



## OPEN A clustering-based survival comparison procedure designed to study the *Caenorhabditis elegans* model

Paul-Marie Grollemund<sup>1,2,4</sup>✉, Cyril Poupet<sup>3,4</sup>✉, Élise Comte<sup>1</sup>, Muriel Bonnet<sup>2</sup>,  
Philippe Veisseire<sup>2</sup> & Stéphanie Bornes<sup>2</sup>

*Caenorhabditis elegans* is highly important in current research, serving as a pivotal model organism that has greatly advanced the understanding of fundamental biological processes such as development, cellular biology, and neurobiology, helping to promote major advances in various fields of science. In this context, the survival of a nematode under various conditions is commonly investigated via statistical survival analysis, which is typically based on hypothesis testing, providing valuable insights into the factors influencing its longevity and response to various environmental factors. The extensive reliance on hypothesis testing is acknowledged as a concern in the scientific analysis process, emphasizing the need for a comprehensive evaluation of alternative statistical approaches to ensure a rigorous and unbiased interpretation of research findings. In this work, we propose an alternative method to hypothesis testing for evaluating differences in nematode survival. Our approach relies on a clustering technique that takes into account the complete structure of survival curves, enabling a more comprehensive assessment of survival dynamics. The proposed methodology helps to identify complex effects on nematode survival and enables us to derive the probability that treatment induces a specific effect. To highlight the application and benefits of the proposed methodology, it is applied to two different datasets, one simple and one more complex.

**Keywords** Survival analysis, Clustering, *Caenorhabditis elegans*, Methodological development, Dataset complexity

*Caenorhabditis elegans* is a nonparasitic nematode found in soil, compost heaps, and rich humus<sup>1</sup>. It feeds on fungi and bacteria. The first work on this animal was carried out in 1900 by Maupas<sup>2</sup>, but it was in the 1960s, that this model was used in laboratories by Brenner for genetic studies. In 1998, it was the first multicellular organism whose genome was fully sequenced<sup>3</sup>. In 2000, Lai and colleagues showed, via comparative proteomic analysis, that *C. elegans* has at least 83% (15,344 sequences out of 18,452 proteins) of human orthologs<sup>4</sup>. Genes, signaling pathways, and basic biological functions are therefore conserved<sup>5</sup>. This is the case for apoptosis<sup>6</sup>, immunological mechanisms<sup>7</sup>, stress response<sup>8</sup>, etc. The perfect knowledge of its genome has allowed the generation of more than 12,000 different genetically characterized mutants, which are available to the scientific community, mainly through the *Caenorhabditis* Genetics Center (Minneapolis, MN, United States of America) and the National Bioresource Project for the Experimental Animal "Nematode *C. elegans*" (Tokyo, Japan). *C. elegans* is also amenable to sophisticated yet convenient genetic techniques, such as RNAi feeding, transgenesis via microinjection, mutagenesis screening, and CRISPR/Cas9 genome editing<sup>9,10</sup>. This allowed mechanistic research to understand human diseases using the worm. Moreover, *C. elegans* has a transparent cuticle, and it is possible to visualize gene expression in transgenic strains with various fluorescent proteins in live organisms. It has many other advantages, including a short lifespan (2–3 weeks), the possibility of having a synchronized isogenic population, low-cost maintenance, and no ethical requirements.

In laboratories, its natural and diversified diet is substituted by *Escherichia coli* OP50 or by microorganisms and/or the extract and the active ingredient of interest. Depending on the study, phenotypes such as lifespan,

<sup>1</sup>University of Clermont Auvergne, CNRS, LMBP, Clermont-Ferrand, France. <sup>2</sup>University of Clermont Auvergne, INRAE, VetAgro Sup, UMRP, Aurillac, France. <sup>3</sup>University of Tours, INRAE, ISP, 37000 Tours, France. <sup>4</sup>Paul-Marie Grollemund and Cyril Poupet contributed equally to this work and share the first authorship. ✉email: paul\_marie.grollemund@uca.fr; cyril.poupet@univ-tours.fr

longevity, and mobility are observed. Over the past thirty years, 38,446 publications have referenced the National Center for Biotechnology Information (NCBI) for the nematode *C. elegans*. In the same time interval, we reference 96,488 publications with the keyword lifespan, including 4,702 for the worm. Nematodes have been used in many fields, including those related to aging<sup>11</sup>, immunity<sup>12</sup>, abiotic stresses such as oxidative stress<sup>13</sup>, hypoxia<sup>14</sup>, heat stress<sup>15</sup>, cold tolerance<sup>16</sup>, osmotic stress<sup>17</sup>, UV light<sup>18</sup>, heavy metal stress<sup>19</sup>, biotic stress (pathogenic bacteria<sup>20</sup> and fungi<sup>21</sup>), toxicity testing<sup>22</sup> and lifespan measurement. It is a setting widely used in biology to visualize the effect of a treatment on an organism. For example, the selection of a new probiotic strain is based on in vitro and in vivo tests before human experiments for the last stages of research. In vivo tests involving laboratory animals such as mice, rats, rabbits, and *C. elegans* are among these models, particularly in the screening and identification of therapeutic targets conserved in humans<sup>23</sup>.

Longevity and survival analyses are key tools for the study of new live therapeutic products<sup>24</sup>. In this case, live worms were counted daily. The lifespan curves represent the evolution of the number of living nematodes over time. In nematodes, as in other preclinical models and human clinical studies, the effect of a treatment can be monitored and analyzed with specific statistical methods.

For a range of experiments, the resulting data are counts of living nematodes over time for which a survival analysis can be used to address standard problems. For example, evaluating the potential impact of a given drug is performed by comparing the survival function of a treatment group against that of a control group. A standard approach was described and studied by Petrascheck and Miller<sup>25</sup>, consisting of deriving the Kaplan-Meier method<sup>26</sup> to obtain survival curve estimates and comparing them according to the log-rank test to evaluate significant differences between curves (which reflects the difference between two underlying hazard functions)<sup>27,28</sup>.

The log-rank test is a widely used method in survival analysis for comparing survival curves between groups. It is asymptotically valid but may perform rather poorly when the proportional hazards (PH) assumption does not hold<sup>29</sup> and suffers from its inability to handle competing risks or time-dependent covariates. It can be determined that PH does not hold according, for instance, by computing Schoenfeld residuals; nevertheless, having survival functions that cross each other is strong evidence of PH violation. In that case, the power of the log-rank test is reduced by PH violation, which increases the possibility of incorrectly detecting significant differences between two survival functions.

To overcome these limitations, several extensions of the log-rank test have been developed. These extensions include, for instance, the stratified log-rank test for adjusting for covariates, the weighted log-rank test for incorporating weights, and the generalized log-rank test for accommodating nonproportional hazards. These advancements in statistical techniques enhance the applicability and robustness of survival analysis, providing insights into the relationships between covariates and survival outcomes. However, each alternative method is not straightforward. The tuning parameters, such as the relevant periods over which the survival curves are compared, must be fixed, which could be challenging to determine.

As another limitation in this framework, when several drugs are tested, multiple combinations of pairwise log-rank tests must be computed to obtain an overview of the relative effects of each drug, for example<sup>21</sup>. The performance of multiple tests on the same dataset drastically increases the risk of the testing procedure and promotes false positive results. Although numerous approaches, such as the Bonferroni correction, have been developed to overcome this issue, recent works support the systematic reliance on this type of approach. Moreover, pairwise comparisons provide only a partial view, whereas we aim to discern broader differences among multiple survival curve groups. In other words, tests that compare pairs of conditions fail to consider the overall joint structure of survival curves.

Another limitation is that the log-rank test does not indicate the direction of any potential difference. While some may argue that this is not the test's intended purpose and that defining a comparison of survival function is unclear, we stress that such considerations are integral to statistical analysis. Typically, they are addressed through graphical interpretation or by comparing summaries such as the half-life time. Depending solely on one-dimensional summaries to compare high-dimensional quantities such as survival functions may yield inconsistent results.

Finally, we highlight that relying mainly on statistical significance to distinguish between positive and negative results is not recommended<sup>30</sup>. Moving beyond statistical significance is induced by recent works<sup>30–32</sup>, which are currently part of various scientific fields, such as Hayat et al.<sup>33</sup>, Erickson and Rattner<sup>34</sup>, Campitelli<sup>35</sup>, and Ciapponi et al.<sup>36</sup>. This movement is also motivated by recalling historical aspects of significance testing<sup>37</sup>, which highlight the limitations and misinterpretations often associated with solely relying on statistical significance as a measure of scientific validity.

In this work, we aimed to assess the different impacts of various treatments on nematode survival to determine the benefits of (i) the use of the probiotic strain *Lactocaseibacillus rhamnosus* Lcr35® for the treatment of *Candida albicans* infection<sup>38,39</sup> or (ii) the administration of extracts from a fermented matrix (cheese) on the longevity of the nematode<sup>40</sup>. These studies required the use of the Kaplan-Meier model and the log-rank test to determine significant differences between the experimental conditions. However, owing to the limitations presented above, the performance of these approaches is not sufficient for a detailed understanding of biological mechanisms. The log-rank test only allows for the detection of an overall significant difference between two experimental conditions and does not provide precise information regarding potential differences in specific portions of the survival curves, such as young versus older. Moreover, in cases where the curves intersect, the test may be insufficiently sensitive in detecting a difference that is present, leading to a loss of information regarding the understanding of underlying mechanisms. To address these issues, further developments are necessary to enable comparisons between survival curves while overcoming the mentioned limitations. In particular, comparing survival curves in scenarios with more than two groups requires the implementation of a method that extends beyond traditional hypothesis testing to complement the log-rank test.

To perform a survival analysis and determine the differences induced by experimental drugs, in this paper, we propose the transition from the null hypothesis testing (NHT) paradigm to a method based on a clustering approach. A clustering-based method offers a viable approach that allows for a comprehensive analysis of the data while circumventing the limitations associated with traditional test procedures. This proposal is inspired by ideas such as those proposed by Kamary and colleagues, which involve reframing hypotheses as components of a mixture model<sup>41</sup>. In this context, the use of an unsupervised classification method involves the use of a nonparametric version of this paradigm. Clustering is used in this context as a method of mapping possible effects and then determining how the treatments studied fit into an estimated cluster distribution. The proposed method should be viewed as a complement to the log-rank test by providing an additional perspective on the type and scale of observed survival differences. Overall, the proposed approach offers a robust and flexible methodology for comparing survival curves in scenarios involving more than two groups, enabling comprehensive analysis of complex survival data.

## Materials

As the method is designed for nematode data analysis, it is tested on two different nematode experiments (see following subsections). The aim is to include both simple (from two experiments) and complex datasets in terms of outcomes. The first database refers to the work of Poupet and colleagues<sup>38,39</sup>, whereas the second relates to the research of Cardin and colleagues<sup>40</sup>.

### Treatment of candidiasis in *C. elegans* with *L. rhamnosus* Lcr35®

*C. albicans* fungal infection was induced in *C. elegans* at the L4/young adult stage, and preventive or curative treatments were administered via *E. coli* OP50 or *L. rhamnosus* Lcr35®. The survival assay was conducted according to the methodology described by de Barros et al.<sup>42</sup>, with certain modifications.

For the control groups, monotypic contamination was induced in *C. elegans* by placing them on plates containing only *C. albicans* (BHI plates), *L. rhamnosus* Lcr35® (NGM plates), or *E. coli* OP50 (NGM plates). For preventive treatment, worms were placed on plates containing *L. rhamnosus* Lcr35® or *E. coli* OP50. The worms were subsequently washed with M9 buffer to remove bacteria before being transferred to *C. albicans* plates. For curative treatment, worms were initially placed on *C. albicans* plates and then transferred to *L. rhamnosus* Lcr35® or *E. coli* OP50 plates after washing. All the incubations were carried out at 20 °C. The nematodes were exposed to *C. albicans* for 2 h, while for *E. coli* OP50 and *L. rhamnosus* Lcr35®, various incubation times were tested (2, 4, 6, and 24 h). All plates were supplemented with 0.12 mM 5-fluorodeoxyuridine to maintain a synchronous population.

The infected nematodes were washed off the plates with M9 buffer and transferred to a 6-well microtiter plate, with approximately 50 worms per well. Each well contained 2 ml of BHI/M9 (20%/80%) liquid assay medium supplemented with 0.12 mM 5-fluorodeoxyuridine (Sigma, Saint-Louis, United States). The microtiter plates were then incubated at 20 °C.

The nematodes were observed daily, and they were considered dead when they did not respond to gentle mechanical stimulation. This assay was conducted as three independent experiments, with three wells per condition.

The data analyzed in this paper are derived from our previous work<sup>38,39</sup>.

### Effect of cheese extracts on *C. elegans* longevity

The effects of dry milk extracts on the lifespan of the *C. elegans* N2 strain were assessed through a longevity assay. An agar medium was prepared by dissolving 3 g of NaCl and 6 g of agarose in 1 L of water. The medium was heated to 40 °C and then supplemented with dried milk extracts, following a procedure described elsewhere<sup>40</sup>, and 0.12 mM FUdR was added. The supplemented medium was added to a 24-well plate at 40 °C. To inhibit significant fungal growth, the aliquot was further supplemented with amphotericin B at a final concentration of 1.6 µg/mL. After being poured, the wells were immediately transferred onto ice to solidify the agar, which was subsequently stored at 4 °C until use. Synchronous L4/young adult stage worms were placed in each well, with approximately 20 worms per well, on supplemented agar medium (or agar medium for the control condition). The worms were fed heat-killed *E. coli* OP50 and maintained at 20 °C throughout the experiment. Food supplementation was performed every 3 days to prevent starvation (20 µL of 100 mg/mL suspension).

The nematodes were observed daily, and they were considered dead when they did not respond to gentle mechanical stimulation. This assay was conducted as three independent experiments, with three wells per condition.

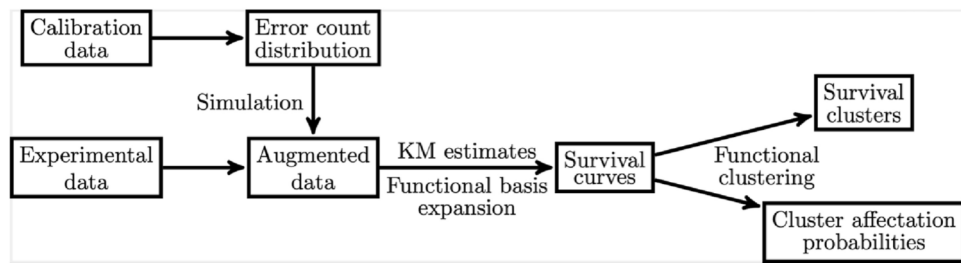
The data analyzed in this paper are derived from our previous work<sup>40</sup>.

## Methods

R software version 4.1.2 was used to implement the proposed methodology. The different steps of the proposed pipeline analysis are described below and are summarized in Fig. 1, which takes, as inputs, experimental data (detailed above 1. Materials section) and the calibration data outlined in Sect. 2.2. The implemented procedure is available at the following address: [https://github.com/pmgrollemund/survival\\_clustering](https://github.com/pmgrollemund/survival_clustering).

### Data augmentation

Nematode survival data are acquired through manual counting under a microscope, which is a meticulous process yielding uncertain outcomes. For example, in situations with numerous nematodes in a sample, a nematode may inadvertently be recounted. Additionally, determining a nematode's vitality is not always definitive and is typically based on movement or characteristic morphology. Consequently, the data exhibit



**Fig. 1.** Pipeline for the proposed survival data analysis.

uncontrolled variability, notably influenced by the experimenter's judgment. Moreover, nematode survival experiments typically involve multiple observers over several days. The counting ability of each experimenter influences raw data and statistical analysis outcomes. Accounting for experimenter effects is challenging, and we propose a method in this section to address this issue. The aim is to ensure that the observed survival differences between experimental conditions are not partly attributed to the presence of multiple experimenters or their respective counting performances. To address this problem in our analysis, we propose adding simulated data to the dataset to mimic this variability. These simulated data reflect what might have been collected by the same or another experimenter and then allow us to model experimenter effects. In the following, these simulated data are used to assess the robustness of the experimental data, particularly by determining whether the simulated data for a specific treatment align similarly to the experimental data for the same treatment.

Prior to obtaining calibration data, the variability in the experimenters' counts must be measured to simulate synthetic data. Therefore, an experiment must be conducted to determine the variability in counting by the experimenters, in which the experimenters counted the same pit over several days and several times a day. This results in a dataset referred to as calibration data. Studying calibration data enables us to infer the distribution of counting errors on the basis of the number of nematodes in the well. By characterizing the distribution of counting errors by experimenters, the aim is to consider the intensity of these errors concerning the analysis of data from other experiments. To accomplish this, we propose in this work to proceed with simulation to augment the database with survival curves that could have plausibly been observed under similar experimental conditions. To achieve this goal, we have developed an algorithm that simulates counting errors on the basis of the estimated distribution while ensuring that a count on the same day cannot be strictly greater than the previous count for the same well. For each datum at a given time point, a counting error is simulated on the basis of the empirical distribution of counting errors. For further details, refer to the GitHub repository containing all the code for this method.

### Data preparation

Survival curves are estimated according to a standard Kaplan–Meier procedure (with the R package *survminer* 0.4.9). To compare the curves for each time, an interpolation is performed so that survival curves are evaluated on the same temporal grid. As the monotonicity of the survival curves is an important feature to preserve, the interpolation method chosen is constrained spline estimation<sup>43</sup>, which is performed with the R package *ConSpline* 1.2. As the aim of this study was to measure the difference in terms of survival between experimental conditions and a control group, we computed the average survival curve (ASC) of the control group, and we derived the deviance survival curve for a given survival curve  $S(\cdot)$  as  $ASC(t) - S(t)$ , at each time  $t$  of the overall considered time period.

### Functional clustering

The deviance survival curves are clustered with the discriminative functional mixture model<sup>44</sup> by using the R package *funFEM* 1.2. To obtain relevant results, covariance models and several clusters are chosen according to the ICL criterion<sup>45</sup>. In this case, the resulting clusters can be viewed as being sufficiently different for certain factors and can therefore be seen as evidence of difference in the same way as a log-rank test. As a major difference, the log-rank test and these weighted alternatives evaluate the difference in (weighted) average survival curves, whereas a functional clustering approach can detect differences at different temporal locations and different temporal scales. In practice, the range of possible cluster numbers is from 2–10, avoiding large numbers since a large cluster number negatively impacts the ability to interpret the results. Note that clustering is performed only on experimental data, not on simulated data, ensuring that the number and simulation method of simulated data do not influence the clusters found. The simulated data are solely used for the postprocessing phase described in Sect. 2.3.

To facilitate interpretation, a label is assigned to each cluster, corresponding to the way the survival curve deviates from ASC. For the sake of interpretation, determining deviation consists of segmenting the time domain into distinct intervals and evaluating whether the average difference from the ASC for each interval exceeds a given threshold  $s$ . For instance, a cluster with an average curve  $D$  is labeled “+ + + / - =” if

$$s < \int_{I_1} D(t) dt \leq 2s \text{ and } 2s < \int_{I_2} D(t) dt \text{ and } -2s \leq \int_{I_3} D(t) dt < -s \text{ and } -s \leq \int_{I_4} D(t) dt \leq s$$

where the indexed intervals  $I_1$  to  $I_4$  correspond to four default intervals used for segmenting the time domain of survival curves. These intervals are predefined but can be manually adjusted as needed. According to the range of survival variation, the parameter  $s$  is by default fixed as 0.15.

### Assessing probability allocation

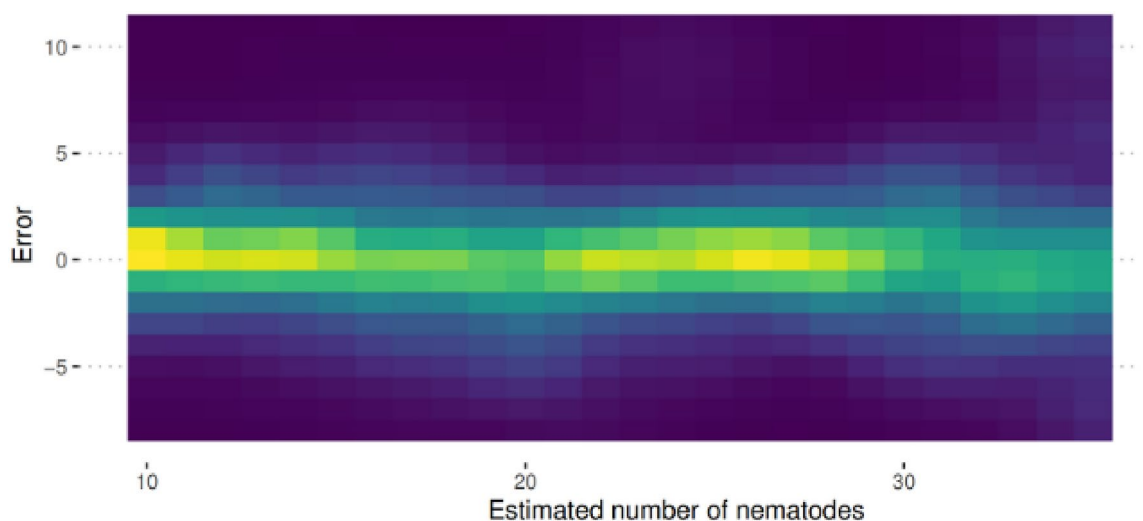
Survival deviation clusters allow us to define the different patterns of variation occurring in the database, considering the temporal dimension of the data. In other words, clusters correspond to “effect groups”, or in other words, to groups representing various potential effects on nematode survival. To link these effect groups with experimental conditions and then determine the impact of treatments on nematode survival, we must ascertain how treatments are distributed across the detected clusters by providing a degree of association between a treatment and the set of detected possible effects. This degree of association is computed via the clustering method since it involves a probabilistic model (Gaussian mixture model), from which we derive the probability of each curve belonging to each cluster. The association of a treatment with a cluster is defined as the average degree of association of each survival curve in the cluster. Note that considering augmented data allows us to more robustly determine the probability of assigning a treatment to a cluster. Indeed, taking into account the variability in counting errors helps to put into perspective the degree of association to which treatment is effectively associated with a possible effect on nematode survival. In practice, such computations rely on the ability to compute the probability of cluster allocation for data not used to fit the clustering model (simulated data), but the current implementation of the method (funFEM 1.2) is not able to compute the cluster allocation probability for this type of data. We then implement a new prediction procedure by retrieving the estimated matrix representation of the functional data and by using the hidden function “estep” of the R package funFEM to obtain the cluster allocation probability. The newly implemented procedure allows us to indicate how each treatment probably distributes on the estimated clusters, according to variability in experimenter counting performance.

## Results

### Variability in experimenter counts

To incorporate experimenter variability into the data analysis, an additional experiment was conducted. During this experiment, two experimenters counted the number of nematodes in the same wells multiple times a day. The experiment spanned 15 days, with 10 counts per day per experimenter. By calculating the mean counts per well, an approximation of the true number of nematodes in the well was obtained, allowing for the determination of counting errors at each attempt. Specifically, it was estimated that the wells contained a maximum of 35 nematodes at the beginning of the experiment and a minimum of 10 nematodes at the end. This is not concerning if the results do not exceed these thresholds since laboratory experiments rarely exceed the maximum threshold, and in addition, we consider that counting errors are minimal below ten nematodes.

The number of counting errors estimated in this manner ranged from  $-8$  to  $11$ , which was consistently observed across both tested experimenters. Notably, the intensity of counting errors only marginally increases with an increase in the number of nematodes in the well. When the estimated number of nematodes is lower than 15, the standard deviation of the counting error is approximately 1.363, and when the estimated number of nematodes is greater than 30, the standard deviation is approximately 2.606. Furthermore, the distribution of counting errors is quite symmetrical: 31.88% of underestimation counts and 34.06% of overestimation counts. Additionally, Fig. 2 provides a graphical representation of the distribution of counting errors on the basis of the estimated number of nematodes in the well. To produce this heatmap, two steps were undertaken. The first step involved estimating the error distribution on the basis of the estimated number of nematodes in the well.



**Fig. 2.** Heatmap of the counting error distribution based on the estimated number of nematodes in the well. The light color represents a high probability, whereas the dark color indicates a low probability.



The obtained distributions were subsequently smoothed to ensure the continuity of the results as the number of nematodes varied.

### Survival clustering

In our study, we employed two distinct datasets to evaluate the performance of our statistical analysis model. Through conventional log-rank analysis, one dataset proved straightforward for interpretation, yielding robust and significant differences. Conversely, the other dataset posed a more intricate challenge because of the crossing survival curves, making interpretation complex. In cases where survival curves are intertwined, the log-rank test loses its power and becomes less suited for analysis. The underlying objective was to ascertain whether clustering analysis could streamline result interpretation, especially in scenarios where traditional log-rank analysis encounters limitations.

### Analysis of the impact of incubation time on worm survival

Figure 3 illustrates the impact of varying incubation times (2, 4, 6, and 24 h) with the nonpathogenic *E. coli* OP50 strain on the survival of *C. elegans*. Figures 3A and D illustrate the evolution of the nematode survival probability for each experimental condition, and it becomes apparent that the incubation time with the bacterium does not seem to have a discernible impact, with the average curves crossing. This outcome is particularly evident following the categorization analysis of the data. As depicted in Fig. 3C, we observed that the four experimental conditions were consistently grouped into four clusters. Within cluster 1, the “6 h” condition predominates, whereas the “4 h” and “24 h” conditions are represented at similar levels. In cluster 2, the “2-h” and “24-h” conditions constitute the majority at approximately 50%, whereas the other two conditions constitute approximately 25%. The last two clusters are at relatively weak levels for every condition, especially for the “2 h” condition in cluster 3 and the “6 h” and “24 h” conditions in cluster 4.

Examining the profiles of the various clusters’ curves (Fig. 3B and E), we note that cluster 4 leads to a subtle rise in the probability of nematode population survival for later times, specifically beyond 8–10 days. In contrast, cluster 2 is more conducive to shorter durations, approximately 5 days. Moreover, clusters 1 and 3 exhibit an intermediate trend, falling between the characteristics of clusters 2 and 4.

In conclusion, the outcomes derived from the clustering tool suggest that incubation time with the control bacterium has no discernible effect on nematode survival. The curves representing different incubation periods overlap consistently, indicating that any chosen incubation time yields indistinguishable results. This implies that, from a survival perspective, the nematode is resilient regardless of the specific duration of exposure to the control bacterium. Importantly, the clustering results align with our previous research<sup>38</sup>, where we utilized the conventional log-rank method for analysis. This consistency in interpretation underscores the robustness of the clustering approach, demonstrating its effectiveness in yielding results comparable to those of established methodologies. This finding reinforces the notion that irrespective of the analytical tool employed, the lack of significant differences in survival probabilities remains a consistent finding across various incubation times with the control bacterium.

Figure 4 shows the effects of various incubation times (2, 4, 6, and 24 h) with the probiotic *L. rhamnosus* Lcr35<sup>+</sup> on the survival of *C. elegans*. As depicted in Fig. 4A, all the conditions resulted in increased longevity compared with the control condition. However, this increase is not uniform, suggesting the presence of a hierarchy among the experimental conditions.

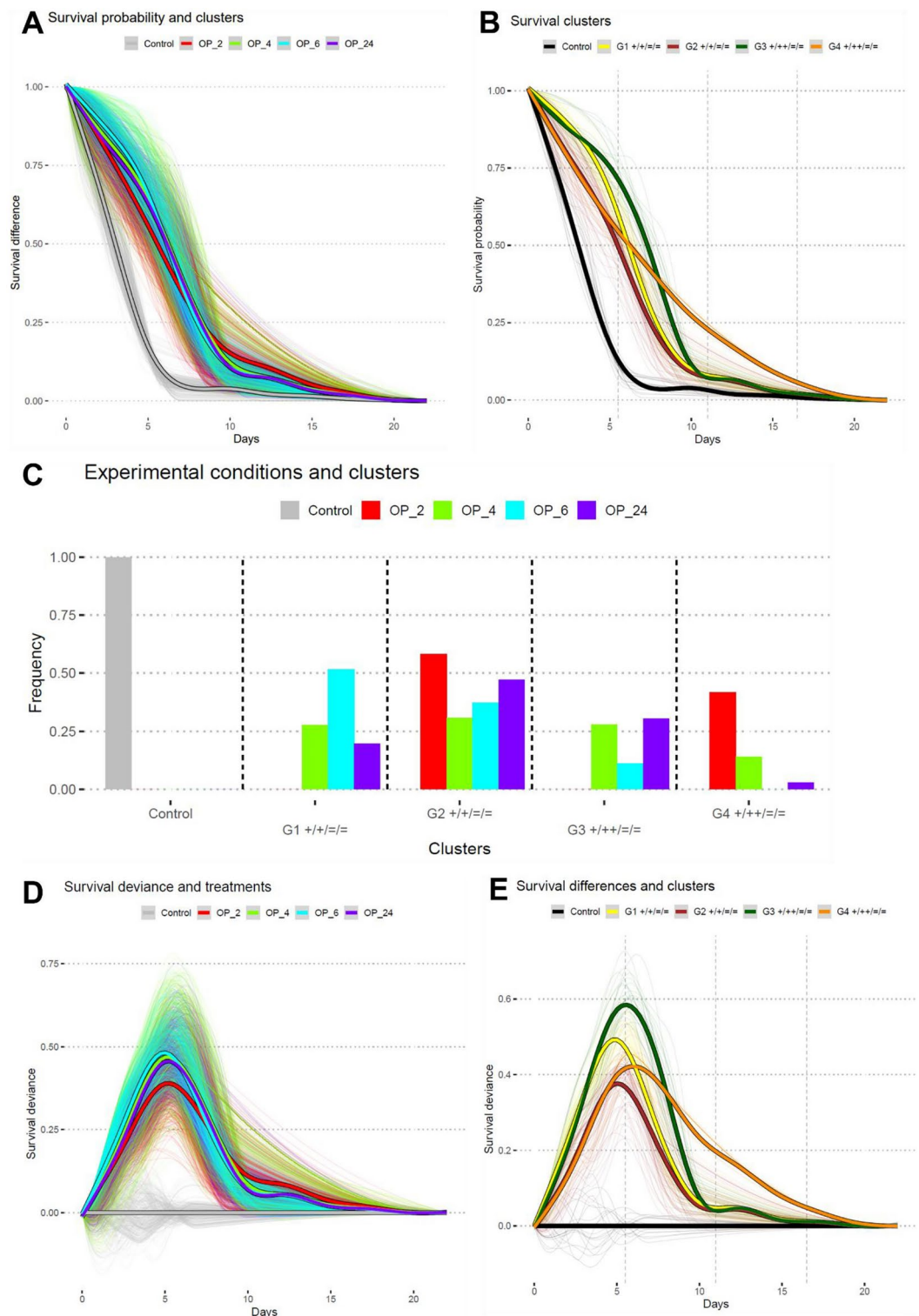
As shown in Fig. 4E, both the 2-h and 4-h conditions clustered together in groups 2, and 3 at relatively similar levels. Notably, the 2-h condition predominates in Group 3, whereas the 4-h condition prevails in Groups 2. Concerning the survival curves, it is apparent that the two conditions display similar patterns. Compared with that of the control, the difference in survival was equal, peaking after approximately 7.5 days (Fig. 4D). However, in groups 2 and 3, there was a noticeable discrepancy in survival, with Group 2 exhibiting a more pronounced and slightly delayed effect. This distinction becomes particularly evident during intermediate and longer time intervals, roughly between 5 and 15 days (Fig. 4B).

In contrast, incubation with *L. rhamnosus* Lcr35<sup>+</sup> for 6 h resulted in a moderate decrease in longevity compared with the previous experimental conditions. Categorization, however, is intricate, as four subpopulations appear to cluster relatively evenly in groups 1, 2, 3, and 4. This finding suggests that this experimental condition represents a pivotal incubation time, marking a transition between optimal incubation times for the nematode and those where the impact is still positive but not relevant.

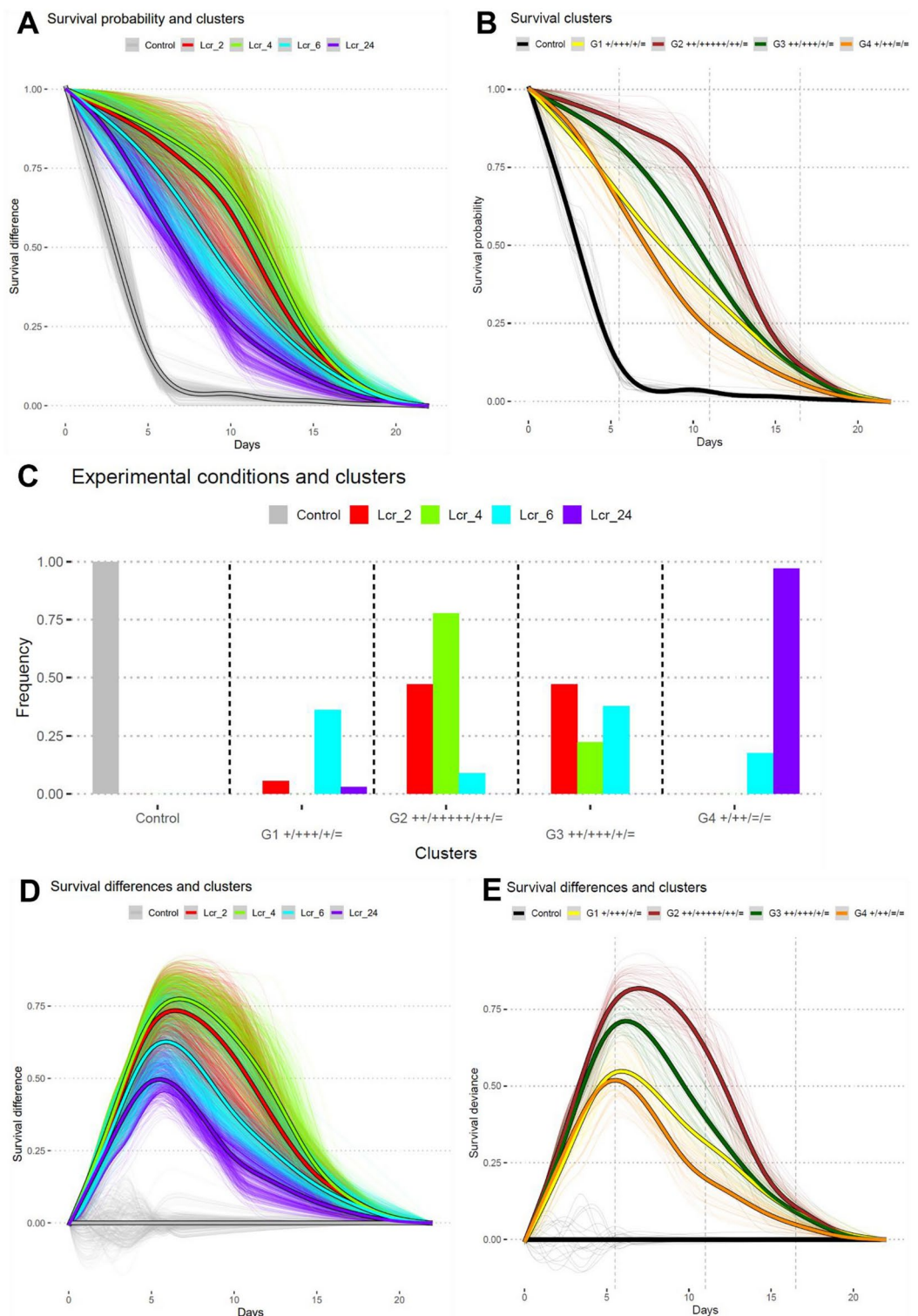
Notably, the condition corresponding to a 24-h incubation time stands out distinctly from the other conditions. According to Fig. 4E, this condition predominantly resides in Group 4, which also includes the 6-h incubation condition, comprising approximately 95% of the group. Graphically, (Fig. 4A–D and, in comparison to the control group, this group had a relatively low probability of survival or a difference in survival from early time points (approximately 0 to 5 days) and significantly reduced levels compared with those of the control groups at intermediate time points (5 to 10 days).

Consequently, it can be inferred that, from a host benefit perspective, the 24-h incubation time is not relevant. The same conclusion applies to the 6-h condition, which is further clustered within multiple groups. Owing to their high similarity, the 2-h and 4-h conditions can ultimately be considered without statistical distinctions.

In conclusion, the findings suggest that the variation in incubation time with the probiotic *L. rhamnosus* Lcr35<sup>+</sup> has a discernible effect on the survival of *C. elegans*. The nuanced differences observed among the experimental conditions, particularly the distinctive patterns in groups 1, 3, and 4, highlight the intricate relationship between incubation duration and nematode longevity. Notably, the 24-h incubation time stands out as having a significantly different effect, leading to reduced survival probabilities compared with shorter incubation periods. Importantly, the clustering of the 6-h condition within multiple groups further emphasizes its complex categorization, suggesting a transitional impact. Interestingly, despite the subtleties uncovered through detailed



**Fig. 3.** Impact of incubation time with *E. coli* OP50 on the survival of *C. elegans*. **(A)** Survival curves (observed and simulated) for nematodes, with each curve color-coded according to incubation time (2, 4, 6, and 24 h). The mean curve for each incubation time is highlighted, and it is calculated on the basis of both observed and simulated data. **(D)** Same as in Plot A, except curves are color-coded on the basis of the allocated cluster. **(C)** Correspondence between each experimental condition and each cluster, indicating how frequently a curve associated with a specific incubation time is allocated to a particular cluster. **(B)** Deviations of survival curves (observed and simulated) from the mean curve of the control group, with each curve color-coded according to incubation time. **(E)** Same as in Plot D, except curves are color-coded on the basis of the allocated cluster. This evaluation helps assess the likelihood of experimental conditions coinciding with a type of effect on nematode survival identified with a specific estimated cluster.



**Fig. 4.** Impact of incubation time with *L. rhamnosus* Lcr35<sup>®</sup> on the survival of *C. elegans*. **(A)** Survival curves (observed and simulated) for nematodes, with each curve color-coded according to incubation time (2, 4, 6, and 24 h). The mean curve for each incubation time is highlighted, and it is calculated on the basis of both observed and simulated data. **(B)** Same as in Plot A, except curves are color-coded on the basis of the allocated cluster. **(C)** Correspondence between each experimental condition and each cluster, indicating how frequently a curve associated with a specific incubation time is allocated to a particular cluster. **(D)** Deviations of survival curves (observed and simulated) from the mean curve of the control group, with each curve color-coded according to incubation time. **(E)** Same as in Plot D, except curves are color-coded on the basis of the allocated cluster. This evaluation helps assess the likelihood of experimental conditions coinciding with a type of effect on nematode survival identified with a specific estimated cluster.



analysis, the ultimate host benefits appear to align with those obtained through traditional statistical methods, such as log-rank analysis, as reported in previous research<sup>38</sup>. This finding supports the robustness and reliability of the statistical clustering approach in yielding results that are consistent with established methodologies.

### Comparative analysis of a complex dataset for the characterization of cheese extracts.

The adult nematodes were exposed to various extracts of goat cheese under two control conditions: heat-inactivated *E. coli* OP50 and heat-inactivated *E. coli* OP50 supplemented with an antifungal agent (some samples being nonsterile). The results are presented in Fig. 5.

Figure 5A depicts the survival probability of nematodes under the experimental conditions. Only the lipid and ethanol fractions presented survival curves that were significantly different from those of the other fractions, particularly from those of the control samples. Specifically, the lipid fraction shows a relative increase in survival (approximately the 12<sup>th</sup> day), whereas the ethanol fraction leads to a decrease in survival during the early stages (approximately the first five days). This fact is indeed very clear in the curve in Fig. 5B. For the remaining experimental conditions, no clear trend emerged from the analysis of their survival curves. The overall appearances of the two control conditions (OP50 and OP50AF) were similar. However, examination of the differences in relative survival presented in Fig. 5D revealed that there was still a discernible difference in behavior induced by the two experimental conditions. The condition of heat-inactivated *E. coli* OP50 in the presence of the antifungal agent appeared to promote better population survival. Thus, although both conditions are classified among the controls (Fig. 5E), it seems unlikely that they can be interchangeably substituted for one another.

Examining the corresponding median curves of the various experimental groups revealed distinct patterns. Compared with the control group, Group 1 tended to have a decreased survival probability, whereas Group 2 had the opposite outcome, with an increased survival probability. By visually comparing the curves of the clusters (Fig. 5B and C) with those of the experimental conditions, we observe that the curves of group 1 closely overlap with those of the conditions corresponding to the ethanol fraction, whereas those of group 2 closely overlap with the conditions of the lipid fraction. These findings demonstrate that these fractions are the predominant representatives of their respective groups.

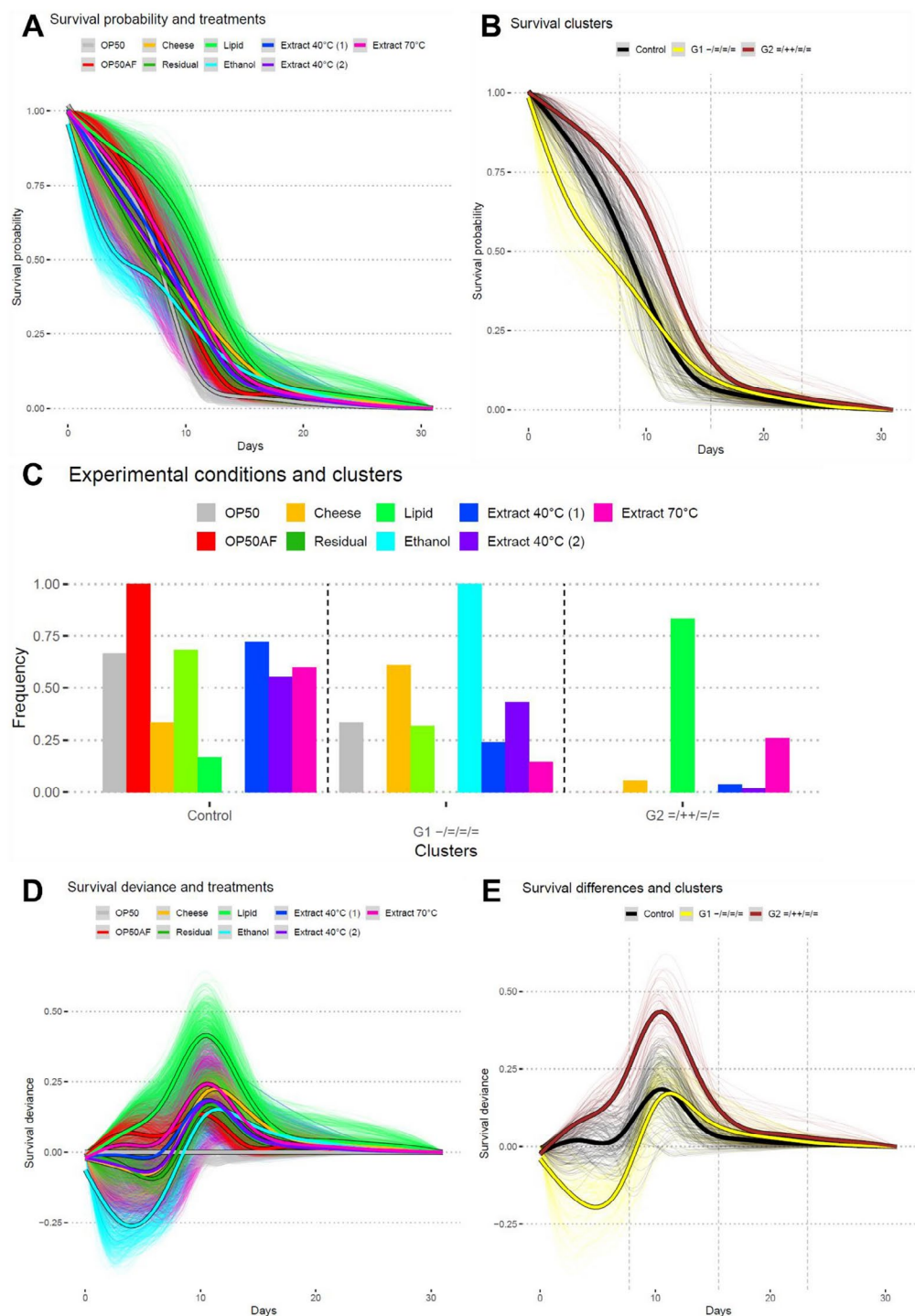
Unlike the results obtained with the previous dataset (i.e., incubation time with *E. coli* OP50 and *L. rhamnosus* Lcr35<sup>+</sup>), Fig. 5E illustrates a markedly different classification of experimental conditions. In this case, our model highlights only two distinct experimental clusters, with six experimental conditions grouped alongside the two controls. According to the model, only cheese extracts obtained from ethanol treatment and the lipid fraction were predominantly found in different clusters, specifically cluster 1 and cluster 2, respectively. Thus, the obtained classification indicates that the majority of cheese extracts do not impact the longevity of nematodes, as they exhibit statistically similar behavior to that of the control.

Comparing the classification obtained from our new analytical model with the interpretation we provided in a previous article<sup>40</sup> regarding the two control conditions (OP50 and OP50AF), we previously reported that the addition of amphotericin B resulted in a significant difference, according to the log-rank test, promoting the impact of this molecule on population survival. However, our initial analysis did not reveal that, overall, the two experimental conditions were very similar, as we demonstrated here. For the other experimental conditions, our initial approach involved pairwise comparisons against the control to highlight significant differences. Consequently, we were unable to perform a comprehensive comparative analysis such as the one conducted in the present study.

## Discussion

A major contribution of this work is the proposal of a new approach for survival data analysis, which is specifically tailored to experiments conducted on the nematode *C. elegans*. By addressing the nuances of complex datasets and potential pitfalls in conventional methods, our approach aims to provide a more insightful and accessible framework for drawing meaningful conclusions from diverse datasets. This approach avoids the use of the conventional hypothesis testing framework, which typically includes the log-rank test. Instead, we propose employing a clustering method to group data into coherent clusters to identify survival effect profiles. By assessing the correspondence between these clusters and the experimental condition groups (defined by treatments), the potential treatment effects and their likelihood intensities are determined. In addition, the integration of simulated data (based on the experimenter's counting performance) helps to account for a source of variability that is commonly ignored in the analysis of nematode survival, even though it may be significant.

As demonstrated earlier in this article, our new method of statistical analysis for survival data provides a significantly more robust approach, facilitating a clearer interpretation of underlying biological mechanisms. When a dataset reveals easily identifiable significant differences through the log-rank test, our derived classification further reinforces these findings. Notably, for complex datasets requiring in-depth analysis, where conventional tests may prove insufficiently powerful or adaptable, our innovative methodology overcomes these challenges by offering visual classification. This greatly streamlines the work for researchers, making the comprehension and thorough exploration of intricacies within complex data much more accessible. This study contributes to a deeper understanding of the temporal dynamics of probiotic-nematode interactions and validates the efficacy of statistical clustering as an alternative analytical tool in this context. Moreover, our current approach allows us to obtain new results and deduce a different interpretation, offering an alternative perspective that leads to partially different conclusions.



**Fig. 5.** Impact of goat cheese extracts on the longevity of the nematode. **(A)** Survival curves (observed and simulated) for nematodes, with each curve color-coded according to cheese extract level (control OP50; control OP50AF with antifungal, raw cheese, cheese residual, lipid fraction, ethanol extraction, extraction at 40 °C (1), extraction at 40 °C (2) and extraction at 70 °C). The mean curve for each cheese extract level is highlighted, and it is calculated on the basis of both observed and simulated data. **(B)** Same as in Plot A, except curves are color-coded on the basis of the allocated cluster. **(C)** Correspondence between each experimental condition and each cluster, indicating how frequently a curve associated with a specific cheese extract level is allocated to a particular cluster. **(D)** Deviations of survival curves (observed and simulated) from the mean curve of the control group, with each curve color-coded according to cheese extract level. **(E)** Same as in Plot D, except curves are color-coded on the basis of the allocated cluster. This evaluation helps assess the likelihood of experimental conditions coinciding with a type of effect on nematode survival identified with a specific estimated cluster.

## Conclusion

In conclusion, our innovative approach to survival comparison using clusters emerges as a robust alternative to traditional hypothesis testing in the study of the *C. elegans* model. This methodology provides a thorough interpretation of results, addressing pivotal questions about whether the survival curve differs from the control curve and, if so, in what manner.

The primary advantage of our procedure lies in its ability to circumvent biases and limitations inherent in conventional hypothesis-testing approaches. On the basis of this work, a comprehensive procedure involves implementing the log-rank test along with the functional clustering process proposed in this paper. The resulting analysis provides a range of relevant evidence for analyzing differences between survival curves, especially when multiple experimental conditions are tested. This proves particularly beneficial for complex datasets, preserving the integrity of log-rank results while providing supplementary insights to inform decisions regarding the experimental condition's effect compared with the control.

Furthermore, as another major advantage, our method considers external sources of variability absent in a standard database concerning nematode lifetime, especially by incorporating experimenter variability into the procedure for determining differences between survival curves. Its adaptability and potential for improvement in future experiments underscore its efficacy as an analytical tool of choice for survival studies involving the *C. elegans* model. In summary, this approach makes a significant contribution to survival analysis methodology, paving the way for deeper investigations and a more nuanced understanding of results across diverse experimental contexts.

Our model is poised for evolution to address various forms of variability, considering factors such as individual experimenters and the integration of datasets. The aim is to create a more robust control condition by merging datasets, thereby enhancing the reliability and generality of our findings. Additionally, the model will undergo refinement to normalize variability, considering diverse sources of variation, which may arise in merging databases resulting from the work of different experimenters. This involves developing mechanisms to systematically account for experiment-to-experiment variations, enabling a more comprehensive understanding of the experimental landscape. By continually adapting to the intricacies of different experimental setups and datasets, our evolving model is geared toward establishing a standardized and versatile platform for survival analysis in the *C. elegans* model. This forward-looking approach ensures that our methodology remains at the forefront of addressing the complex and dynamic nature of biological experiments, ultimately contributing to more accurate and interpretable outcomes in future studies.

One current limitation of this work, which could inspire further research, is that the method for classifying survival curves is not specific to survival curves but applies more generally to any dataset containing functional data. However, survival curves are a particular type of functional data because they are constrained through their integral value, representing a specific subset of the functional data space. The purpose would then be to develop a functional clustering approach that takes this constraint into account to better discriminate survival curves. In our view, this should not necessarily involve working with a parametric version of survival curves, as this would significantly reduce the richness of survival curves. One potential approach could be to investigate the constraints this would impose on the coefficients of a suitable basis function, such as those in a B-spline basis. This could help determine the representation of survival curves in a coefficient space specific to survival data analysis. It may then be possible to use a standard clustering method on the coefficients of survival curves expressed in this specific basis function space.

Another lead of work involves deploying the proposed analysis into a more comprehensive and automated analysis process. This process aims to define a protocol for conducting experiments on nematode lifetime and to build a database to collect experimental results. This will include assessing experimenter counting performance, log-rank test results, and the complete analysis conducted in this article. Additionally, it involves searching for databases where the experimental conditions for the control group are similar. This is intended to enable comparison with results from other experiments and to augment the control group with new data, thereby enhancing the ability to effectively discriminate survival curves. Notably, including control group data from other experiments or comparing different experiments can be facilitated by normalizing databases on the basis of experimenter count performance.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 6 May 2024; Accepted: 13 November 2024

Published online: 16 November 2024

## References

- Hodgkin, J. & Doniach, T. Natural Variation and Copulatory Plug Formation in *Caenorhabditis elegans*. *Genetics* **146**(1), 149–164. <https://doi.org/10.1093/genetics/146.1.149> (1997).
- D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess, Eds., *C. elegans II*, 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 1997. Accessed: Mar. 25, 2024. [Online]. Available: <http://www.ncbi.nlm.nih.gov/books/NBK19997/>
- The C. elegans Sequencing Consortium\*, “Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology,” *Science*, **282**5396, 2012–2018 <https://doi.org/10.1126/science.282.5396.2012> (1998).
- C.-H. Lai, C.-Y. Chou, L.-Y. Chang, C.-S. Liu, and W. Lin, “Identification of Novel Human Genes Evolutionarily Conserved in *Caenorhabditis elegans* by Comparative Proteomics,” *Genome Res.*, vol. 10, no. 5, pp. 703–713, May 2000, <https://doi.org/10.1101/gr.10.5.703>.
- Leung, M. C. K. et al. *Caenorhabditis elegans*: An Emerging Model in Biomedical and Environmental Toxicology. *Toxicol. Sci.* **106**(1), 5–28. <https://doi.org/10.1093/toxsci/kfn121> (2008).

6. J. Z. Malin and S. Shaham, “Cell Death in *C. elegans* Development,” *Curr. Top. Dev. Biol.* **114**, 1–42 <https://doi.org/10.1016/bs.ctdb.2015.07.018> (2015).
7. Ermolaeva, M. A. & Schumacher, B. Insights from the worm: The *C. elegans* model for innate immunity. *Semin. Immunol.* **26**(4), 303–309. <https://doi.org/10.1016/j.smim.2014.04.005> (2014).
8. Rodriguez, M., Snoek, L. B., De Bono, M. & Kammenga, J. E. Worms under stress: *C. elegans* stress response and its relevance to complex human disease and aging. *Trends Genet.* **29**(6), 367–374. <https://doi.org/10.1016/j.tig.2013.01.010> (2013).
9. Park, H.-E.H., Jung, Y. & Lee, S.-J.V. Survival assays using *Caenorhabditis elegans*. *Mol. Cells* **40**(2), 90–99. <https://doi.org/10.14348/molcells.2017.0017> (2017).
10. Eroglu, M., Yu, B. & Derry, W. B. Efficient CRISPR /Cas9 mediated large insertions using long single-stranded oligonucleotide donors in *C. elegans*. *FEBS J.* **290**(18), 4429–4439. <https://doi.org/10.1111/febs.16876> (2023).
11. Wang, X. et al. Ageing induces tissue-specific transcriptomic changes in *Caenorhabditis elegans*. *EMBO J.* **41**(8), e109633. <https://doi.org/10.15252/embo.2021109633> (2022).
12. Goswamy, D., Gonzalez, X., Labeled, S. A. & Irazoqui, J. E. *C. elegans* orphan nuclear receptor NHR-42 represses innate immunity and promotes lipid loss downstream of HLH-30/TFEB. *Front. Immunol.* **14**, 1094145. <https://doi.org/10.3389/fimmu.2023.1094145> (2023).
13. Cardin, G. et al. A Mechanistic Study of the Antiaging Effect of Raw-Milk Cheese Extracts. *Nutrients* **13**(3), 897. <https://doi.org/10.3390/nu13030897> (2021).
14. Powell-Coffman, J. A. Hypoxia signaling and resistance in *C. elegans*. *Trends Endocrinol. Metab.* **21**(7), 435–440. <https://doi.org/10.1016/j.tem.2010.02.006> (2010).
15. R. N. Plagens, I. Mossiah, K. S. Kim Guisbert, and E. Guisbert, “Chronic temperature stress inhibits reproduction and disrupts endocytosis via chaperone titration in *Caenorhabditis elegans*,” *BMC Biol.* **19**(1), 75 <https://doi.org/10.1186/s12915-021-01008-1> (2021).
16. Takagaki, N. et al. The mechanoreceptor DEG-1 regulates cold tolerance in *Caenorhabditis elegans*. *EMBO Rep.* **21**(3), e48671. <https://doi.org/10.15252/embr.201948671> (2020).
17. D. Chandler-Brown et al., “Sorbitol treatment extends lifespan and induces the osmotic stress response in *Caenorhabditis elegans*,” *Front. Genet.* **6**, <https://doi.org/10.3389/fgene.2015.00316> (2015).
18. Deng, J., Bai, X., Tang, H. & Pang, S. DNA damage promotes ER stress resistance through elevation of unsaturated phosphatidylcholine in *Caenorhabditis elegans*. *J. Biol. Chem.* **296**, 100095. <https://doi.org/10.1074/jbc.RA120.016083> (2021).
19. S. Moyson, R. M. Town, K. Vissenberg, and R. Blust, “The effect of metal mixture composition on toxicity to *C. elegans* at individual and population levels,” *PLOS ONE*, vol. 14, no. 6, p. e0218929, Jun. 2019, <https://doi.org/10.1371/journal.pone.0218929>.
20. Veisseire, P. et al. Investigation into In Vitro and In Vivo *Caenorhabditis elegans* Models to Select Cheese Yeasts as Probiotic Candidates for their Preventive Effects against *Salmonella* Typhimurium. *Microorganisms* **8**(6), 922. <https://doi.org/10.3390/microorganisms8060922> (2020).
21. Poupet, C. et al. In vivo investigation of Lcr35<sup>+</sup> anti-candidiasis properties in *Caenorhabditis elegans* reveals the involvement of highly conserved immune pathways. *Front. Microbiol.* **13**, 1062113. <https://doi.org/10.3389/fmicb.2022.1062113> (2022).
22. Hunt, P. R. The *C. elegans* model in toxicity testing. *J. Appl. Toxicol.* **37**(1), 50–59. <https://doi.org/10.1002/jat.3357> (2017).
23. Basic, M. et al. Approaches to discern if microbiome associations reflect causation in metabolic and immune disorders. *Gut Microbes* **14**(1), 2107386. <https://doi.org/10.1080/19490976.2022.2107386> (2022).
24. Poupet, C., Chassard, C., Nivoliez, A. & Bornes, S. *Caenorhabditis elegans*, a Host to Investigate the Probiotic Properties of Beneficial Microorganisms. *Front. Nutr.* **7**, 135. <https://doi.org/10.3389/fnut.2020.00135> (2020).
25. Petrascheck, M. & Miller, D. L. Computational Analysis of Lifespan Experiment Reproducibility. *Front. Genet.* **8**, 92. <https://doi.org/10.3389/fgene.2017.00092> (2017).
26. Kaplan, E. L. & Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **53**(282), 457–481. <https://doi.org/10.1080/01621459.1958.10501452> (1958).
27. Pletcher, “Model fitting and hypothesis testing for age-specific mortality data,” *J. Evol. Biol.* **12**(3), 430–439 <https://doi.org/10.1046/j.1420-9101.1999.00058.x> (1999).
28. Ziehm, M. & Thornton, J. M. Unlocking the potential of survival data for model organisms through a new database and online analysis platform: SurvCurv. *Aging Cell* **12**(5), 910–916. <https://doi.org/10.1111/accel.12121> (2013).
29. Uno, H., Tian, L., Cai, T., Kohane, I. S. & Wei, L. J. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat. Med.* **32**(14), 2430–2442. <https://doi.org/10.1002/sim.5647> (2013).
30. Amrhein, V., Greenland, S. & McShane, B. Scientists rise up against statistical significance. *Nature* **567**(7748), 305–307. <https://doi.org/10.1038/d41586-019-00857-9> (2019).
31. McShane, B. B., Gal, D., Gelman, A., Robert, C. & Tackett, J. L. Abandon Statistical Significance. *Am. Stat.* **73**(sup1), 235–245. <https://doi.org/10.1080/00031305.2018.1527253> (2019).
32. Wasserstein, R. L., Schirm, A. L. & Lazar, N. A. Moving to a World Beyond ‘ $p < 0.05$ ’. *Am. Stat.* **73**(sup1), 1–19. <https://doi.org/10.1080/00031305.2019.1583913> (2019).
33. Hayat, M. J. et al. Moving nursing beyond  $p < 0.05$ . *Res. Nurs. Health* **42**(4), 244–245. <https://doi.org/10.1002/nur.21954> (2019).
34. Erickson, R. A. & Rattner, B. A. Moving Beyond  $p < 0.05$  in Ecotoxicology: A Guide for Practitioners. *Environ. Toxicol. Chem.* **39**(9), 1657–1669. <https://doi.org/10.1002/etc.4800> (2020).
35. G. Campitelli, “Retiring Statistical Significance from Psychology and Expertise Research,” vol. 2, 2019.
36. A. Ciapponi, J. M. Belizán, G. Piaggio, and S. Yaya, “There is life beyond the statistical significance,” *Reprod. Health* **18**(1), pp. 80, s12978–021–01131-w <https://doi.org/10.1186/s12978-021-01131-w> (2021).
37. G. Shafer, “Testing by Betting: A Strategy for Statistical and Scientific Communication,” *J. R. Stat. Soc. Ser. A Stat. Soc.* **184**(2), 407–431 <https://doi.org/10.1111/rssa.12647> (2021).
38. Poupet, C. et al. *Lactobacillus rhamnosus* Lcr35 as an effective treatment for preventing *Candida albicans* infection in the invertebrate model *Caenorhabditis elegans*: First mechanistic insights. *PLOS ONE* **14**(11), e0216184. <https://doi.org/10.1371/journal.pone.0216184> (2019).
39. Poupet, C. et al. Curative Treatment of Candidiasis by the Live Biotherapeutic Microorganism *Lactobacillus rhamnosus* Lcr35<sup>+</sup> in the Invertebrate Model *Caenorhabditis elegans*: First Mechanistic Insights. *Microorganisms* **8**(1), 34. <https://doi.org/10.3390/microorganisms8010034> (2019).
40. Cardin, G. et al. Development of an innovative methodology combining chemical fractionation and in vivo analysis to investigate the biological properties of cheese. *PLOS ONE* **15**(11), e0242370. <https://doi.org/10.1371/journal.pone.0242370> (2020).
41. K. Kamary, K. Mengersen, C. P. Robert, and J. Rousseau, “Testing hypotheses via a mixture estimation model,” 2014, <https://doi.org/10.48550/ARXIV.1412.2044>.
42. De Barros, P. P. et al. *Lactobacillus paracasei* 28.4 reduces in vitro hyphae formation of *Candida albicans* and prevents the filamentation in an experimental model of *Caenorhabditis elegans*. *Microb. Pathog.* **117**, 80–87. <https://doi.org/10.1016/j.micpath.2018.02.019> (2018).
43. M. C. Meyer, “Inference using shape-restricted regression splines,” *Ann. Appl. Stat.*, vol. 2, no. 3, Sep. 2008, <https://doi.org/10.1214/08-AOAS167>.
44. C. Bouveyron, E. Côme, and J. Jacques, “The discriminative functional mixture model for a comparative analysis of bike sharing systems,” *Ann. Appl. Stat.* **9**(4), <https://doi.org/10.1214/15-AOAS861> (2015).



45. Biernacki, C., Celeux, G. & Govaert, G. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(7), 719–725. <https://doi.org/10.1109/34.865189> (2000).

## Acknowledgements

Some strains were provided by the CGC, which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440).

## Author contributions

PMG: developed the general strategy and implemented the programming and wrote the paper. CP: provided datasets, has participated in the emergence of the problematic, interpreted the results and wrote the paper. EC: developed the general strategy and implemented the programming. MB: has participated in the emergence of the problematic, interpreted the results and wrote the paper. PV: has participated in the emergence of the problematic, provided critical feedback and wrote the paper. SB: has participated in the emergence of the problematic, supervised the project and provided critical feedback.

## Funding

This work was supported by the Emergence Program, I-Site Clermont, Clermont Auvergne Project CAP 2025.

## Declaration

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to P.-M.G. or C.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024