ethology *international journal of behavioural biology* **WILEY**

Fully flexible analysis of behavioural sequences based on parametric survival models with frailties—A tutorial

Lorenz Gygax[1] 🔗 | **Yvonne R. A. Zeeland[2]** | **Christina Rufener[3]** 🔗

[1]Faculty of Life Sciences, Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences, Animal Husbandry & Ethology, Humboldt-Universität zu Berlin, Berlin, Germany

[2]Department of Clinical Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

[3]Department of Animal Science, Center of Animal Welfare, University of California, Davis, Davis, USA

Correspondence
Lorenz Gygax, Humboldt-Universität zu Berlin, Faculty of Life Sciences, Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences, Animal Husbandry & Ethology, Unter den Linden 6, 10099 Berlin, Germany, Office: Philippstraße 13 (Haus 10), 10115 Berlin, Germany.
Email: lorenz.gygax@hu-berlin.de

Present address
Christina Rufener, Centre for Proper Housing of Ruminants and Pigs, Federal Food Safety and Veterinary Office (FSVO), Agroscope Tänikon, Ettenhausen, Switzerland

Funding information
Ministerie van Economische Zaken en Klimaat; Bundesamt für Lebensmittelsicherheit und Veterinärwesen

Abstract

Recent automated systems allow collecting continuous data on individual animals with high accuracy over a long time. During this time, animals can be traced across different (discrete) types of behavioural states, with the duration in each state being known. Nevertheless, analyses of such sequences of states or behaviours may prove difficult. Classic Markov-chain methods have limitations in respect to incorporating "memory" (effects of past states), the duration in the states and accounting for dependencies. Dependencies occur in many data sets, where, for example a variety of individuals from different groups are observed and/or when an experiment is divided in different crossover treatment phases. So-called parametric survival analysis with frailties can incorporate aforementioned aspects in one coherent model. The time spent in a specific state (performing a specific behaviour) can be modelled in dependence of the subsequent state (transition probabilities) while incorporating how these transitions are influenced by experimental treatments. In addition, prior states can be used as predictor variables (accounting for past behaviour). Finally, random effects can be included to account for dependencies according to, for example individual identity, group/farm/laboratory or experimental period. Using interactions between random and fixed effects, the within- and between-subject variability of the transition probabilities can be estimated to indicate variation between and consistency within individual subjects (individuality and personality). Moreover, relative hazards describing transitions from one state to several potential follow-up states can be estimated. Behavioural sequences and their modulation by experimental situations can be studied accordingly. Using two exemplary data sets, the data type and structure adequate for parametric survival analysis are introduced and advice is given on how to specify and run such models. Overall, parametric survival analysis with frailties presents a modern and versatile approach that can revive sequential analysis. This will facilitate more detailed use of behavioural data and accordingly detect more subtle aspects of behaviour.

KEYWORDS
brms, fixed effects, R, random effects, sequential behavioural analysis, STAN

1 | INTRODUCTION

Behaviour occurs in sequences, and every single behaviour is related to other behaviours. The sequence of behaviours reflects moment-to-moment decisions by animals and is, therefore, a potential approach to studying everyday decision-making and understanding behavioural control mechanisms (e.g., Grafen, 2002; Gygax, 2017). Accordingly, incorporating the sequence of different behavioural states in behavioural analyses would be crucial to reflect animal behaviour appropriately. Yet, the application of "sequential analyses" is quite rare (Asher et al., 2009). Specifically with the increased use of automated data collection methods (e.g., Leos-Barajas et al., 2017; Rufener et al., 2018), data on longer behavioural sequences of individual animals are available that are more precisely measured. Current approaches that summarise this sequence data, for example analysing the proportion of time spent performing specific behaviours, do not make full use of this type of data because they ignore the durations of single bouts and the sequence of the behaviours. Therefore, methods that deal with this data adequately are needed more urgently than ever and the field of statistics has, of course, progressed since the time of, for example Haccou and Meelis (1992). In their book, they introduced a series of methods to deal with behavioural data based on "time-structured" models. These rely mostly on Markov-Chain models (see also next section in respect to limitations of these models) and seem to have been rarely applied (see section on applications). All in all, we think it is timely to have another look at what modern statistical approaches can offer in respect to analysing behavioural sequences, what questions they might answer and how such analyses can be implemented in animal behaviour research. Our tutorial gives an overview of methodological approaches, summarises applications of sequential analysis in the past and its potential in the future, and outlines some basic modelling considerations before providing two examples of a fully flexible analysis of real data sets using multi-state parametric survival models.

2 | METHODOLOGICAL APPROACHES

Classical approaches to deal with sequential data have mainly focused on methods based on Markov-Chains (e.g., Berchtold & Sackett, 2002; Ivanouw, 2007; Macdonald & Raubenheimer, 1995). At least in their basic application, these methods have several restrictions that are either unrealistic or impractical for the evaluation of data from behavioural experiments (Helske & Helske, 2019). First, the main assumption of Markov-Chains is the absence of a "memory"; that is the current state alone (and no further states in the [recent] past) determines the transition probabilities to the next following states (type of behaviour). With respect to behaviour organised in longer chains within a given context (Casarrubea et al., 2015), this seems an unreasonable assumption. For instance, the probability that a sleeping animal will start to drink after waking up most likely depends on the thirst of the animal, that is whether the animal drank

before sleeping. Moreover, the duration in a given state may be an important aspect in addition to the type itself (Metz et al., 1983). Finally, it is difficult to reflect experimental paradigms including dependencies due to repeated measurements of individuals, crossover designs with phases involving different treatments or hierarchical structuring (e.g., observing individuals in different groups; Arnqvist, 2020) in a Markov-Chain model.

Whereas Markov-Chain models have a heavy focus on the type of state (i.e., behaviour A vs. behaviour B), another class of models has an additional focus on the duration of each state (duration of a behavioural bout). The classical approaches to deal with durations until an event are survival models. The basic implementation of these models deals with the time until the occurrence of a single event ("alive" until death occurs, hence the name survival analysis). These models can be extended to accommodate several states, in which the duration in a specific state is modelled until a given other state occurs. Technically, they include a mandatory explanatory variable defining the type of transition (current state to follow-up state, e.g., "alive to sick" vs. "sick to dead" vs. "alive to dead") to start with.

As so often, the parametric varieties of such survival models are simpler to estimate and are more flexible in their application. Accordingly, multi-state parametric survival models can overcome the limitations mentioned for the Markov-Chain models. In addition to incorporating the multiple states, they provide the flexibility of linear models in respect to the inclusion of predictors (fixed effects) such as not only treatment variables, but also prior states. In addition, they allow for the inclusion of random effects reflecting the experimental design. Such random effects are classically called "clustering" and their implementation "frailties" in the jargon of survival modelling (e.g., Duchateau & Janssen, 2008; Munda et al., 2012). In short, parametric survival models can be considered as a special case of generalised linear-mixed-effects models with all their possibilities, for example using error distributions other than the Gaussian.

Here, for instance, the duration of time until an animal transitions from one behaviour to another is expected to be right-skewed as most behaviours will be performed for a short time only before the next behaviour in a sequence follows (and as is expected for durations more generally). As the term "parametric" suggests already, these models make relatively strong assumptions on the distributions of the frailties (random effects) and the distribution of event times (the durations in the different states). At the same time, simpler estimation techniques can be used in comparison with, for example semi-parametric models (e.g., Munda et al., 2012), such that even quite complex models become numerically tractable.

The raw data for these models basically consist in the durations that animals spend in the different states given a specific follow-up state («duration of stay»). The parametric assumptions then imply what the shape of the distribution of these durations for each possible type of transition (e.g., behaviour A to behaviour B vs behaviour A to behaviour C) is like. Moreover, the survival curve or the hazard curve for each possible transition can be directly deduced based on

WILEY

the parameters describing the distribution of the durations (see example 1).

The survival curve describes the proportion of subjects that have and have not performed a specific transition after a given elapsed time. For instance, a certain proportion of animals will have transitioned to feeding after 2 h of sleep. The remaining animals will either remain sleeping or have transitioned to another behaviour. The hazard curve indicates the "risk" to perform a specific transition at a given elapsed time; that is how likely an animal is to switch from sleeping to feeding at any given time since the animal went to sleep; for example this likeliness may increase the longer the animal has been in the state "sleeping". Finally, two hazard curves can be related to each other resulting in relative hazard curves: A sleeping animal might be more likely to start feeding than to start resting again after waking up, at least if it had been in a sleeping state for a certain period of time.

In short, duration of stay, survival curve and hazard curve are all given by the estimated parameters in parametric survival models, and the expressions can be used interchangeably in this sense: All are illustrations of the same estimated parameters, and accordingly, it does not matter mathematically whether one focuses on the duration in a state, the survival curve or the hazard curve. This is because each of these can be deduced based on the parameter(s) of the chosen error distribution; that is one can be converted in the other (see example 1).

Given that parametric survival models estimate these relevant aspects for all the different types of transitions, the complete and most likely behavioural sequences can be deduced and interpreted based on these models (see next section). Applied to our theoretic example, parametric survival models can estimate (a) whether a sleeping animal is indeed more likely to feed than to perform other behaviours after waking up, (b) how long it takes to transition from one behavioural state to a specific follow-up state and (c) whether previous states (e.g. did the animal feed before sleeping?) or treatment effects (e.g. is the animal housed in a restrictive environment) affect behavioural sequences.

3 | APPLICATIONS IN THE PAST AND POSSIBILITIES FOR THE FUTURE

Before we show some past applications of sequential analyses and sketch some potential questions that can be addressed in the future, we would like to point out two formal issues. In the literature on sequences, the term "state" is often used for the different conditions that subjects find themselves in. Here, we use the term synonymously with different types of behaviour. We use the term behaviour in a loose sense, too, and also consider staying in specific locations as a behaviour. Given the wide use of terms like "analyses of sequence" in the prolific field of molecular genetics on the one hand and the few and patchy citations of seminal papers on analyses of behavioural sequences (such as e.g., Haccou & Meelis, 1992) on the other hand, it is no easy task

to find past applications of the analysis of behavioural sequences in the literature.

If behavioural sequences are observed in a relatively undisturbed ("feral") situation, the result of a multi-state parametric survival model can be viewed as a basic description of the behavioural mechanism at work in a given species. This may allow understanding the overall behavioural organisation in species more complex than those considered so far (mostly invertebrates, e.g., leech behaviour by Garcia-Perez et al., 2005). This complete information on behavioural sequences can moreover be compared between two or more (experimental) situations. Looking at such a model in more detail, one can observe whether, for example only single transitions have been affected by the treatments or whether the duration of a specific behaviour (independent of the possible follow-up behaviours) has changed.

Whereas the survival curves may have a more illustrative character and may serve in the assessment of a model (see example 1), the hazard curves lay the basis for more detailed assessments of behavioural transitions. Approaches using this view have previously helped in classifying the context of behaviour in Hector's dolphins (Slooten, 1994), investigating conflict resolution (Egge et al., 2011) and predation risk in stalk-eyed flies (Worthington & Swallow, 2011), contest duration in spiders (Moya-Laraño & Wise, 2000), the sequential organisation of nightingale songs (Ivanitskii et al., 2017) and the transition between different phases of training and the endpoints of either qualifying as a guide dog or withdrawal from training (Asher et al., 2017).

As we suggest to accommodate dependencies in the data set based on random effects, these random effects can also be used to assess repeatability within individuals (Nakagawa & Schielzeth, 2010) and, thus, contribute to the uprising field of animal personality (e.g. Carter et al., 2013; Dall & Griffith, 2014; Sih et al., 2004a, 2004b). For instance, we can specifically ask whether only some behavioural transitions seem to be part of the animals' personality as indicated by large within-individual repeatability.

Given that the field of sequential analysis is at least 30 years old, the citations in the paragraphs above are relatively scarce. This illustrates the lack of use of the information available from behavioural sequences and indicates that many researchers studying animal behaviour are not equipped with the necessary tools to investigate behavioural sequences comprehensively. Multi-state parametric models can provide in-depth information about behavioural sequences and reflect the potential of their future application in animal behaviour research.

4 | BASIC PRACTICAL MODELLING CONSIDERATIONS

The type of analysis addressed here can be applied if sequences of different types of (behavioural) states and the durations spent in each state are recorded ("bout" durations). Accordingly, the current state, the time duration spent in the current state, whether this

duration is censored, and the type of transition as defined by the current and the subsequent state are the indispensable information for a multi-state parametric survival model (Jackson, 2016; Putter et al., 2007). With respect to these variables, a few modelling considerations have to be taken into account.

First, the type of transition is used as a factor variable with a different level for each possible transition and included as a fixed effect. Accordingly, this variable can be generated by combining the current and the subsequent state (e.g., for the transition from behaviour A to behaviour B: AB). If transitions are possible among all states, the main effects plus the statistical interaction of the current and subsequent state can alternatively be used to specify the type of transition. The information on the type of transition can be complemented by additional predictors such as variables reflecting the experimental design (random effects) and by further fixed effects that reflect, for instance, the experimental treatment and the history of states (former states and/or the time spent in those former states). These fixed effects are thought to modulate the effect of the variable "type of transition" on the duration in a given state. Therefore, the additional fixed effects will usually occur in a statistical interaction with the variable type of transition.

Second, the duration spent in a state can be either censored or uncensored. Censoring means that the subject in a specific state under consideration was observed, but the event of interest has not occurred; that is the transition to the next state did not occur during the observation time. The observed duration is, therefore, censored and provides only a minimal estimate of the duration in a given state (e.g., Bressers et al., 1991). In multi-state data sets such as continuous behavioural sequences, many or even all transitions that are considered are usually observed as a subject moves from one state to another. Therefore, the observation of the time spent in a given state is mostly uncensored because the state has been observed until the next state was reached. The last recorded state in an observation period may be an exception because the next following state is not observed, for example the last behaviour on an observation day. These (partly censored) durations are then used as the outcome in a statistical model.

Modelling durations that are spent in specific states (survival times for the specific state transitions) need special statistical treatment because durations do not usually follow a normal distribution. As a result, the distribution of such durations is normally right-skewed (has a long right tail) and is sometimes called the "shape of the baseline hazard".

Third, to implement the multi-state parametric survival model, we use the package *brms* (Bürkner, 2017, 2018) in R Versions 3.6.3, 4.0.2 and 4.0.3 (R Core Team, 2020) here. It provides all aspects listed above, specifically the flexible use of fixed and random effects as well as some special (error) distributions needed to model durations. The Weibull, the Gamma, the univariate lognormal and the exponential distribution are currently implemented in brms and can reflect the expected right-skewed distributions of the durations. The package provides an interface to STAN (Carpenter et al., 2017)
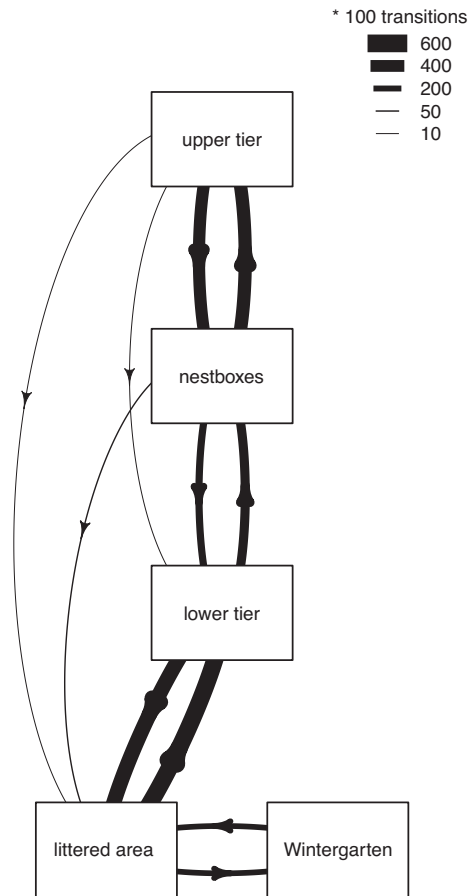
that allows a modern Bayesian estimation of such complex models using an MCMC (Markov-Chain Monte Carlo) method for parameter estimation. Moreover, *brms* uses an R syntax very similar to one of the simpler mixed-model approaches as, for example implemented in *nlme* (Pinheiro et al., 2019) or *lme4* (Bates et al., 2015). Therefore, its syntax looks easily recognisable to readers who are familiar with those methods. See Supporting Information Text S1 for alternative R packages and more details on *brms*.

Data and R code for running the examples below are available separately for each example in the Supporting Informations section of the online version of this article (Data S1 and S2, Code S1 and S2).

5 | EXAMPLE 1: INDIVIDUAL MOVEMENTS OF HENS IN AVIARIES

In this first example, we would like to show how a multi-state parametric survival model is set up, how specific behavioural transitions can be compared based on the hazard curve and the hazard ratio, and how random effects can be used to assess repeatability in the context of animal personality. At the same time, we address some more practical issues that arise when dealing with a real data set.

Here, we use a data set that recorded the movements of individual hens (*Gallus gallus domesticus*) between different zones in aviaries. Rufener et al. (2018) describe the type of data in detail and the data set re-used here is the one used in Rufener et al. (2019). In short, the hens could move from an outside area ("wintergarden", zone 1) to a littered area within the barn on the ground level (zone 2), and up into the aviary to the lower tier (zone 3), the nest boxes (zone 4) and the upper tier (zone 5; Figure 1). Twenty focal hens were observed in six groups of 225 hens each. In half of the pens, the focal hens were white among brown hens, in the other half of the pens, the focal hens were brown among white hens (for easing additional visual observations). All focal hens were observed using an automated tracking system for six consecutive days at 11 time points throughout their laying period, and at any one time, half of the hens were observed simultaneously (for details see Table 1 in Supporting Information Text S1 and also on how we dealt with "flickering"). As in Rufener et al. (2019), we only use the data from 30 min before lights on (01:30) until 30 min after lights off (17:30). More precisely, this means that we included the data from the first zone change that occurred after 01:30 h until the last zone change that occurred before 17:30 to avoid censoring for those transitions that were observed on a given day. In addition, we imputed the unobserved but potentially possible transitions as being censored using the total duration of the complete observations for this hen on each given day (the hen was observed for this duration of time without these censored transitions occurring). For instance, if a hen did not move from the nest boxes (zone 4) to the upper tier (zone 5) on a given day, the transition "4–5" was manually imputed as a single data point with the duration of observation for this given hen on this given day and marked as being censored

FIGURE 1 Observed transitions between the different states reflecting different zones in an aviary in the hen example. For the thickness of the arrows, all observed transitions across all observations were summed. Numbers give the overall number of transitions in 100 s

stay duration was reflected in full seconds in the original records, some final stay durations were rounded to 0 automatically. This was a problem for running the model because the available error distributions all include link functions that involve some form of calculating the logarithm, which is not possible for zeros. In accordance with a "first aid" transformation for counts, we added 0.5 s to all stay times in order to avoid zero values for statistical evaluation (see model code below). Some recordings showed transitions within the zones, that is starting from one zone and ending in the same zone. In these cases, the durations of stay in a given zone were summed and a single stay in the zone with the summed duration was kept in the data. On the way down the aviary, hens sometimes jumped or fell across zones (Figure 1). Therefore, transitions across several tiers were possible and were kept in the data set.

The following variables were kept in the data set for evaluation and are available in the Supporting Information Data S1:

penID	an identifier for each of the six pens;
henID	an identifier for each of the 120 hens (nested in the pens);
henDate	an identifier for each observation day for each hen (nested in hens);
date	calendar date of the recording days (crossed with all the above);
time.in	the time of day when the hen entered the zone;
secSMN	time.in as the number of seconds since midnight;
stay	the duration of stay in the given zone in seconds;
cens	a variable indicating whether the observed duration was censored;
zone	the zone of residence;
zone.next	the zone that followed the given observation;
type.trans	the combination of the current and the next zone, that is the possible transitions;
zone.prev	the previously visited zone;
stay.prev	the duration of stay in the previously visited zone;
Focal	colour of the focal animals;
ageInDay	hen age [days];
Week	observational week (1-11).

All codes necessary to produce the following output are available in the Supporting Information Code S1. In Table 1, the summary of this data illustrates the size of the data set as well as the structure and the data types as described in the list above.

Based on this data, we set up an initial model that was expanded in a few follow-up steps. As the hazard function, we chose one of the somewhat more flexible two-parameter families, the Weibull, as an example. The Weibull distribution can reflect the expected right-skewed distribution appropriately (see above).

In respect to the random effects, we followed a pragmatic approach in that we kept the random effect simple in this initial model using an intercept-only random effect (see also Supporting

(the transition was not observed). The possible transitions are those that are visualised in Figure 1. If we had not included these censored observations, no information would be available for the unobserved transitions for a given hen and day. Yet, we do actually know that these transitions did not occur during the observation period of that day. Therefore, using the possibility to include censored data adds this available information to the statistical model. Omitting these censored data points would potentially lead to an underestimation of the durations because the information that these transitions did not occur in the length of an observation day would not be available to the model. Similarly, not treating these observations as censored would lead to the same effect but to a lesser extent.

When moving up the aviary, the hens could not jump zones and had to move from the littered area via the lower tier and the nest boxes to the upper tier. Similarly, when the hens moved to or from the wintergarden, they had to cross the littered area. There were some instances in the raw data where not all these transitions were recorded. In those cases, short stays in the zones that the hens needed to pass through were imputed. Because these imputations were shorter than 1 s in certain instances and the

Information Text S1). In a second step, we subsequently extended this random effect to address a specific question related to this data set. Nevertheless, we included the full random structure reflecting the experimental design including `penID`, `henID`, `henDate` and `date` starting with this initial starting with this initial model. Though the first three random effects were nested and all crossed with calendar date, they could be specified in the model formula independently because for instance, the identifier for hens was unique and not re-used in the different pens. In this case, the nesting structure is implicitly defined by the fact that each hen identifier was only observed in one pen. Calendar date accounted for the fact that half of the hens were observed on exactly the same days, and the observations were in this sense dependent. This accounted also for the fact that events in the barn may have affected all hens similarly on a given day. In the initial model, we used the simplest fixed effect possible, namely the type of transition as the sole fixed effect. All these considerations resulted in the following first model (Table 2, model 1).

`stay + 0.5` defined the outcome variable of this model and `| cens ()` defined the variable that indicated which observations

were censored. On the last line, some technical aspects were defined. `cores= 4` indicated the number of CPUs available and allowed running four MCMC chains in parallel (four chains was the default setting and, therefore, did not need to be specified). Here, a chain each was run on the four cores. It is shown explicitly (though this is the default also) that a total number of 2000 iterations were ran per MCMC chain, and that the first 1,000 iterations were not considered in the evaluation (warm-up). The seed (with no default value) indicated the starting point for the internal random number generator and allows the exact replication of the model (internally, random numbers are used by the method).

This model ran for about 72 h on a Windows 10, 64 bit computer with 32 GB RAM and with 4 Intel i5-4570, 3.2 GHz CPU cores (one chain per core using the parallelisation capabilities of *brms/STAN* as visible in the model code). The run time was so long presumably due to the size of this data set and could potentially be reduced by implementing the model directly in STAN and/or by using more informative priors. Due to the additional complexity involved in both these potential solutions, we did not attempt to do so here.

TABLE 1 Structure and size of data set of example 1, individual movements of hens in aviaries. See text for further explanations

```
str (hens.df)
'data.frame':     319484 obs. of  16 variables:
 $ penID     : Factor w/ 6 levels "A.1","A.2","A.3",..: 1 1 1 1 1 1 1 1 ...
 $ henID     : Factor w/ 120 levels "A.1","A.10","A.11",..: 1 1 1 1 1 1 ...
 $ henDate   : Factor w/ 5705 levels "A.1 01.06.2017",..: 1 1 1 1 1 1 1 ...
 $ date      : Factor w/ 132 levels "01.06.2017","01.11.2016",..: 1 1 1 ...
 $ time.in   : POSIXct, format: "2017-06-01 04:02:05" ...
 $ secSMN    : int  14525 14792 15282 15578 16665 16857 17011 17708 ...
 $ stay      : int  267 490 296 1087 192 154 697 41 170 489 ...
 $ cens      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ zone      : Factor w/ 5 levels "5","4","3","2",..: 2 1 2 1 2 1 2 1 ...
 $ zone.next : Factor w/ 5 levels "5","4","3","2",..: 1 2 1 2 1 2 1 2 ...
 $ type.trans: Factor w/ 11 levels "1 2","2 1","2 3",..: 8 11 8 11 8 11 ...
 $ zone.prev : Factor w/ 5 levels "5","4","3","2",..: NA 2 1 2 1 2 1 2 ...
 $ stay.prev : int  NA 267 490 296 1087 192 154 697 41 170 ...
 $ focal     : Factor w/ 2 levels "brown","white": 2 2 2 2 2 2 2 2 2 2 ...
 $ ageInDays : int  357 357 357 357 357 357 357 357 357 357 ...
 $ week      : int  9 9 9 9 9 9 9 9 9 ...
```

TABLE 2 R code for the different models for example 1, individual movements of hens in aviaries. See text for further explanations

No.	R model code
1	```hens.initial.brm <- brm ((stay+0.5)	cens (cens) ~ type.trans +` ` (1 | penID) + (1 | henID) + (1 | henDate) +` ` (1 | date),` ` hens.df, family= weibull (),` ` cores= 4, warmup= 1000, iter= 2000, seed= 5781)```
2	```hens.indiv.brm <- brm ((stay+0.5)	cens (cens) ~ -1 + type.trans +` ` (1 | penID) + (-1 + type.trans | henID) +` ` (-1 + type.trans | henDate) +` ` (1 | date),` ` hens.df, family= weibull (),` ` cores= 4, warmup= 1000, iter= 2000, seed= 5781)```
3	```hens.extFix.brm <- brm ((stay+0.5)	cens (cens) ~ focal * type.trans +` ` s (ageInDays, by= type.trans) +` ` (1 | penID) + (1 | henID) + (1 | henDate) +` ` (1 | date),` ` data= hens.df, family= weibull (),` ` cores= 4, warmup= 5, iter= 10, seed= 5612)```
4	```hens.shape.brm <- brm (brmsformula ((stay+0.5)	cens (cens) ~ -1 + type.trans +` ` (1 | penID) + (-1 + type.trans | henID) +` ` (-1 + type.trans | henDate) +` ` (1 | date),` ` shape ~ -1 + type.trans),` ` data= hens.df, family= weibull (),` ` cores= 4, warmup= 5, iter= 10, seed= 6417)```

A model calculated based on an MCMC estimation needs to be checked in respect to convergence. Because we use a parametric model here, distributional assumptions should additionally be checked (see Supporting Information Text S1 for some general remarks on how these checks work and some specific information for the current model). In the current case, convergence and distributional assumptions seemed to have been met reasonably.

Having done these checks, we could turn to the interpretation of this model for which the package *tidybayes* (Kay, 2020) provided some useful tools Oct 27. In survival analysis, the duration of stay in the different states is basically modelled. Therefore, a meaningful figure was the comparison of the raw durations observed with the estimated duration (Figure 2). The relative differences in the durations that hens stayed in a specific zone given where they go next looked similar in the raw data and the model estimates. The much longer estimates for the transitions nest to litter (4 → 2), upper tier to litter (5 → 2) and upper tier to lower tier (5 → 3) can be explained by the fact that many of these observations were censored (e.g., medians of 5 → 2, 5 → 3 are at the maximally observed duration). However, this censoring was not visualised directly in the raw data. This shift between the relatively shorter raw data and the longer estimation of these durations (including the information of censoring) is the exact reason why specific methods should be applied if there are censored observations in a data set. If this is not done, the durations of the observations are underestimated.
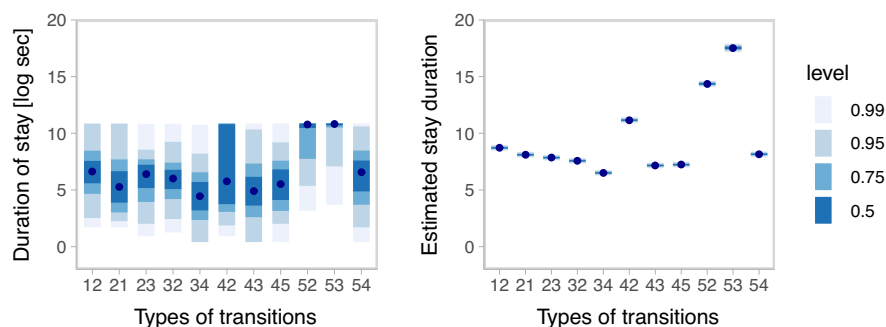
The model estimates directly reflect; that is they correspond to the parameters of the Weibull distribution. The estimates for the types of transitions correspond to the logarithm of what is usually called the "location" parameter of the Weibull distribution. In the standard case here, a common shape parameter was estimated for all types of transitions. Based on these location and shape parameters (as estimated in each sample of the MCMC chain), survival and hazard functions could be estimated for each transition type (and visualised based on their credible intervals, e.g., Figure 3). The

survival and hazard functions can be calculated based on the location (L) and shape (S) parameter of the Weibull distribution and the potential durations (X corresponding to the values on the X-axis in Figure 3) as follows Oct 27 (similar formulas are available for other error distributions):
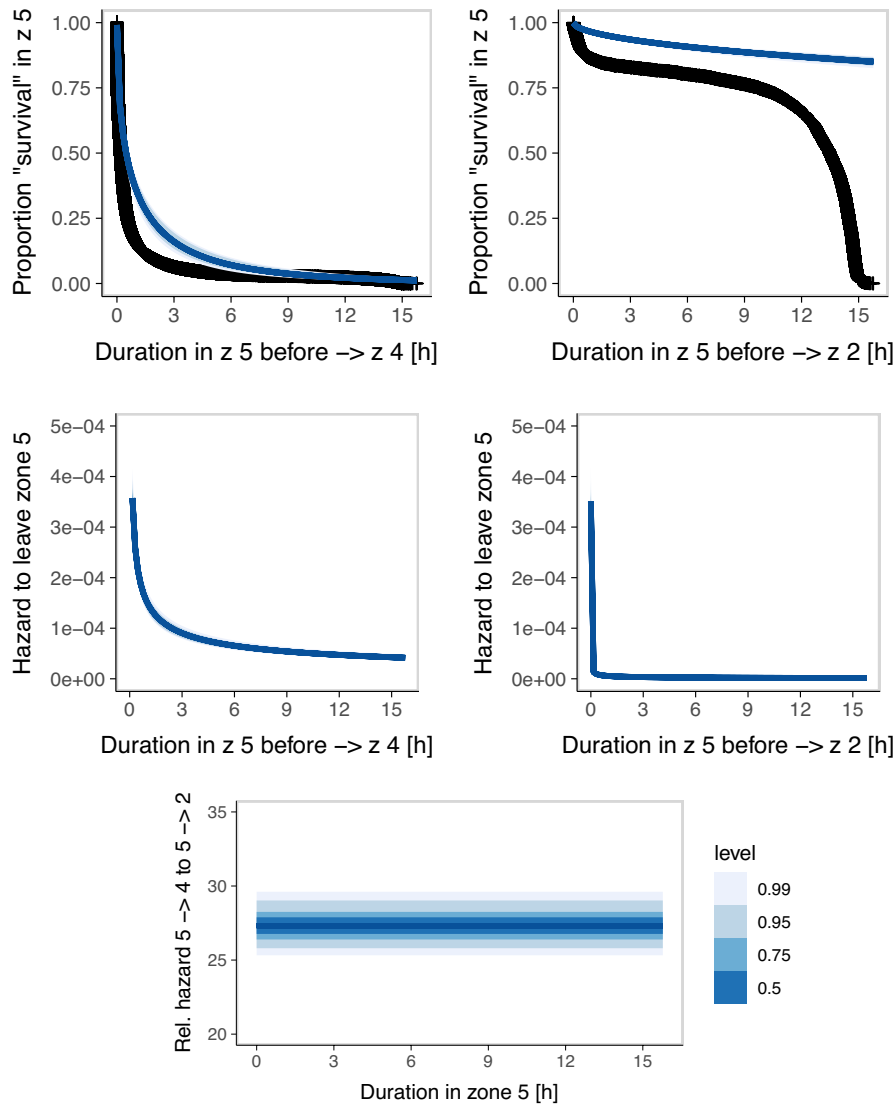
$$Survival = e^{-\left(\frac{X}{e^L}\right)^S} ; Hazard = \frac{1}{e^L} \times S \times \left(\frac{X}{e^L}\right)^{S-1}$$

Depending on the question of interest, relative hazard curves can also be drawn. The relative hazard is the ratio of two hazards and indicates the relative "risk" of one transition over another. For instance, the relative hazard to move to different states from a given one (e.g., 5 → 4 versus 5 → 2, that is the likelihood of a controlled descent to the level of the nest boxes from the top tier in comparison with the risk of falling to the littered area from the top tier). As an example, these visualisations are shown in Figure 3 for the transitions 5 → 4 versus 5 → 2. In the survival curves, the raw observations were included (in black) and it is visible that the parametric estimates (in blue) may not (yet) fit the data very well, specifically for the transition 5 → 2. Not only this is due to the large amount of censored data on the one hand, but also because only one single shape parameter was estimated for all the transition types. Specifically, the transition 5 → 2 seemed to have a different shape also in the survival curves for the different hens (see Supporting Information Figure S1). We will illustrate further down how the model can be extended to allow for an individual shape parameter for each transition type. Given the common shape parameter implemented so far, the general shape of the survival and hazard curve is the same for each type of transition and the relative hazard is, therefore, constant (Figure 3). Biologically, this indicates that a controlled descent is about 25 times more likely than a fall from zone 5 at any one time.

Next, we turned to a specific question raised by the original authors. The authors were interested in knowing more about the individuality of the hens based on this data set (Rufener et al., 2018).

FIGURE 2 Observed (left) and estimated (right; including credible intervals with several levels of coverage) stay durations of the hens depending on the current zone (first number) and where they moved next (second number). 1 = outside area ("wintergarden"), 2 = littered area within the barn on the ground level, 3 = lower tier, 4 = nest boxes, 5 = upper tier. The relative differences between the types of transitions are the same in both parts of the figures. The longer durations in the right panel partly reflect the fact that some of the observations were actually censored (which is not directly visible in the plot of the raw data). Different proportions of the data / estimates (levels) are reflected by the different shading. The point and the dark blue shading correspond to the median and the 50% interval, respectively, and to the box of a classical boxplot accordingly [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 3 Estimated survival (top) and hazard (middle) curves for the transitions 5 → 4 (left) and 5 → 2 (right) and the relative hazard of the 5 → 4 (controlled descent from the upper tier to the nest box level) versus the 5 → 2 transition (fall from the upper tier to the littered area; bottom). The observed survival curves are indicated in the top panels by the black line with small vertical lines reflecting censored observations. These observations are so dense that they are visible as a thick black line [Colour figure can be viewed at wileyonlinelibrary.com]

We have used the variance components estimated in this type of model to address the repeatability of the behaviour within individual hens. If between-hen variability is larger than within-hen variability, the notion of individuality is indeed supported. We can, therefore, ask whether the variability in stay durations for the different transition types varies more between hens or between days within hens. We generally advise to follow the recommendation to use sum contrasts for all factor variables (Levy, 2014) such that the intercept corresponds to an overall average effect and the effect of the different levels are expressed as the deviations from this average. In the current model with only one fixed effect, this was not necessary because we could force the model to directly estimate the stay durations for the different transition types if we omitted the intercept. This resulted in the extended model 2 (Table 2).

This model ran for roughly 16 days on the same computer as mentioned above. Two of the chains did clearly not converge in the number of iterations used and with the standard thinning of 1. For full convergence, both these numbers would possibly need to be increased (increasing accordingly the run time). Two of the chains

mixed well, and at least one of the other two chains approached the values of the well-mixed chains towards the end of the run for all estimated parameters. Therefore, the following exemplary evaluation is based on the two well-mixed chains only (and the model was not re-run). In this model, we have estimated the standard deviation of each transition type between and within each hen. Squaring this standard deviation allows calculation of the ratio of the between- to the within-hen variance (Figure 4). This ratio is clearly >1 for the transitions 5 → 4 and 4 → 5 (at around 3) and clearly <1 for all other transitions, indicating that an analysis of individuality in these hens should focus on the zone changes between the upper tier and the level of the nest boxes.

So far, we have treated this data set as if the hens' decisions to move from one zone to the next did not change with age (or time). Moreover, we have not explicitly considered that the two breeds observed may differ in a systematic way, which could, potentially, explain some part of the between-hen variation. Because only one breed was observed in a given pen, the de-facto number of replicates for the variable `focal` is 6 (pens). With this small sample size, strong

FIGURE 4 Ratio of the between-hen variance divided by the within-hen variance for the different types of transitions. Values above the value of one indicate some level of individuality (the between-hen variability is larger than the within-hen variability). Transitions between the zones 1 = outside area ("wintergarden"), 2 = littered area within the barn on the ground level, 3 = lower tier, 4 = nest boxes, 5 = upper tier [Colour figure can be viewed at wileyonlinelibrary.com]

conclusions in relation to a (main) effect of the type of hybrid cannot be drawn from this data set, and an inclusion of this variable should rather be viewed as an illustration of the potentials of the statistical approach presented here. To account for the time effect, either `ageInDay` or `week` can potentially be used as an additional fixed effect. Because the observational weeks do not have any pre-defined meaning for the hens and because it can be assumed that changes in behaviour will show a smooth development, a smooth function (spline) of `ageInDay` could be used here for the time course (as in generalised additive mixed-effects models; Young, 2016). These two additional fixed effects (`focal`, `ageInDay`) could be included in an interaction with the type of transition in a model sketched as model 3 (Table 2; the number of iterations was chosen such that it is just possible to show that the model is, in principal, running using this syntax).

With this model, we basically reflect the difference in the transition probabilities between the two breeds and how the probabilities changed across time. If we wanted to graphically illustrate this, we would need to show the change across time of the transition probabilities for 22 cases (11 types of transitions × 2 breeds). This is without considering any of the potential interactions between fixed and random effects and thus shows one of the problems with the approach followed here: A high number of parameters in the statistical model was estimated, and it is no trivial task to make all this information accessible in an easy and straightforward way.

The fixed effects in this model could even be further expanded with including information on, for example previous states (´memory´). In the simplest case, the fixed effect would be extended by `… * zone.prev * stay.prev + …`. In doing so, we would ask whether the transitions between zones also depended on which zone had been visited previously and for how long. This could reflect, for instance, upward or downward movements of the hens across several levels in the aviary. We leave this as a potential future avenue for research in this example.

So far, the shape parameter of the Weibull distribution has been considered fixed for all observations. However, the function `brm`

even allows to model a dependency of such a parameter that is usually considered fixed on one or several explanatory variables. The respective model is sketched as model 4 (Table 2).

To sum up, we have seen some of the flexibility in the parametric approach to survival analysis as can be implemented with the package *brms*. Whereas we have only superficially sketched out some of the possibilities, we had a deeper look into a central question that was asked based on this data set: the individuality of the hens. This question has been addressed in Rufener et al. (2018), starting out from some exemplary location graphs of individual hens (Figure 2 in Rufener et al., 2018) in which the hens had seemingly different patterns. Based on the additional evaluation presented here, we can conclude specifically that an analysis of individuality in these hens should focus on the zone changes between the upper tier and the level of the nest boxes, whereas for all other zone changes, the variability within hens was larger than between hens.

6 | EXAMPLE 2: MULTIPLE-CHOICE FOR ENRICHMENTS IN FERRETS

In this second example, we focus on how changes in experimental conditions can be included in the model, how they can be illustrated and interpreted. Moreover, we address issues that arise when past states are to be evaluated and the relationship between model complexity and the amount of data available.

Here, we use a data set from a study on ferrets (*Mustela putorius furo*; Reijgwart et al., 2016). In this study, individual ferrets lived in an experimental chamber system, where they could reach six different enrichments and a control from a corridor which served as their home base (where food, water and an opportunity to sleep were present). Chambers could be accessed by pushing open a weighted door, while returning back to the home base through a cat flap was free of charge. To asses the value of the different enrichments, the weight of the doors to the chambers was increased gradually from day to day. The raw data included each visit to the different resources and the time of day that the resource pens were entered and abandoned. This data set was converted into a format that reflected the sequence of visits and also included the length of stays in each chamber. Based on the distribution of the stay durations in the choice chamber (Figure 2 in Supporting Information Text S1), we considered stay times of less than 120 sec as only passing through the home corridor in order to reach a novel enrichment chamber. Therefore, these relatively short visits were omitted in the analysis. The 120 sec is a rough estimate and more sophisticated means to decide on this boundary could potentially be followed. Given the omission of passing through events from the analysis, the animals could reach any chamber from any other chamber; that is all transitions between the different chambers were theoretical possible (Figure 5). The sole exception was visiting the home corridor directly after being in the home corridor because every visit to a resource chamber was deemed meaningful independent of its duration. Only observation days up to a door weight of 2000 g were included because
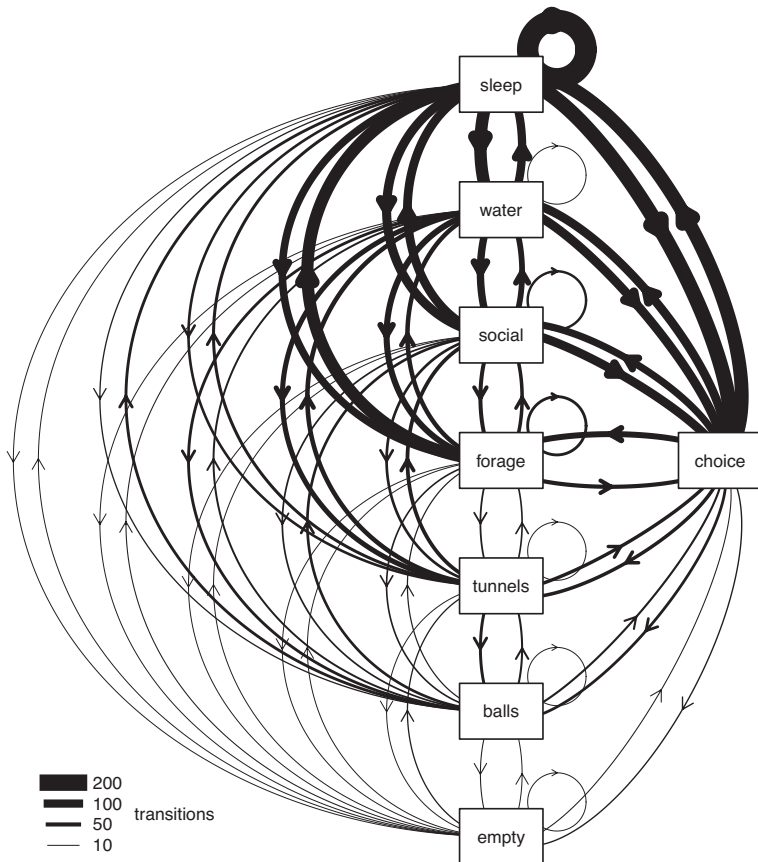
FIGURE 5 Observed transitions between the different states in the ferret example. For the thickness of the arrows, all observed transitions across all observations were summed

TABLE 3 Structure and size of data set of example 2, multiple choice for enrichments in ferrets. See text for further explanations

```
str (ferrets.df)
'data.frame':     6641 obs. of  12 variables:
 $ ferrID   : Factor w/ 7 levels "fret1","fret2",..: 1 1 1 1 1 1 1 1 1 ...
 $ ferrDays : Factor w/ 76 levels "fret1 1","fret1 10",..: 1 1 1 1 1 1 ...
 $ stay     : num  56 15 7 15 121 19 49 15 16 7 ...
 $ cens     : num  0 0 0 0 0 0 0 0 0 ...
 $ context  : Factor w/ 8 levels "balls","choice",..: 8 6 5 8 5 5 4 6 7
...
 $ loc.next : Factor w/ 8 levels "balls","choice",..: 6 5 8 5 5 4 6 7 3
...
 $ type.trans: Factor w/ 63 levels "balls balls",..: 61 44 39 60 36 35 ...
 $ weight   : int  0 0 0 0 0 0 0 0 0 ...
 $ loc.prev : Factor w/ 8 levels "balls","choice",..: 2 8 6 5 8 5 5 4 6
...
 $ dur.prev : num  NA 56 15 7 15 121 19 49 15 16 ...
 $ loc.snd  : Factor w/ 8 levels "balls","choice",..: NA 2 8 6 5 8 5 5 ...
 $ dur.snd  : num  NA NA 56 15 7 15 121 19 49 15 ...
```

TABLE 4 R code for the different models for example 2, multiple choice for enrichments in ferrets. See text for further explanations

No.	R model code
1	ferrets.initial.brm <- brm (stay | cens (cens) ~ type.trans + s (weight, by= type.trans) + (type.trans | ferrID) + (1 | ferrDays), ferrets.df, family= weibull (), cores= 4, warmup= 5, iter= 10, seed= 9167)
2	ferrets.initial.brm <- brm (stay | cens (cens) ~ type.trans * weight + (1 | ferrID) + (1 | ferrDays), ferrets.df, family= weibull (), cores= 4, warmup= 1000, iter= 2000, seed= 7819)

the ferrets hardly visited any resource cages if weights were heavier, and therefore, only very few transitions could be observed on those days (Table 2 in Supporting Information Text S1). On any given day,

all non-observed but potentially possible transitions from one chamber to another were included as censored observations using the total observation time of a given animal on a given day (see below).

The data set for evaluation contains the following variables and is available in the Supporting Information Data S2:

ferrID	an identifier for each ferret;
ferrDays	an identifier for each observation day of a given ferret (nested in ferret);
stay	the duration of time spent in the different chambers;
cens	an indicator of whether the stay time was censored;
context	the type of the current chamber;
loc.next	the type of the chamber that was visited next;
type.trans	the type of transition, that is the combination of the current context and the context of the next chamber;
weight	the weight of the doors on a given experimental day;
loc.prev	the chamber that had been visited previously;
dur.prev	the duration of this previous visit;
loc.snd	the second to last type of chamber visited;
dur.snd	the duration of this visit.

All the codes for the following evaluations are available in the Supporting Information Code S2. A first summary of the size of the data set and the variables is given in Table 3.

One problem with including former states as potential predictors becomes immediately visible here: The former states are unknown for the first state on each day and for all the unobserved transitions on a given day. These transitions could not be observed, and therefore, the state prior to these transitions could not be observed, either. All observations with unknown prior state are lost if the effect of past states was of interest (or an imputation scheme would need to be followed).

As with the hen data, we did not attempt to develop the best possible model for this data set here, but would like to illustrate some additional possibilities in applying parametric survival modelling. Therefore, the model development is not described step by step and we refer the readers to the recommendations and directions given above in respect to checks that need to be performed once a model has been calculated.

Based on a-priori considerations, a first model on this ferret data could include the transition type and the weight of the doors as fixed effects because we wanted to know whether adding weight changed the transition probabilities of the visits to the different chambers. The model was complemented with the random effects of the ferret days nested in the ferrets. Because only one ferret was tested at any one time, no crossed effect for the different test days was needed. In addition, one could include the interaction between the fixed effects of transition type and ferret ID. This assumes that there is individuality in the decisions to visit the different enrichments. To restrict the complexity of the model, one could also assume that the influence on these decisions due to the weight of the doors is similar for all ferrets (no interaction between weight and ferret ID). The effect of weight could be assumed to be smooth but potentially non-linear and modelled as a spline ("generalised additive mixed-effects model"). This would result in a model shown as model 1 in Table 4.
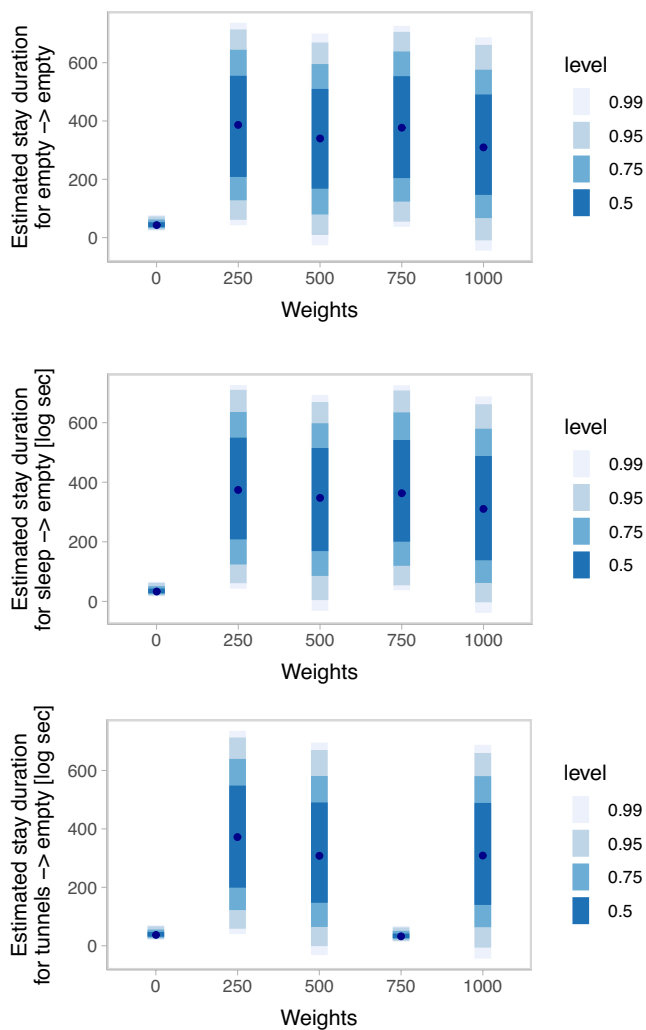
It became visible quickly that this model was over-specified to the extent that at least the standard procedure for setting up priors and initial values failed. One reason is likely to be that in theory, 63 different transition types were possible but only a maximum of 75% of these transitions were observed on any observation day of the individual ferrets. With increasing weight of the doors, this percentage dropped rapidly. Therefore, we restricted the data set to weights up to and including 1000 g. With only five different weights at fixed values (see Young, 2016), the spline could no longer be estimated and we included weight as a factor variable (with sum contrasts). Finally, the interaction between the transition type and ferret ID was not numerically tractable, either, and, therefore, we could not estimate the variability in making behavioural decisions among individual ferrets. This simplified model resulted in model 2 (Table 4).

Based on the R-hat, the effective sample size and the plots of the Markov-Chain, this model did clearly not converge. Therefore, the iteration parameters were increased to `warmup = 10000`, `iter = 20000`, `thin = 10`, further to `warmup = 80000`, `iter = 100000`, `thin = 20` and finally to `warmup = 200000`, `iter = 300000`, `thin = 100`. This final model again ran over seven days in spite of the relatively small sample size. Still, the model did not converge and some of the chains in the plots actually diverged after periods of convergence. This clearly indicated a problem in model estimation, which is likely due to the large proportion of non-observed transitions (and too many censored observations). This high number of non-observed transitions was likely due to several causes. First, the number of potential transitions was high to begin with, and, therefore, many movements between the chambers by the ferrets would have been needed to start with to cover all possible changes with a certain likelihood. The ferrets had potentially explored some of the resources at the start of the experiment due to their novelty but became less interested quickly and even more so due to the increase in the door weight. Eventually, some of the resources were hardly being visited after some time. This illustrates that experiments to be evaluated with multi-state survival models need to be designed accordingly from the start and using the approach in a post hoc manner as we did here may not always be possible.

Keeping the problems in convergence in mind, we will nevertheless show how changes in stay times in relation to the door weight can be shown graphically for three transitions. Surprisingly, it was not the most common transitions (as seen in Figure 5) that resulted in the best estimates based on the R-hat value and the effective sample size. Many transitions with the empty cage yielded the highest effective sample sizes and an R-hat value most close to one. The stay durations in the empty and in the sleep chamber were much shorter with a door weight of zero compared with the higher weights when going to the empty chamber (again) afterwards (Figure 6, top and middle). The stay durations in the chamber with tunnels when

transiting to the empty chamber showed a similar pattern expect for a duration when door weight was 750 g similar to when the door was not weighed.

Theoretically, it would be of interest to further expand this model to include the type (and the duration) of one or several preceding states to investigate longer chains of behaviours. Given the relatively low percentage of the potential transitions that were observed in this data set to start with, this seems futile. If that had been the aim of the study, a much larger data set (e.g. several days with the same weight) would have needed to be collected. Similarly, as with the hen example, it becomes obvious quickly that the number of statistical parameters that are estimated increases fast with model complexity. Therefore, a large sample size is needed for such detailed analyses as well as a well laid-out strategy of how to present and interpret the results from such analyses.

FIGURE 6 Estimated stay durations of the ferrets for three exemplary behavioural transitions depending on the weight of the doors. Different coverage probabilities of the estimates (levels) are reflected by the different shading. The point and the dark blue shading correspond to the median and the 50% interval, respectively, and, to the box of a classical boxplot accordingly [Colour figure can be viewed at wileyonlinelibrary.com]

In the original evaluation by Reijgwart et al. (2016), day-to-day changes seemed to be fairly gradual with increasing the weight of the doors. At least the three transitions that we presented in some detail here nevertheless suggest that there was a marked shift in behaviour by adding any weight at all. The increased stay durations with some weight on the doors (Figure 6) indicated that the ferrets remained longer with a given resource if a cost was involved and visited each of the resources more rarely. The primary goal of the original study was to determine a preference for one specific enrichment. The evaluation sketched here showed how one could determine whether one resource influences the other, for example whether animals would be more inclined to visit chambers where they could rest or sleep after having eaten, or accessing certain types of resources on certain parts of the day, which could be very interesting in relation to an animal's daily routine, for example.

7 | POTENTIAL DIFFICULTIES

The flexibility of parametric survival modelling as shown here is both the strength and potential weakness of the approach. The number of potential models is vast for any given data set and, therefore, the researcher´s degree of freedom is huge (e.g. Tong, 2019). This means that the maximum model to be explored should be well defined in advance of starting such an evaluation. Even if this aspect has been taken care of, the models resulting from this approach contain a multiplicity of estimated parameters. This means that the amount of data needed to sensibly estimate such models is quite large. Moreover, the run times of these models are high, specifically with (the necessary) larger sample sizes. Therefore, care and patience are needed in estimating and developing these models (and/or powerful computers). Finally, it is not easy to present the results of estimating so many parameters in a way easily digestible for a reader.

8 | CONCLUDING REMARKS

We hope that this tutorial shows how parametric survival modelling with frailties offers a unique approach if repeated behavioural sequences have been observed even if the modelling approach is admittedly far from trivial. These models take full advantage of all the information contained in behavioural sequence data (type of behaviour, duration during which behaviour is performed, sequence of behavioural states), accommodate any conceivable fixed effects structure including treatment variables and past behaviour, and incorporate random effects to reflect a wide variety of experimental designs and, therefore, avoiding pseudo-replication and increase power in detecting effects. The syntax of *brms* is very similar to *lme4*, and therefore, starting out with *brms* is easy for anyone familiar with *lme4*. *brms* includes a much wider choice of families for the error distribution than those needed for censored durations as used in survival analysis, and for instance, families needed for zero-inflated models. In addition, *brms* offers flexibility in multi-parameter

distributions because each parameter can be modelled independently by a unique set of predictors. In this sense, *brms* can be considered an all-purpose Swiss-Army knife for a huge variety of complex mixed models.

In respect to ethological questions, we could also show that, starting from an overall model describing the flow of behaviour, the analysis can be directed to specific types of behavioural transitions that, for example change with experimental conditions or show a high repeatability within individuals. In addition, the statistical approach can aid to predict what type of behaviours are likely to follow a given state, which can also help to gain further insight in relationships between certain behavioural activities. Moreover, an a-priori focus can be laid on specific aspects such as asking about the relative probability of different behavioural transitions starting from the same state. All in all, parametric survival modelling with frailties is a promising approach for researchers interested in behaviour and should become part of their toolbox.

ACKNOWLEDGEMENTS

ETHICS STATEMENT

The data for this study were re-used from previous publications. In Rufener et al.'s studies (2018; 2019), all experimental animal work was approved by the Veterinary Office of the Canton of Bern (Switzerland, approval number BE31/15) and the work complied with Swiss regulations regarding the treatment of experimental animals. Reijgwart et al.'s study (2016) was ethically approved by the Animal Care and Use Committee of Intravacc, Bilthoven, The Netherlands (DEC201400137).

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

ORCID

Lorenz Gygax https://orcid.org/0000-0001-8546-2930
Christina Rufener https://orcid.org/0000-0003-2690-1520

REFERENCES

Arnqvist, G. (2020). Mixed models offer no freedom from degrees of freedom. *Trends in Ecology & Evolution*, 35, 329–335. https://doi.org/10.1016/j.tree.2019.12.004

Asher, L., Collins, L. M., Ortiz-Pelaez, A., Drewe, J. A., Nicol, C. J., & Pfeiffer, D. U. (2009). Recent advances in the analysis of behavioural organization and interpretation as indicators of animal welfare. *Journal of the Royal Society Interface*, 6, 1103–1119. https://doi.org/10.1098/rsif.2009.0221

Asher, L., Harvey, N. D., Green, M., & England, G. C. W. (2017). Application of survival analysis and multistate modeling to understand animal behavior: Examples from guide dogs. *Frontiers in Veterinary Science*, 4, 116. https://doi.org/10.3389/fvets.2017.00116

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Berchtold, A., & Sackett, G. (2002). Markovian models for the developmental study of social behavior. *American Journal of Primatology*, 58, 149–167. https://doi.org/10.1002/ajp.10056

Bressers, M., Meelis, E., Haccou, P., & Kruk, M. (1991). When did it really start or stop: The impact of censored observations on the analysis of duration. *Behavioural Processes*, 23, 1–20. https://doi.org/10.1016/0376-6357(91)90102-6

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R rackage brms. *The R Journal*, 10, 395–411. https://doi.org/10.32614/RJ-2018-017

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76, 1–32. https://doi.org/10.18637/jss.v076.i01

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? *Biological Reviews*, 88, 465–475. https://doi.org/10.1111/brv.12007

Casarrubea, M., Jonsson, G. K., Faulisi, F., Sorbera, F., Di Giovanni, G., Benigno, A., Crescimanno, G., & Magnusson, M. S. (2015). T-pattern analysis for the study of temporal structure of animal and human behavior: A comprehensive review. *Journal of Neuroscience Methods*, 239, 34–46. https://doi.org/10.1016/j.jneumeth.2014.09.024

Dall, S., & Griffith, S. (2014). An empiricist guide to animal personality variation in ecology and evolution. *Frontiers in Ecology and Evolution*, 2. https://doi.org/10.3389/fevo.2014.00003

Duchateau, L., & Janssen, P. (2008). *The frailty model*. Springer Science+Business Media, LLC.

Egge, A. R., Brandt, Y., & Swallow, J. G. (2011). Sequential analysis of aggressive interactions in the stalk-eyed fly *Teleopsis dalmanni*. *Behavioral Ecology and Sociobiology*, 65, 369–379. https://doi.org/10.1007/s00265-010-1054-5

Garcia-Perez, E., Mazzoni, A., Zoccolan, D., Robinson, H. P. C., & Torre, V. (2005). Statistics of decision making in the leech. *Journal of Neuroscience*, 25, 2597–2608. https://doi.org/10.1523/JNEUROSCI.3808-04.2005

Grafen, A. (2002). A state-free optimization model for sequences of behaviour. *Animal Behaviour*, 63, 183–191. https://doi.org/10.1006/anbe.2001.1871

Gygax, L. (2017). Wanting, liking and welfare: The role of affective states in proximate control of behaviour in vertebrates. *Ethology*, 123, 689–704. https://doi.org/10.1111/eth.12655

Haccou, P., & Meelis, E. (1992). *Statistical analysis of behavioural data – An approach based on time-structured models*. Oxford University Press.

Helske, S., & Helske, J. (2019). Mixture hidden markov models for sequence data: The seqHMM package in R. *Journal of Statistical Software*, 88, 1–32. https://doi.org/10.18637/jss.v088.i03

Ivanitskii, V. V., Marova, I. M., & Antipov, V. A. (2017). Sequential organization in the song of thrush nightingale (*Luscinia luscinia*): clustering and sequential order of the song types. *Bioacoustics*, 26, 199–215. https://doi.org/10.1080/09524622.2016.1239132

Ivanouw, J. (2007). Sequence analysis as a method for psychological research. *Nordic Psychology*, *59*, 251–267. https://doi.org/10.1027/1901-2276.59.3.251

Jackson, C. H. (2016). flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, *70*, 8. https://doi.org/10.18637/jss.v070.i08

Kay, M. (2020). *tidybayes: Tidy data and geoms for Bayesian models*. R package version 2.0.1. http://mjskay.github.io/tidybayes/. https://doi.org/10.5281/zenodo.1308151

Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T. A., Watanabe, Y. Y., Murgatroyd, M., & Papastamatiou, Y. P. (2017). Analysis of animal accelerometer data using hidden Markov models. *Methods in Ecology and Evolution*, *8*, 161–173. https://doi.org/10.1111/2041-210X.12657

Levy, R. (2014). Using R formulae to test for main effects in the presence of higher-order interactions. *arXiv*. https://arxiv.org/abs/1405.2094

Macdonald, I. L., & Raubenheimer, D. (1995). Hidden Markov models and animal behaviour. *Biometrical Journal*, *37*, 701–712. https://doi.org/10.1002/bimj.4710370606

Metz, H. A., Dienske, H., de Jonge, G., & Putters, F. A. (1983). Continuous-time Markov chains as models for animal behaviour. *Bulletin of Mathematical Biology*, *45*, 643–658. https://doi.org/10.1007/BF02459596

Moya-Laraño, J., & Wise, D. H. (2000). Survival regression analysis: A powerful tool for evaluating fighting and assessment. *Animal Behaviour*, *60*, 307–313. https://doi.org/10.1006/anbe.2000.1495

Munda, M., Rotolo, F., & Legrand, C. (2012). parfm: Parametric frailty models in R. *Journal of Statistical Software*, *51*, 1–20. https://doi.org/10.18637/jss.v051.i11

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for gaussian and non-gaussian data: A practical guide for biologists. *Biological Reviews*, *85*, 935–956. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2019). *nlme: Linear and Nonlinear mixed effects models*. R package version 3.1-140, https://CRAN.R-project.org/package=nlme

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, *26*, 2389–2430. https://doi.org/10.1002/sim.2712

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reijgwart, M. L., Vinke, C. M., Hendriksen, C. F. M., van der Meer, M., Schoemaker, N. J., & van Zeeland, Y. R. A. (2016). Ferrets' (*Mustela putorius furo*) enrichment priorities and preferences as determined in a seven-chamber consumer demand study. *Applied Animal Behaviour Science*, *180*, 114–121. https://doi.org/10.1016/j.applanim.2016.04.022

Rufener, C., Abreu, Y., Asher, L., Berezowski, J. A., Sousa, F. M., Stratmann, A., & Toscano, M. J. (2019). Keel bone fractures are associated with individual mobility of laying hens in an aviary system. *Applied Animal Behaviour Science*, *217*, 48–56. https://doi.org/10.1016/j.applanim.2019.05.007

Rufener, C., Berezowski, J., Maximiano Sousa, F., Abreu, Y., Asher, L., & Toscano, M. J. (2018). Finding hens in a haystack: Consistency of movement patterns within and across individual laying hens maintained in large groups. *Scientific Reports*, *8*, 12303. https://doi.org/10.1038/s41598-018-29962-x

Sih, A., Bell, A., & Johnson, J. C. (2004a). Behavioral syndromes: An ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*, 372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004b). Behavioral syndromes: An integrative overview. *The Quarterly Review of Biology*, *79*, 241–277. https://doi.org/10.1086/422893

Slooten, E. (1994). Behavior of Hector's dolphin: Classifying behavior by sequence analysis. *Journal of Mammalogy*, *75*, 956–964. https://doi.org/10.2307/1382477

Tong, C. (2019). Statistical inference enables bad science; Statistical thinking enables good science. *The American Statistician*, *73*(sup1), 246–261. https://doi.org/10.1080/00031305.2018.1518264

Worthington, A. M., & Swallow, J. G. (2011). Sequential analysis reveals behavioral differences underlying female-biased predation risk in stalk-eyed flies. *Ethology*, *117*, 829–837. https://doi.org/10.1111/j.1439-0310.2011.01941.x

Young, M. E. (2016). The problem with categorical thinking by psychologists. *Behavioural Processes*, *123*, 43–53. https://doi.org/10.1016/j.beproc.2015.09.009

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.