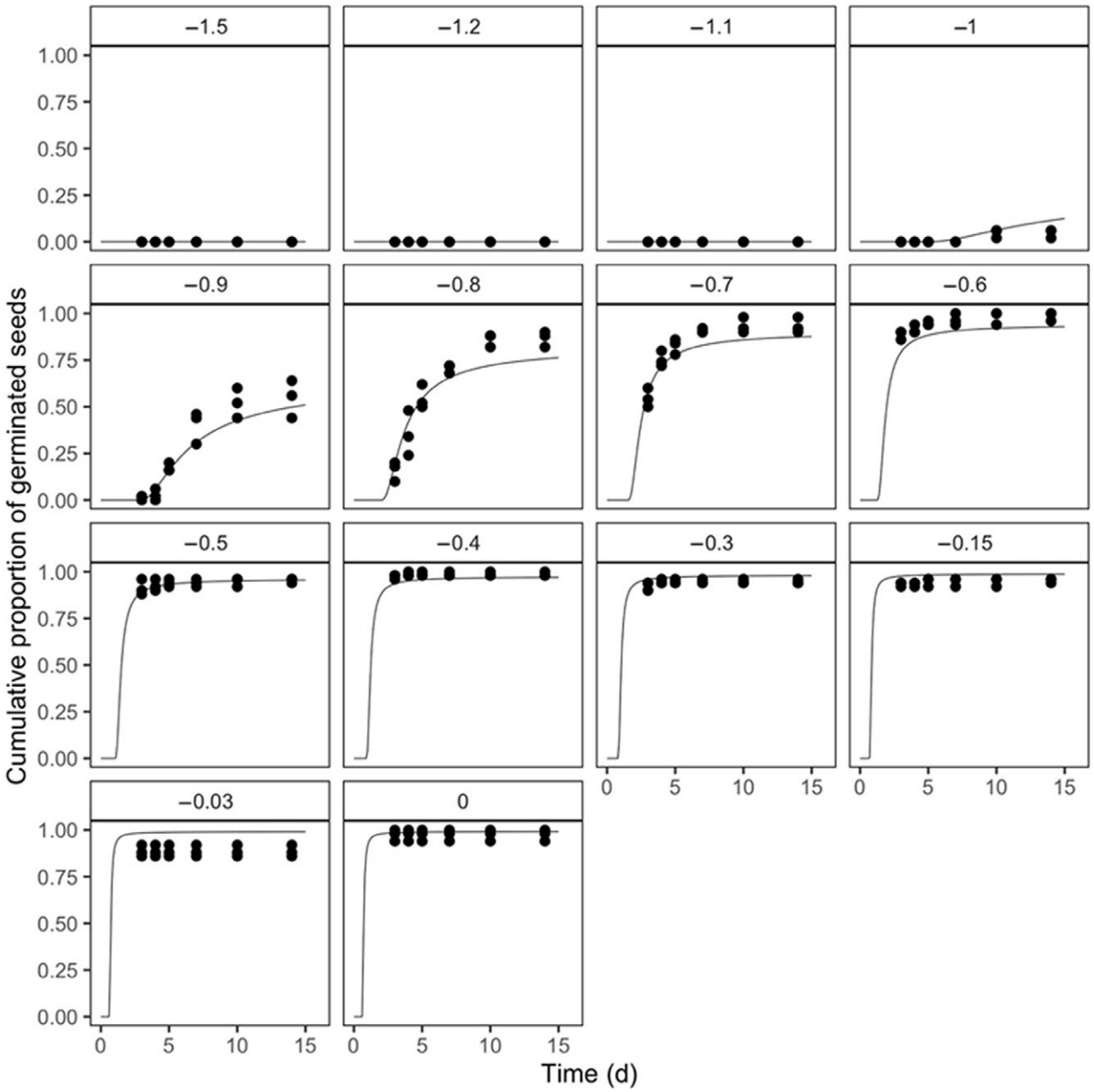


Figure 5. Hydro-time-to-event model (Mesgaran et al. 2013) fit to a germination assay with rapeseed (Brassica napus L. var. oleifera 'Excalibur'), at different water potential levels (MPa) in the substrate (Example 4). Symbols show the observed data, and dotted lines show the fitted values.

sandwich SEs, in particular, for the location parameter *e*, which defines the median emergence time for the emerged fraction.

Predictions

The time-to-event curve (either parametric or nonparametric) can be used to predict the cumulative proportion of individuals that experienced the event at any given time *t*. For parametric time-to-event curves and KDEs, predictions are straightforward, as the curve is continuous for any time level from 0 to infinity. For NPMLEs, we need to take into account that the time-to-event curve is not specified within each time interval, as shown in Figure 2. Thus, we have to make a reasonable assumption as to how the cumulative probability of events will increase within the time from the left to the right of interval margin. A reasonable choice is to assume that the probability increases progressively within each interval ("interpolation" method; Figure 2); however, it is also very common in literature to assume that the probability increases abruptly either at the beginning of the interval ("left-side" prediction method) or at the middle of the interval ("midpoint" prediction method; see Therneau 2021) or at the end of the interval ("right-side" prediction method).

As a convention, predictions need to be reported with a measure of variability. For parametric models, SEs for predictions can be derived by the delta method (Efron 1981), along with a cluster-robust sandwich estimator for the variance–covariance matrix (Zeileis et al. 2020), if this is required by the experimental design. For nonparametric models, we recommend the bootstrap method,

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Table 3. Estimates for model parameters relating to a lognormal time-to-event model, for an assay with Lactuca serriola (Example 5).

Parameter	Estimate	Naive SE	Sandwich SE
b	4.088	0.199	0.301
d	0.837	0.021	0.019
e	4.161	0.067	0.252

and the possible constraints to randomization imposed by the experimental design can be accounted for by resampling randomization units rather than individuals (Goldstein 1999).

In the DRCTE package, predictions are obtained from the model object by using the 'predict()' method and by providing a set of times when predictions need to be obtained. Similar to the 'summary()' method, the 'units' argument can be used to provide a variable coding for the randomization units, which is used to derive cluster robust SEs, according to the type of fitted model. For NPMLEs, the 'type' argument can be used to select the prediction method (interpolation is the default choice).

Example 3 Continued

For the data set on the three species of the genus *Verbascum*, it may be relevant to ask what cumulative proportion of emergence can be expected on the 5th and 10th days. The predictions are easily obtained in the DRCTE package, along with cluster-robust SEs, using the codes reported in the Supplementary Material (Table 4).

Quantiles

Very often, we are also interested in some specific percentiles from the estimated distribution of event times, that is, the times taken to reach a certain proportion of individuals with the event. In seed germination assays, these percentiles are usually defined as, for example,T_{10} (time to 10% germination), T_{30} (time to 30% germination), T_{50} (time to 50% germination), and so on. The reciprocals of these quantities ($1/T_x$, with x being the germination percentage) are also of interest and are often referred to as the germination rates (e.g., GR_{10}, GR_{30}, and GR_{50}), and they represent the daily progress to germination and provide an intelligible measure of "speed."

The concept of quantiles for time-to-event studies is closely related to the concept of effective doses (ED) in dose–response assays, although we discourage the use of this term for time-to-event studies, as the main predictor is not, in strict terms, a dose, but rather time. As a general suggestion for time-to-event assays, quantiles should be calculated for the whole sample, and they should not be restricted to the fraction of individuals that have experienced the event, especially when the purpose is to make comparisons across treatment groups (Bradford 2002; Keshtkar et al. 2021).

In order to calculate quantiles in DRCTE, we have included the *ED()* function for compatibility with the DRC package, although we recommend the use of the 'quantile()' method, which is specifically tailored to deal with statistical distributions. Quantiles can be calculated considering the whole population of individuals (default choice) or they can be restricted to the fraction of individuals with the event (use the 'restricted = T' argument), although this second choice (corresponding to 'type = "relative"' in DRC) is discouraged. Quantiles can also be calculated for rates, by using the 'rate = T' argument. In all cases, SEs are calculated according to the same techniques as discussed before for predictions, and the 'units' argument can also be used to account for the experimental design.

Table 4. Prediction for the cumulative proportion of germinated seeds from the data of Example 3 (Figure 4). Cluster robust bootstrap SEs are given.

Verbascum species	Time	Prediction	SEs
	day		
V. arcturus	5	0.00	0.0000000
V. arcturus	10	0.05	0.0315233
V. blattaria	5	0.73	0.0559562
V. blattaria	10	0.84	0.0405031
V. creticum	5	0.97	0.0237803
V. creticum	10	0.97	0.0237803

Example 3 Continued

Here we compare the three *Verbascum* species in terms of the germination rates for the 10th and 50th percentiles. Bootstrap estimators with cluster-robust SEs are reported in Table 5.

Discussion

Time-to-event data are very common in agriculture, but they have been very rarely recognized as such, in contrast to the medical sciences, where their evaluation has almost become a standard procedure. Consequently, no unified framework for data analyses has been established or proposed to date. The framework of survival analysis does not appear to be immediately transferable to agriculture, due to differences in terms of aims, terminology, and experimental approaches. A brief survey of literature shows that, with time-to-event data, agricultural scientists have used or proposed a lot of different methods that are not always correct and/or efficient.

In this article, we have proposed a unified framework that consists of both a procedure and the software facilities to accomplish the challenging task of analyzing time-to-event data. The procedure is based on time-to-event curve fitting, which is the best method to unambiguously describe and compare the whole time course of events. The proposed methodology is not new, and it has been implemented in ecotoxicology with dose–response studies. However, it is necessary to reinforce the idea that the statistical approaches and terminology for time-to-event studies need to be different from those used in ecotoxicological dose–response studies, in order to comply with the typical traits of the observed data (e.g., censoring). Furthermore, time-to-event data may need the use of additional and specific methods that are not commonly used with other types of experimental data in agriculture.

The adoption of a unified framework for data analyses, at least as the starting point, should improve communication among scientists and practitioners and favor the exchange of results and ideas. In this respect, the framework we provide is based on a few selected methods that have been available for several years and have already proven reliable not only in medicine, but also in agriculture or other contexts. The innovation provided by our DRCTE package is that all those methods were placed together in the same "container," using a consistent syntax and data structure.

The results obtained from our framework can also be seen as the starting point for other analyses (second-step analyses). For example, the GR values (or model parameters) obtained in the first step can be used as the response variables to fit ANOVA or linear/nonlinear regression models or they can be used for the application of principal component analyses or other multivariate methods. In this second step, wherever possible, it would be important not to discard relevant information from the first step, such as SEs (Jensen et al. 2017; Ritz et al. 2019), which are needed when

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Table 5. Quantiles for germination rates as derived for the data of Example 3 (Figure 4).

Verbascum species	Index	Estimate	SEs[a]
		day^{-1}	
V. arcturus	GR10	0.0911807	0.0050252
V. blattaria	GR10	0.3337697	0.0276939
V. creticum	GR10	0.4309032	0.0176747
V. arcturus	GR50	0.0006217	0.0065312
V. blattaria	GR50	0.2334323	0.0123284
V. creticum	GR50	0.3043018	0.0127946

[a]Cluster robust bootstrap SEs are given.

adopting a meta-analytic technique using R packages such as METAFOR (Viechtbauer 2010); the preceding references contain several interesting examples.

Obviously, we do not claim that our framework is flawless and gives totally unambiguous results in all settings. On the contrary, there are several limitations that should be addressed in future work. For example, we concentrated only on a few indicators of seed germination velocity/capability/uniformity, but the usefulness of other indices might be considered in future work (e.g., the indices proposed in Cao et al. 2011).

Another limitation is that the tests for comparing nonparametric time-to-event curves have seldom been used in agriculture, and future work should help elucidate which test is advisable in which settings. In particular, it has been shown that rank tests for NPMLE methods may fail to detect significance when two time-to-event curves intersect each other (Gomez et al. 2009). Furthermore, depending on the testing method, early or late differences between time-to-event curves might be given different weights, which might affect the resulting P-level (Fay and Shaw 2010).

Further limitations may appear with complex experimental designs. In this article, we proposed a "marginal" approach, in which variances are modified to comply with possible deviations with respect to the independence assumption (in loose language: cluster-robust SEs). Such an approach is relatively easy to apply when the design structure is simple, whereas the bootstrap approach for nonparametric models may be difficult to implement with multilevel designs. In this case, we suggest that, after fitting the time-to-event curves and deriving the necessary information, one should proceed with the meta-analytic approach proposed by Jensen et al. (2017), which is tailored to comply with all types of designs. An alternative and elegant approach would be to fit a time-to-event mixed model, which has already been proposed in the Bayesian framework (Onofri et al. 2019). Although Bayesian methods might be potentially useful for time-to-event experiments (Humplik et al. 2020), a strong change in attitude toward statistical inference is required for its adoption, which might not be easy for those less familiar with statistics. However, implementing Bayesian methods within the DRCTE package might represent a future challenge to be considered.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/wsc.2022.8

Acknowledgments. The research was partially supported by the funds for the Basic Research of the University of Perugia (Italy). The authors declare that they have no conflict of interest.

References

Barreiro-Ures D, Francisco-Fernández M, Cao R, Fraguela BB, Doallo R, González-Andújar JL, Reyes M (2019) Analysis of interval-grouped data in weed science: the binnednp Rcpp package. Ecol Evol 9:10903–10915

Bolker BM (2008) Ecological Models and Data in R. Princeton, NJ: Princeton University Press. 408 p

Bradford KJ (1995) Water relations in seed germination. Pages 351–396 *in* Kigel J, Galili G, eds. Seed Development and Germination. New York: Marcel Dekker

Bradford KJ (2002) Applications of hydrothermal time to quantifying and modeling seed germination and dormancy. Weed Sci 50:248–260

Brien CJ, Bailey RA (2006) Multiple randomizations. J R Stat Soc Series B Stat Methodol 68:571–609

Brown RF, Mayer DG (1988a) Representing cumulative germination. 1. A critical analysis of single-value germination indices. Ann Bot 61:117–125

Brown RF, Mayer DG (1988b) Representing cumulative germination: 2. The use of the Weibull function and other empirically derived curves. Ann Bot 61:127–138

Cao R, Francisco-Fernández M, Anand A, Bastida F, González-Andújar JL (2011) Computing statistical indices for hydrothermal times using weed emergence data. J Agric Biol Environ Stat 149:701–712

Cao R, Francisco-Fernández M, Anand A, Bastida F, González-Andújar JL (2013) Modeling *Bromus diandrus* seedling emergence using nonparametric estimation. J Agric Res 18:64–86

Carroll RJ, Wang S, Simpson DG, Stromberg AJ, Ruppert D (1998) The Sandwich (Robust Covariance Matrix) Estimator. Department of Statistics Technical Report. College Station TX: Texas A&M University. 16 p

Catara S, Cristaudo A, Gualtieri A, Galesi R, Impelluso C, Onofri A (2016) Threshold temperatures for seed germination in nine species of *Verbascum* (Scrophulariaceae). Seed Sci Res 26:30–46

Efron B (1981) Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 68:589–599

Fay MP, Shaw PA (2010) Exact and asymptotic weighted logrank tests for interval censored data: the interval R package. J Stat Softw 36:1–34

Gianinetti A (2020) Basic features of the analysis of germination data with generalized linear mixed models. Data 5:6

Goldstein H (1999) Multilevel Statistical Models. London: Institute of Education, Multilevel Models Project. 192 p

Gomez G, Colle ML, Oller R, Langohr K (2009) Tutorial on methods for interval-censored data and their implementation in R. Stat Modelling 9:259–297

Gonzalez-Andujar JL, Francisco-Fernandez M, Cao R, Reyes M, Urbano JM, Forcella F, Bastida F (2016) A comparative study between nonlinear regression and nonparametric approaches for modelling *Phalaris paradoxa* seedling emergence. Weed Sci 56:367–376

Hay FR, Mead A, Bloomberg M (2014) Modelling seed germination in response to continuous variables: use and limitations of probit analysis and alternative approaches. Seed Sci Res 24:165–186

Humplik JF, Dostal J, Ugena L, Spichal L, De Diego N, Vencalek O, Fuerst T (2020) Bayesian approach for analysis of time-to-event data in plant biology. Plant Methods 16:14

Jensen SM, Andreasen C, Streibig JC, Keshtkar E, Ritz C (2017) A note on the analysis of germination data from complex experimental designs. Seed Sci Res 27:321–327

Keshtkar E, Kudsk P, Mesgaran MB (2021) Perspective: common errors in dose–response analysis and how to avoid them. Pest Manag Sci 77: 2599–2608

Kleinbaum DG, Klein M (2005) Survival Analysis. New York: Springer Science

Lambert PC, Thompson JR, Weston CL, Dickman PW (2007) Estimating and modeling the cure fraction in population-based cancer survival analysis. Biostatistics 8:576–594

Manso R, Fortin M, Calama R, Pardos M (2013) Modelling seed germination in forest tree species through survival analysis. The *Pinus pinea* L. case study. For Ecol Manag 289:515–524

McCullagh P, Nelder JA (1998) Generalized Linear Models. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC

McNair JN, Sunkara A, Frobish D (2012) How to analyse seed germination data using statistical time-to-event analysis: non-parametric and semi-parametric methods. Seed Sci Res 22:77–95

Mesgaran MB, Mashhadi HR, Alizadeh H, Hunt J, Young KR, Cousens RD (2013) Importance of distribution function selection for hydrothermal time models of seed germination. Weed Sci 53:89–101

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Mesgaran MB, Onofri A, Mashhadi HR, Cousens RD (2017) Water availability shifts the optimal temperatures for seed germination: a modelling approach. Ecol Modell 351:87–95
Onofri A (2021) Time-to-Event Methods for Agriculture. https://github.com/OnofriAndreaPG/drcte. Accessed: February 10, 2022
Onofri A, Benincasa P, Mesgaran MB, Ritz C (2018) Hydrothermal-time-to-event models for seed germination. Eur J Agron 101:129–139
Onofri A, Gresta F, Tei F (2010) A new method for the analysis of germination and emergence data of weed species. Weed Sci 50:187–198
Onofri A, Mesgaran MB, Neve P, Cousens RD (2014) Experimental design and parameter estimation for threshold models in seed germination. Weed Sci 54:425–435
Onofri A, Mesgaran MB, Tei F, Cousens RD (2011) The cure model: an improved way to describe seed germination? Weed Sci 51:516–524
Onofri A, Piepho H-P, Kozak M (2019) Analysing censored data in agricultural research: a review with examples and software tips. Ann Appl Biol 174:3–13
Pace R, Benincasa P, Ghanem ME, Quinet M, Lutts S (2012) Germination of untreated and primed seeds in rapeseed (*Brassica napus* var *oleifera* Del.) under salinity and low matric potential. Exp Agric 48:238–251
Ranal MA, Santana DGD (2006) How and why to measure the germination process? Rev Bras Bot 29:1–11
R Core Team (2021) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing
Ritz C (2010) Toward a unified approach to dose-response modeling in ecotoxicology. Environ Toxicol Chem 29:220–229
Ritz C, Jensen SM, Gerhard D, Streibig JC (2019) Dose-response analysis using R. Boca Raton, FL: CRC Press. 214 p
Ritz C, Pipper CB, Streibig JC (2013) Analysis of germination data from agricultural experiments. Eur J Agron 45:1–6
Ritz C, Pipper C, Yndgaard F, Fredlund K, Steinrucken G (2010) Modelling flowering of plants using time-to-event methods. Eur J Agron 32:155–161
Romano A, Stevanato P (2020) Germination data analysis by time-to-event approaches. Plants 9:617
Scott SJ, Jones RA, Williams WA (1984) Review of data analysis methods for seed germination. Crop Sci 24:1192–1199
Soltani E, Ghaderi-Far F, Baskin CC, Baskin JM (2015) Problems with using mean germination time to calculate rate of seed germination. Aust J Bot 63:631
Therneau TM (2021) A Package for Survival Analysis in R. https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf. Accessed: February 10, 2022
Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. J Stat Softw 36:1–48
Wickham H, Grolemund G (2016) R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 1st ed. Sebastopol, CA: O'Reilly. 520 p
Winkler AM, Webster MA, Vidaurre D, Nichols TE, Smith SM (2015) Multi-level block permutation. Neuroimage 123:253–268
Yu B, Peng Y (2008) Mixture cure models for multivariate survival data. Comput Stat Data Anal 52:1524–1532
Zeileis A, Köll S, Graham N (2020) Various versatile variances: an object-oriented implementation of clustered covariances in R. J Stat Softw 95:1–36

Downloaded From: https://bioone.org/journals/Weed-Science on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use