

**Title:** The interplay between extinction intensity and selectivity:  
2 correlation in trait effects on taxonomic survival

**Running title:** Variation in trait effects on taxonomic survival

4 **Author:** Peter D Smits, psmits@uchicago.edu, Committee on Evolutionary  
Biology, University of Chicago, IL, USA.

6 **Keywords:** macroevolution, extinction, macroecology, Bayesian, brachiopods

**Word count:** approximately 5800.

8 **Table count:** 1.

**Figure count:** 5.

10 **Data archival location:** If accepted, all data and code necessary to duplicate  
this analysis will be made available on DRYAD.

## Abstract

While the effect of geographic range on extinction risk is well documented, how other traits may increase or decrease extinction risk is less well known. I analyze patterns of Paleozoic brachiopod genus durations and their relationship to geographic range, affinity for epicontinental seas versus open ocean environments, and body size. Additionally, I allow for environmental affinity to have a nonlinear effect on duration. Using a hierarchical Bayesian approach, I also model the interaction between the effects of biological traits and a taxon's time of origination. My analysis framework eschews the traditional distinction between background and mass extinction, instead the entire time period is analyzed where these are part of the same continuum. I find evidence that as baseline extinction risk increases, the effect of geographic range increases but the effect of environmental preference tends to decrease. Additionally, I find strong evidence for correlation between the effects of geographic range and the non-linear aspect of environmental preference which may help explain this pattern. For parts of the Paleozoic I find support for a "survival of the generalists" scenario, though there are times where this relationship is absent or even reversed. These results support the hypothesis that as extinction intensity increases, overall extinction selectivity decreases.

## 1 Introduction

How do biological traits affect extinction risk? Biological traits are defined here as descriptors of a taxon's adaptive zone, which is the set of all biotic–biotic and biotic–abiotic interactions that a taxon can experience (Simpson, 1944). In effect, these are descriptors of a taxon's broad-sense ecology. Jablonski (1986) observed that during a mass extinction event, the effects of biological traits on

taxonomic survival decreased in size. However, this pattern was not the case for  
40 the effect of geographic range on survival (Jablonski, 1986).

Jablonski (1986) phrased his conclusions in terms of background versus mass  
42 extinction, but this scenario is readily transferable to a continuous variation  
framework as there is no obvious distinction in terms of extinction rate between  
44 these two states (Wang, 2003). Additionally, the Jablonski (1986) scenario has  
strong model structure requirements in order to test its proposed  
46 macroevolutionary mechanism; not only do the taxon trait effects need to be  
modeled, but the correlation between trait effects need to be modeled as well.

48 There are two end-member macroevolutionary mechanisms which may underlie  
the pattern observed by Jablonski (1986): the effect of geographic range on  
50 predictive survival remains constant and those of other biological traits decrease,  
or the effect of geographic range in predicting survival increases and those of  
52 other biological traits stay constant. Reality, of course, may fall somewhere  
along this continuum.

54 Conceptually, taxon survival can be considered an aspect of “taxon fitness”  
along with expected lineage specific branching/origination rate (Cooper, 1984,  
56 Palmer and Feldman, 2012). A taxon with a beneficial trait should persist for  
longer, on average, than a taxon without that beneficial trait. Here I model  
58 brachiopod taxon durations because trait based differences in extinction risk  
should manifest as differences in taxon durations. Brachiopods are an ideal  
60 group for this study as they are well known for having an exceptionally  
complete fossil record (Foote, 2000). I focus on the brachiopod record from most  
62 of the Paleozoic, from the start of the Ordovician (approximately 485 Mya)  
through the end Permian (approximately 252 Mya) as this represents the time  
64 of greatest global brachiopod diversity (Alroy, 2010).

The analysis of taxon durations, or time from origination to extinction, falls  
66 under the purview of survival analysis, a field of applied statistics commonly  
used in health care (Klein and Moeschberger, 2003) but has a long history in  
68 paleontology (Simpson, 1944, 1953, Van Valen, 1973, 1979). I adopt a  
hierarchical Bayesian survival modeling approach, which represents both a  
70 conceptual and statistical unification of the paleontological dynamic and cohort  
survival analytic approaches (Baumiller, 1993, Foote, 1988, Raup, 1975, 1978,  
72 Simpson, 2006, Van Valen, 1973, 1979). By using a Bayesian framework I am  
able to quantify the uncertainty inherent in the estimates of the effects of  
74 biological traits on survival, especially in cases where the covariates of interest  
(i.e. biological traits) are themselves known with error.

### 76 **1.1 Factors affecting brachiopod survival**

Geographic range is widely considered the most important taxon trait for  
78 estimating differences in extinction risk at nearly all times, with large  
geographic range associated with low extinction risk (Jablonski, 1986, 1987,  
80 Jablonski and Roy, 2003, Payne and Finnegan, 2007), though Foote and Miller  
(2013) find that this generalization does not hold in the Mesozoic. For the  
82 Paleozoic, however, I expect this to hold true for the entire period analyzed.

Epicontinental seas are a shallow-marine environment where the ocean has  
84 spread over the surface of a continental shelf with a depth typically less than  
100m. In contrast, open-ocean coastline environments have much greater  
86 variance in depth, do not cover the continental shelf, and can persist during  
periods of low sea level. Because of this, it is strongly expected that taxa which  
88 favor epicontinental seas would be at great risk during periods of low sea levels,  
such as during glacial periods, when epicontinental seas are drained. During the

90 Paleozoic (approximately 541–252 My), epicontinental seas were widely spread  
91 globally but declined over the Mesozoic (approximately 252–66 My) and have  
92 nearly disappeared during the Cenozoic (approximately 66–0 My) as open-ocean  
93 coastlines became the dominant shallow-marine setting (Johnson, 1974, Miller  
94 and Foote, 2009, Peters, 2008).

Miller and Foote (2009) demonstrated that during several mass extinctions taxa  
96 associated with open-ocean environments tend to have a greater extinction risk  
97 than those taxa associated with epicontinental seas. During periods of  
98 background extinction, however, they found no consistent difference between  
99 taxa favoring either environment. These two environment types represent the  
100 primary environmental dichotomy observed in ancient marine systems (Miller  
101 and Foote, 2009, Peters, 2008, Sheehan, 2001).

102 Given these findings, I predict that as extinction risk increases, the extinction  
103 risk associated with favoring open-ocean environments should generally increase.  
104 Additionally, there is a possible nonlinear relationship between environmental  
105 preference and taxon duration. A long standing hypothesis is that generalists or  
106 unspecialized taxa will have greater survival than specialists (Baumiller, 1993,  
107 Liow, 2004, 2007, Nürnberg and Aberhan, 2013, 2015, Simpson, 1944). In this  
108 analysis I allowed for environmental preference to have a parabolic effect on  
109 taxon duration

110 Body size, measured as shell length, was also considered as a potential trait that  
111 influences extinction risk (Payne et al., 2014). Body size is a proxy for metabolic  
112 activity and other correlated life history traits (Payne et al., 2014). There is no  
113 strong hypothesis of how body size effects extinction risk in brachiopods, such  
114 that a positive, negative, or zero effect are all plausible.

## 2 Materials and Methods

### 116 2.1 Fossil occurrence information

The dataset analyzed here was sourced from the Paleobiology Database  
118 (<http://www.paleodb.org>) which was then filtered based on taxonomic,  
temporal, stratigraphic, and other occurrence information that was necessary  
120 for this analysis. These filtering criteria are very similar to those from Foote and  
Miller (2013) with an additional constraint of being present in the body size  
122 data set from Payne et al. (2014). Epicontinental versus open-ocean assignments  
for each fossil occurrence are partially based on those from Miller and Foote  
124 (2009), with additional occurrences assigned similarly (Miller and Foote,  
personal communication).

126 Fossil occurrences were analyzed at the genus level which is common for  
paleobiological, macroevolutionary, or macroecological studies of marine  
128 invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2013, Kiessling  
and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,  
130 Payne and Finnegan, 2007, Simpson and Harnik, 2009, Vilhena et al., 2013).  
Although species diversity dynamics tend to be of much greater interest than  
132 those of higher taxa, the nature of the fossil record makes accurate and precise  
taxonomic assignments at the species level for all occurrences extremely difficult  
134 if not impossible. Additionally, there is evidence of real differences in biological  
patterns at the genus level versus the species level (Jablonski, 1987). As such,  
136 the choice to analyze genera as opposed to species was in order to assure a  
minimum level of confidence and accuracy in the data analyzed here.

138 Genus duration was calculated as the number of geologic stages from first  
appearance to last appearance, inclusive. Durations were based on geologic

<sup>140</sup> stages as opposed to millions of years because of the inherently discrete nature  
of the fossil record; dates are not assigned to fossils themselves but instead  
<sup>142</sup> fossils are known from a geological interval which represents some temporal  
range. Stages act as effectively irreducible globally consistent temporal intervals  
<sup>144</sup> in which taxa occur. Of note, however, is that stages are of variable length  
which may cause origination cohorts to have artificially larger or smaller  
<sup>146</sup> membership. The hierarchical Bayesian framework used here may help control  
for this potentiality as cohorts with smaller samples or effects will be drawn  
<sup>148</sup> towards to over all mean instead of standing on their own (Gelman et al., 2013).

Genera with a last occurrence in or after the Changhsingian stage were right  
<sup>150</sup> censored at the Changhsingian. Genera with a duration of only one stage were  
left censored (Appendix A). The covariates used to model genus duration were  
<sup>152</sup> geographic range size ( $r$ ), environmental preference ( $v, v^2$ ), and body size ( $m$ ).

Geographic range was calculated using an occupancy approach. First, all  
<sup>154</sup> occurrences were projected onto an equal-area cylindrical map projection. Each  
occurrence was then assigned to one of the cells from a  $70 \times 34$  regular raster  
<sup>156</sup> grid placed on the map. Each grid cell represents approximately 250,000 km<sup>2</sup>.  
The map projection and regular lattice were made using shape files from  
<sup>158</sup> <http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans,  
2015).

<sup>160</sup> For each stage, the total number of occupied grid cells, or cells in which a fossil  
occurs, was calculated. Then, for each genus, the number of grid cells occupied  
<sup>162</sup> by that genus was calculated. Dividing the genus occupancy by the total  
occupancy gives the relative occupancy of that genus. Mean relative genus  
<sup>164</sup> occupancy was then calculated as the mean of the per stage relative occupancies  
of that genus.

<sup>166</sup> Body size data was sourced directly from Payne et al. (2014). Because those  
measurements are presented without error, a measurement error model similar  
<sup>168</sup> to the one for environmental affinity could not be implemented (Appendix A).

Prior to analysis, geographic range and body size were transformed and  
<sup>170</sup> standardized in order to improve interpretability of the results. Geographic  
range, which can only vary between 0 and 1, was logit transformed. Body size,  
<sup>172</sup> which is defined for all positive real values, was natural log transformed. These  
covariates were then standardized by mean centering and dividing by two times  
<sup>174</sup> their standard deviation following Gelman and Hill (2007).

## 2.2 Analytical approach

<sup>176</sup> Hierarchical modelling is a statistical approach that explicitly takes into account  
the structure of the observed data in order to model both the within and  
<sup>178</sup> between group variance (Gelman et al., 2013, Gelman and Hill, 2007). The units  
of study (e.g. genera) each belong to a single grouping (e.g. origination cohort).  
<sup>180</sup> These groups are considered separate draws from a shared probability  
distribution (e.g. all cohorts, observed and unobserved). The group-level  
<sup>182</sup> parameters are then estimated simultaneously with the other parameters of  
interest (e.g. covariate effects) (Gelman et al., 2013). The subsequent estimates  
<sup>184</sup> are partially pooled together, where parameters from groups with large samples  
or effects remain large while those of groups with small samples or effects are  
<sup>186</sup> pulled towards the overall group mean.

This partial pooling is one of the greatest advantages of hierarchical modeling.  
<sup>188</sup> By letting the groups “support” each other, parameter estimates better reflect  
our statistical uncertainty. Additionally, this partial pooling helps control for  
<sup>190</sup> multiple comparisons and possibly spurious results as effects with little support

are drawn towards the overall origination cohort mean (Gelman et al., 2013,  
192 Gelman and Hill, 2007).

All covariate effects (regression coefficients), as well as the intercept term  
194 (baseline extinction risk), were allowed to vary by group (origination cohort).  
The covariance/correlation between covariate effects was also modeled. This  
196 hierarchical structure allows inference for how covariates effects may change  
with respect to each other while simultaneously estimating the effects  
198 themselves, propagating our uncertainty through all estimates.

Additionally, instead of relying on point estimates of environmental affinity, I  
200 treat environmental affinity as a continuous measure of the difference between  
the taxon's environmental occurrence pattern and the background occurrence  
202 pattern (Appendix A). Background occurrence pattern for a single genus is  
defined as the distribution of environmental occurrences for all other genera  
204 present during the duration of the genus of interest.

### 2.3 Survival model

206 Genus durations were assumed to follow either an exponential or Weibull  
distribution, both of which make different assumptions about how a taxon's  
208 duration may effect its instantaneous extinction risk (Klein and Moeschberger,  
2003). The exponential distribution assumes that extinction risk is independent  
210 of duration. In contrast, the Weibull distribution allows for age dependent  
extinction via the shape parameter  $\alpha$ , though only as a monotonic function of  
212 duration. Importantly, the Weibull distribution is equivalent to the exponential  
distribution when  $\alpha = 1$ .

214 The following variables are here defined:  $y_i$  is the duration of genus  $i$  in geologic  
stages,  $X$  is the matrix of covariates including a constant term,  $B_j$  is the vector

<sup>216</sup> of regression coefficients for origination cohort  $j$ ,  $\Sigma$  is the covariance matrix of  
 the regression coefficients,  $\tau$  is the vector of scales (standard deviations) of the  
<sup>218</sup> between-cohort variation in regression coefficient estimates,  $\Omega$  is the correlation  
 matrix of the regression coefficients, and  $\alpha_j$  is the shape parameter for cohort  $j$   
<sup>220</sup> with  $a$  is the overall mean shape parameter and  $\pi$  is the variance between  
 estimates of  $\alpha_j$ .

<sup>222</sup> The exponential model is defined

$$y_i \sim \text{Exponential}(\lambda)$$

$$\lambda_i = \exp(\mathbf{X}_i B_{j[i]})$$

$$B \sim \text{MVN}(\vec{\mu}, \Sigma)$$

$$\Sigma = \text{Diag}(\vec{\tau}) \Omega \text{Diag}(\vec{\tau})$$

$$\mu_k \sim \begin{cases} \mathcal{N}(0, \psi_k \nu) & \text{if } k \neq r, \text{ or} \\ \mathcal{N}(-1, 1) & \text{if } k = r \end{cases} \quad (1)$$

$$\tau_k \sim C^+(1)$$

$$\psi_k \sim C^+(1) \text{ if } k \neq r$$

$$\nu \sim C^+(1)$$

$$\Omega \sim \text{LKJ}(2).$$

Similarly, the Weibull model is defined

$$\begin{aligned}
y_i &\sim \text{Weibull}(\alpha_{j[i]}, \sigma) \\
\sigma_i &= \exp \left( \frac{-(\mathbf{X}_i B_{j[i]})}{\alpha_{j[i]}} \right) \\
B &\sim \text{MVN}(\vec{\mu}, \Sigma) \\
\Sigma &= \text{Diag}(\vec{\tau}) \Omega \text{Diag}(\vec{\tau}) \\
\log(\alpha) &\sim \mathcal{N}(a, \pi) \\
\mu_k &\sim \begin{cases} \mathcal{N}(0, \psi_k \nu) & \text{if } k \neq r, \text{ or} \\ \mathcal{N}(-1, 1) & \text{if } k = r \end{cases} \\
\tau_k &\sim \text{C}^+(1) \\
a &\sim \mathcal{N}(0, 1) \\
\pi &\sim \text{C}^+(1) \\
\psi_k &\sim \text{C}^+(1) \text{ if } k \neq r \\
\nu &\sim \text{C}^+(1) \\
\Omega &\sim \text{LKJ}(2).
\end{aligned} \tag{2}$$

<sup>224</sup> The principal difference between the exponential and Weibull models is the inclusion of the shape parameter  $\alpha$ . Note that  $\sigma$  is approximately equivalent to <sup>226</sup>  $1/\lambda$ .

<sup>228</sup> For an explanation of how these two models were developed, parameter explanations, and choice of priors, please see Appendix B. Note that these models as written (Eq. 1, 2) do not include how the uncertainty in <sup>230</sup> environmental affinity is included nor how censored observations are included.

For an explanation of both of these aspects, see Appendices A and C.

<sup>232</sup> **2.4 Parameter estimation**

The joint posterior was approximated using a Markov chain Monte Carlo routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic programming language Stan (Stan Development Team, 2014). The posterior distribution was approximated from four parallel chains run for 10,000 draws each, split half warm-up and half sampling and thinned to every 10th sample for a total of 5000 posterior samples. Chain convergence was assessed via the scale reduction factor  $\hat{R}$  where values close to 1 ( $\hat{R} < 1.1$ ) indicate approximate convergence, meaning that the chains are approximately stationary and the samples are well mixed (Gelman et al., 2013).

**2.5 Model evaluation**

Models were evaluated using both posterior predictive checks and an estimate of out-of-sample predictive accuracy. The motivation behind posterior predictive checks as tools for determining model adequacy is that replicated data sets using the fitted model should be similar to the original data (Gelman et al., 2013). Systematic differences between the simulations and observations indicate weaknesses of the model fit. An example of a technique that is very similar would be inspecting the residuals from a linear regression.

The strategy behind posterior predictive checks is to draw simulated values from the joint posterior predictive distribution,  $p(y^{rep}|y)$ , and then compare those draws to the empirically observed values (Gelman et al., 2013). To accomplish this, for each replicate, a single value is drawn from the marginal posterior distributions of each regression coefficient from the final model as well as estimates of  $\alpha_j$  for the Weibull model (Eq. 1, 2). Then, given the covariate

information  $\mathbf{X}$ , a new set of  $n$  genus durations are generated giving a single  
258 replicated data set  $y^{rep}$ . This is repeated 1000 times in order to provide a distribution of possible values that could have been observed given the model.

260 In order to compare the fitted model to the observed data, various graphical comparisons or test quantities need to be defined. The principal comparison  
262 used here is a comparison between non-parameteric approximation of the survival function  $S(t)$  as estimated from both the observed data and each of the  
264 replicated data sets. The purpose of this comparison is to determine if the model approximates the same survival/extinction pattern as the original data.

266 The exponential and Weibull models were compared for out-of-sample predictive accuracy using the widely-applicable information criterion (WAIC) (Watanabe,  
268 2010), a more fully Bayesian alternative to AIC or DIC (Gelman et al., 2013). Out-of-sample predictive accuracy is a measure of the expected fit of the model  
270 to new data. However, because the Weibull distribution reduces to the exponential distribution when  $\alpha = 1$ , my interest is not in choosing between  
272 these models. Instead, comparisons of WAIC values are useful for better understanding the effect of model complexity on out-of-sample predictive  
274 accuracy. The calculation of WAIC used here corresponds to the “WAIC 2” formulation recommended by Gelman et al. (2013). For an explanation of how  
276 WAIC is calculated, see Appendix D. Lower values of WAIC indicate greater expected out-of-sample predictive accuracy than higher values.

### 278 3 Results

As stated above, posterior approximations for both the exponential and Weibull  
280 models achieved approximate stationarity after 10,000 steps, as all parameter estimates have an  $\hat{R} < 1.1$ .

<sup>282</sup> Comparisons of the survival functions estimated from 1000 posterior predictive  
<sup>283</sup> data sets to the estimated survival function of the observed genera demonstrates  
<sup>284</sup> that Weibull models approximately capture the observed pattern of extinction  
<sup>285</sup> (Fig. 1). This is partially expected given that the unit of analysis is genus  
<sup>286</sup> duration instead of species duration, which can alter the shape of  $S(t)$  (Foote,  
<sup>287</sup> 2001, Raup, 1975, 1978, 1985). Additionally, the Weibull model is expected to  
<sup>288</sup> have better out-of-sample predictive accuracy than the exponential model  
<sup>289</sup> (WAIC 4534 versus 4604). Because the difference in WAIC between these two  
<sup>290</sup> models is large, only those results from the Weibull model will be discussed.

Estimates of the overall mean covariate effects  $\mu$  can be considered  
<sup>292</sup> time-invariant generalizations for brachiopod survival during the Paleozoic  
<sup>293</sup> (Table 1). Consistent with prior expectations, geographic range size has a  
<sup>294</sup> negative effect on extinction risk: genera with large ranges are expected to have  
greater durations than genera with small ranges.

<sup>296</sup> I infer that body size has no distinguishable effect on brachiopod taxonomic  
<sup>297</sup> survival. While the mean estimate of the effect of body size on extinction risk is  
<sup>298</sup> negative, this estimate is within 2 standard deviations of 0 (mean  $\mu_m = -0.07$ ,  
standard deviation 0.08; Table 1).

<sup>300</sup> Interpretation of the effect of environmental preference  $v$  on duration is slightly  
<sup>301</sup> more involved. Because a quadratic term is the equivalent of an interaction  
<sup>302</sup> term, both  $\mu_v$  and  $\mu_{v^2}$  have to be interpreted together because it is illogical to  
<sup>303</sup> change values of  $v$  without also changing values  $v^2$ . To determine the nature of  
the effect of  $v$  on duration I calculated the multiplicative effect of environmental  
preference on extinction risk.

<sup>306</sup> Given mean estimated extinction risk  $\tilde{\sigma}$ , we can define the extinction risk

multiplier of an observation with environmental preference  $v_i$  as

$$f(v_i) = \exp\left(\frac{-(\mu_v v_i + \mu_{v^2} v^2)}{\exp(a)}\right). \quad (3)$$

This function  $f(v_i)$  has a y-intercept of  $\exp(0)$  or 1 because it does not have a non-zero intercept term. Equation 3 can be either concave up or down. A concave down  $f(v_i)$  may indicate that genera of intermediate environmental preference have greater durations than either extreme, and *vice versa* for concave up function.

The expected effect of environmental preference as a multiplier of expected extinction risk can then be visualized (Fig. 2). This figure depicts 1000 posterior predictive estimates of Eq. 3 across all possible values of  $v$ . The number indicates the posterior probability that the function is concave down, with generalists having lower extinction risk/greater duration than either type of specialist. Note that the inflection point of Figure 2 is approximately  $x = 0$ , something that is expected given the estimate of  $\mu_v$  (Table 1). Given the amount of variation (Fig. 2), the interpretation is that while evidence for “survival of the unspecialized” is weak it is still more probable than not (posterior probability 74%).

Coefficients with greater values of  $\tau$  have greater between-cohort variation (Table 1). The covariate effects with the greatest between origination cohort variation are  $\beta_r$ ,  $\beta_v$ , and  $\beta_{v^2}$ . Estimates of  $\beta_m$  have negligible between cohort variation, as there is less between cohort variation than the between cohort variation in baseline extinction risk  $\beta_0$ .

The correlations between the intercept term  $\beta_0$  and the effects of the taxon traits are of particular interest for evaluating the Jablonski (1986) scenario (Fig. 3a first column/last row). The correlation terms of  $\Omega$  (Fig. 3a) describe the

relationship between the coefficients and how their estimates may vary together  
332 across cohorts. Keep in mind that when  $\beta_0$  is low, extinction risk is low; and  
conversely, when  $\beta_0$  is high, then extinction risk is high.

334 Marginal posterior probabilities of the correlations between the level of baseline  
extinction risk  $\beta_0$  and the effects of the taxon traits indicate that the correlation  
336 between expected extinction risk and both geographic range  $\beta_r$  and  $\beta_{v^2}$  are of  
particular note (Fig. 3b).

338 There is an approximately 85% probability that  $\beta_0$  and  $\beta_r$  are negatively  
correlated (Fig. 3b), meaning that as extinction risk increases, the  
340 effect/importance of geographic range on genus duration may increase. There is  
a 93% probability that  $\beta_0$  and  $\beta_{v^2}$  are negatively correlated (Fig. 3b), meaning  
342 that as extinction risk decreases, the peakedness of  $f(v_i)$  may increase and the  
relationship tends towards concave down. Additionally, there is a 94%  
344 probability that values of  $\beta_r$  and  $\beta_{v^2}$  are positively correlated (Mean correlation  
0.45, standard deviation 0.26).

346 While the overall group level estimates are of particular importance when  
defining time-invariant differences in extinction risk, it is also important and  
348 useful to analyze the individual level parameter estimates in order to better  
understand how parameters actually vary across cohorts.

350 In comparison to the overall mean extinction risk  $\mu_0$ , cohort level estimates  $\beta_0$   
show some amount of variation through time as expected by estimates of  $\tau_0$   
352 (Fig. 4). A similar, if slightly greater, amount of variation is also observable in  
cohort estimates of the effect of geographic range  $\beta_r$  (Fig. 4). Again, smaller  
354 values of  $\beta_0$  correspond to lower expected extinction risk. Similarly, smaller  
values of  $\beta_r$  correspond to greater decrease in extinction risk with increasing  
356 geographic range

How the effect of environmental affinity varies between cohorts can be observed  
358 by using the cohort specific coefficients estimates. Following the same procedure  
used earlier (Eq. 3), but substituting cohort specific estimates of  $\beta_v$  and  $\beta_{v^2}$  for  
360  $\mu_v$  and  $\mu_{v^2}$ , the cohort specific effect of environmental preference as a multiplier  
of mean extinction risk can be calculated. This was done only for the Weibull  
362 model, though the observed pattern should be similar for the exponential model.

As expected based on the estimates of  $\tau_v$  and  $\tau_{v^2}$ , there is greater variation in  
364 the peakedness of  $f(v_i)$  than there is variation between convave up and down  
functions (Fig. 5). Twelve of the 33 cohorts have less than 50% posterior  
366 probability that generalists are potentially expected to be shorter lived than  
specialists, though two of those cases have approximately a 50% probability of  
368 being either concave up or down. This is congruent with the 0.74 posterior  
probability that  $\mu_{v^2}$  is positive/ $f(v_i)$  is concave down.

370 Additionally, for some cohorts there is a quite striking pattern where the effect  
of environmental preference  $v$  has a nearly-linear relationship (Fig. 5). These are  
372 scenarios where one of the end member preferences is expected to have a greater  
duration than either intermediate or the opposite end member preference.

374 Whatever curvature is present in these nearly-linear cases is due to the  
definition of  $f(v)$  as it is not defined for non-negative values of  $\sigma$  (Eq. 3). For all  
376 stages between the Givetian through the Viséan, inclusive, intermediate  
preferences are expected to be of intermediate extinction risk when compared  
378 with epicontinental specialists (lowest risk) or open-ocean specialists (highest  
risk). This time period represents most of the Devonian through the early  
380 Carboniferous.

Interestingly, the estimates of  $\alpha_j$  have low between cohort variation with most  
382 cohorts having approximately the same estimate as the overall estimate of  $\alpha$   
(Fig. 4). Also, for the majority of the time analyzed there is evidence for

<sup>384</sup> accelerating extinction with taxon age. If this is due to biological causes  
(Rosindell et al., 2015, Wagner and Estabrook, 2014) or an artifact of  
<sup>386</sup> preservation/the minimum observable duration of a genus (Sepkoski, 1975) is  
unknown. While it is expected that a non-exponential model will better fit  
<sup>388</sup> genus-level data, this would be reflected as decelerating extinction risk with  
taxon age as opposed to accelerating (Foote, 2001, Raup, 1975, 1978, 1985).  
<sup>390</sup> This uncertainty remains an open issue for future analysis.

## 4 Discussion

<sup>392</sup> My results demonstrate that both the effects of geographic range and the  
peakedness/concavity of environmental preference are both negatively  
<sup>394</sup> correlated with baseline extinction risk, meaning that as baseline extinction risk  
increases the effect size of geographic range increases but the curvature of the  
<sup>396</sup> effect of environmental preference decreases (Fig. 3b).

However, the interaction between geographic range and origination cohort,  
<sup>398</sup> although weak, indicates that neither of the two proposed end-member  
macroevolutionary mechanisms is strictly right. The evidence for a correlation  
<sup>400</sup> between changes in the effect of geographic range and the curvature of the effect  
of environmental preference may help explain this result. I hypothesize that  
<sup>402</sup> because taxa with large geographic ranges encompass more possible  
environments, when extinction intensity is high there is little coherent difference  
<sup>404</sup> in the environmental preference among the surviving taxa. The intensity  
decreases the selectivity such that the effect of environmental preference is  
<sup>406</sup> effectively washed out by the strength of the effect of geographic range; this is  
consistent with similar hypotheses presented in Raup (1991).

<sup>408</sup> For the approximately 233 My period analyzed there is an approximate 74%

posterior probability that brachiopod genera with intermediate environmental  
410 preferences are expected to have a lower extinction risk than either end  
members. However, the over all curvature of  $f(v_i)$  is not very peaked meaning  
412 that when averaged over the entire Phanerozoic this relationship may not lead  
to large differences in extinction risk (Fig. 2). Note that the duration of the  
414 period analyzed is approximately four times then length of the Cenozoic (e.g.  
time since the extinction of the non-avian dinosaurs). This result gives very  
416 weak support for the universality of the hypothesis that environmental  
generalists have greater survival than environmental specialists (Liow, 2004,  
418 Nürnberg and Aberhan, 2013, 2015, Simpson, 1944).

The posterior variance in the estimate of overall  $f(v_i)$  reflects the large between  
420 cohort variance in cohort specific estimates of  $f(v_i)$  (Fig. 5). Given that there is  
only a 74% posterior probability that the expected overall estimate of  $f(v_i)$  is  
422 concave down, it is not surprising that there are some stages where the  
estimated relationship is in fact the reverse of the prior expectation.  
424 Additionally, some of those same stages where  $f(v_i)$  does not resemble the prior  
expectation of a concave down nonlinear relation are instead is highly skewed  
426 and effectively linear (Fig. 5). These results demonstrate that, while the  
group-level estimate may only weakly support one hypothesis, the cohort-level  
428 estimates may exhibit very different characteristics.These results are also  
consistent with aspects of Miller and Foote (2009) who found that the effect of  
430 environmental preference on extinction risk was quite variable and without  
obvious patterning during times of background extinction.

432 There are two mass extinction events that are captured within the time frame  
considered here: the Ordovician-Silurian and the Frasnian-Famennian. The  
434 cohorts bracketing these events are worth considering in more detail.

The proposed mechanism for the end Ordovician mass extinction is a decrease

436 in sea level and the draining of epicontinental seas due to protracted glaciation  
(Johnson, 1974, Sheehan, 2001). My results are broadly consistent with this  
438 scenario with both epicontinental and open-ocean specialists having a much  
lower expected duration than intermediate taxa (Fig. 5). All of the stages  
440 between the Darriwillian and the Llandovery, except the Hirnantian, have a  
high probability (90+) that  $f(v)$  is concave down. The pattern for the  
442 Darriwillian, which proceeds the supposed start of Ordovician glacial activity,  
demonstrates that taxa tend to favor open-ocean environments are expected to  
444 have a greater duration than either intermediate or epicontinental specialists, in  
decreasing order.

446 For nearly the entire Devonian estimates of  $f(v)$  indicate that one of the  
environmental end members is favored over the other end member of  
448 intermediate preference (Fig. 5). For almost the entirety the Givetian through  
the end of the Devonian and into the Viséan, I find that epicontinental favoring  
450 taxa are expected to have a greater duration than either intermediate or  
open-ocean specialists. Additionally, for nearly the entire Devonian and through  
452 to the Visean, the cohort-specific estimates of  $f(v)$  are concave-up. This result,  
however, seems to reflect the intensity of the seemingly nearly-linear difference  
454 in expected duration across the range of  $v$  as opposed to an inversion of the  
weakly expected curvilinear pattern.

456 The use of genera as the unit of the study and how to exactly interpret the  
effects of the biological traits is a remaining concern. For example, if any of the  
458 traits analyzed here are associated with increases in speciation rates, this might  
increase the duration of genera through self-renewal (Raup, 1991, 1994), which  
460 would be an example of the difference in biological pattern between species and  
genera (Jablonski, 1987, 2007, 2008). This could lead to a trait appearing to  
462 decrease generic level extinction risk by increasing species level origination rate

instead of decreasing species level extinction risk. However, given the nature of  
464 the fossil record and maintaining a minimum level of data consistency/quality,  
there is no simple solution to decreasing this uncertainty in the interpretations  
466 of how the biological traits studied at the genus-level may translate to the  
species-level.

468 The model used here could be improved through either increasing the number of  
analyzed taxon traits, expanding the hierarchical structure of the model to  
470 include other major taxonomic groups of interest, and the inclusion of explicit  
phylogenetic relationships between the taxa in the model as an additional  
472 hierarchical effect. An example taxon trait that may be of particular interest is  
the affixing strategy or method of interaction with the substrate of the taxon,  
474 which has been found to be related to brachiopod survival where, for  
cosmopolitan taxa, taxa that are attached to the substrate are expected to have  
476 a greater duration than those that are not (Alexander, 1977).

It is theoretically possible to expand this model to allow for comparisons within  
478 and between major taxonomic groups. This approach would better constrain the  
brachiopod estimates while also allowing for estimation of similarities and  
480 differences in cross-taxonomic patterns. The major issue surrounding this  
particular expansion involves finding an similarly well sampled taxonomic group  
482 that is present during the Paleozoic. Example groups include Crinoidea,  
Ostracoda, and other members of the “Paleozoic fauna” (Sepkoski Jr., 1981).

484 Taxon traits like environmental preference or geographic range (Hunt et al.,  
2005, Jablonski, 1987) are most likely heritable, at least phylogenetically  
486 (Housworth et al., 2004, Lynch, 1991). Without phylogenetic context, this  
analysis assumes that differences in extinction risk between taxa are  
488 independent of the shared evolutionary history of those taxa (Felsenstein, 1985).  
In contrast, the origination cohorts only capture shared temporal context. For

example, if taxon duration is phylogenetically heritable, then closely related taxa may have more similar durations as well as more similar biological traits.

Without taking into account phylogenetic similarity the effects of these biological traits would be inflated solely due to inheritance. The inclusion of phylogenetic context as an additional individual-level hierarchical effect independent of origination cohort would allow for determining how much of the observed variability is due to shared evolutionary history versus shared temporal context versus actual differences associated with biological traits (Smits, 2015).

In summary, patterns of Paleozoic brachiopod survival were analyzed using a fully Bayesian hierarchical survival modelling approach while also eschewing the traditional separation between background and mass extinction. I find that as baseline extinction risk increases, the form of the selectivity of extinction changes such that during periods of low extinction risk the effect environmental preference is expected to change from nonlinear to potentially linear or even absent while the effect of geographic range increases. In particular, the correlation between the effect of geographic range and the curvature of the effect of environmental preference on taxon survival supports the hypothesis that during periods of high extinction intensity of the effect of geographic range effectively washes out the effects of other biological traits (Jablonski, 1987, Raup, 1991). Finally, I find weak support for “survival of the unspecialized” (Liow, 2004, 2007, Nürnberg and Aberhan, 2013, 2015, Simpson, 1944) as a general characterization of the effect of environmental preference on extinction risk (Fig. 2), most origination cohorts conforming to this hypothesis (Fig. 5).

## Acknowledgements

514 I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for and  
M. Marites Villarosa Garcia for draft comments. Additionally, thank you A.  
516 Miller for the epicontinental versus open-ocean assignments. This entire study  
would not have been possible without the Herculean effort of the many  
518 contributors to the Paleobiology Database. In particular, I would like to thank  
J. Alroy, M. Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim, A. Hendy,  
520 S. Holland, L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T. Olszewski, P.  
Novack-Gottshall, M. Patzkowsky, M. Uhen, L. Villier, and P. Wager. This work  
522 was supported by a NASA Exobiology grant (NNX10AQ446) to A. Miller and  
M. Foote. I declare no conflicts of interest. This is Paleobiology Database  
524 publication XXX.

## References

- 526 Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation  
to the mode of stabilization on the substrate. *Palaeogeography,*  
528 *Palaeoclimatology, Palaeoecology* 21:209–226.
- Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal  
530 Groups. *Science* 329:1191–1194.
- Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of  
532 filter morphology on evolutionary rates. *Paleobiology* 19:304–321.
- Cooper, W. S., 1984. Expected time to extinction and the concept of  
534 fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- Felsenstein, J., 1985. Phylogenies and the comparative method. *American  
536 Naturalist* 125:1–15.
- Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites.  
538 *Paleobiology* 14:258–271.
- , 2000. Origination and extinction components of taxonomic diversity:  
540 Paleozoic and post-Paleozoic dynamics. *Paleobiology* 26:578–605.
- , 2001. Evolutionary Rates and the Age Distributions of Living and  
542 Extinct Taxa. Pp. 245–294, *in* J. B. Jackson, S. Lidgard, and F. K.  
McKinney, eds. *Evolutionary Patterns: Growth, Form, and Tempo in the  
544 Fossil Record*. University of Chicago Press, Chicago.
- Foote, M. and A. I. Miller, 2013. Determinants of early survival in marine  
546 animal genera. *Paleobiology* 39:171–192.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B.

- 548 Rubin, 2013. Bayesian data analysis. 3 ed. Chapman and Hall, Boca Raton,  
FL.
- 550 Gelman, A. and J. Hill, 2007. Data Analysis using Regression and  
Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- 552 Harnik, P. G., C. Simpson, and J. L. Payne, 2013. Long-term differences in  
extinction risk among the seven forms of rarity. Proceedings of the Royal  
Society B: Biological Sciences 282.
- 554
- Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL  
<http://CRAN.R-project.org/package=raster>. R package version 2.3-24.
- 556 Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively  
Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine  
Learning Research 15:1351–1381.
- 560 Housworth, E. A., P. Martins, and M. Lynch, 2004. The Phylogenetic Mixed  
Model. The American Naturalist 163:84–96.
- 562 Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a  
comment on "On the heritability of geographic range sizes". American  
Naturalist 166:129–135.
- 564
- Jablonski, D., 1986. Background and mass extinctions: the alternation of  
macroevolutionary regimes. Science 231:129–133.
- 566 ———, 1987. Heritability at the species level: analysis of geographic ranges of  
cretaceous mollusks. Science 238:360–363.
- 568 ———, 2007. Scale and hierarchy in macroevolution. Palaeontology 50:87–109.
- 570 ———, 2008. Species Selection: Theory and Data. Annual Review of Ecology,  
Evolution, and Systematics 39:501–524.

- 572 Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and  
living molluscs. *Proceedings. Biological sciences / The Royal Society*  
574 270:401–6.
- Johnson, J. G., 1974. Extinction of Perched Faunas. *Geology* 2:479–482.
- 576 Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine  
benthic biodiversity dynamics through Triassic–Jurassic time. *Paleobiology*  
578 33:414–434.
- Klein, J. P. and M. L. Moeschberger, 2003. *Survival Analysis: Techniques for  
580 Censored and Truncated Data*. 2nd ed. Springer, New York.
- Liow, L. H., 2004. A test of Simpson's "rule of the survival of the relatively  
582 unspecialized" using fossil crinoids. *The American Naturalist* 164:431–43.
- , 2007. Does versatility as measured by geographic range, bathymetric  
584 range and morphological variability contribute to taxon longevity? *Global  
Ecology and Biogeography* 16:117–128.
- 586 Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary  
biology. *Evolution* 45:1065–1080.
- 588 Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings:  
the kinetics of mass extinction and origination. *Science* 326:1106–9.
- 590 Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range  
predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*  
592 39:360–372.
- , 2015. Interdependence of specialization and biodiversity in Phanerozoic  
594 marine invertebrates. *Nature communications* 6:6602.

- Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental  
596 than evolvability. *PLoS one* 7:e38025.
- Payne, J. L. and S. Finnegan, 2007. The effect of geographic range on  
598 extinction risk during background and mass extinction. *Proceedings of the  
National Academy of Sciences* 104:10506–11.
- 600 Payne, J. L., N. A. Heim, M. L. Knope, and C. R. McClain, 2014. Metabolic  
dominance of bivalves predates brachiopod diversity decline by more than 150  
602 million years. *Proceedings of the Royal Society B* 281:20133122.
- Peters, S. E., 2008. Environmental determinants of extinction selectivity in the  
604 fossil record. *Nature* 454:626–9.
- Raup, D. M., 1975. Taxonomic survivorship curves and Van Valen's Law.  
606 *Paleobiology* 1:82–96.
- , 1978. Cohort Analysis of generic survivorship. *Paleobiology* 4:1–15.
- 608 ———, 1985. Mathematical models of cladogenesis. *Paleobiology* 11:42–52.
- , 1991. *Extinction: Bad Genes or Bad Luck?* Norton, New York.
- 610 ———, 1994. The role of extinction in evolution. *Proceedings of the National  
Academy of Sciences* 91:6758–6763.
- 612 Rosindell, J., L. J. Harmon, and R. S. Etienne, 2015. Unifying ecology and  
macroevolution with individual-based theory. *Ecology letters* 18:472–82.
- 614 Sepkoski, J. J., 1975. Stratigraphic biases in the analysis of taxonomic  
survivorship. *Paleobiology* 1:343–355.
- 616 Sepkoski Jr., J. J., 1981. A factor analytic description of the Phanerozoic  
marine fossil record. *Paleobiology* 7:36–53.

- 618 Sheehan, P., 2001. The late Ordovician mass extinction. *Annual Review of*  
Earth and Planetary Sciences 29:331–364.
- 620 Simpson, C., 2006. Levels of selection and large-scale morphological trends.  
Ph.D. thesis, University of Chicago.
- 622 Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine  
bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- 624 Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University  
Press, New York.
- 626 ———, 1953. *The Major Features of Evolution*. Columbia University Press,  
New York.
- 628 Smits, P. D., 2015. Expected time-invariant effects of biological traits on  
mammal species duration. *Proceedings of the National Academy of Sciences* .
- 630 Stan Development Team, 2014. Stan: A c++ library for probability and  
sampling, version 2.5.0. URL <http://mc-stan.org/>.
- 632 Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.  
———, 1979. Taxonomic survivorship curves. *Evolutionary Theory* 4:129–142.
- 634 Vilhena, D. A., E. B. Harris, C. T. Bergstrom, M. E. Maliska, P. D. Ward, C. A.  
Sidor, C. A. E. Strömberg, and G. P. Wilson, 2013. Bivalve network reveals  
636 latitudinal selectivity gradient at the end-Cretaceous mass extinction.  
*Scientific Reports* 3:1790.
- 638 Wagner, P. J. and G. F. Estabrook, 2014. Trait-based diversification shifts  
reflect differential extinction among fossil taxa. *Proceedings of the National  
640 Academy of Sciences* 111:16419–16424.

- Wang, S. C., 2003. On the continuity of background and mass extinction.  
642 Paleobiology 29:455–467.
- Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and  
644 Widely Applicable Information Criterion in Singular Learning Theory.  
Journal of Machine Learning Research 11:3571–3594.

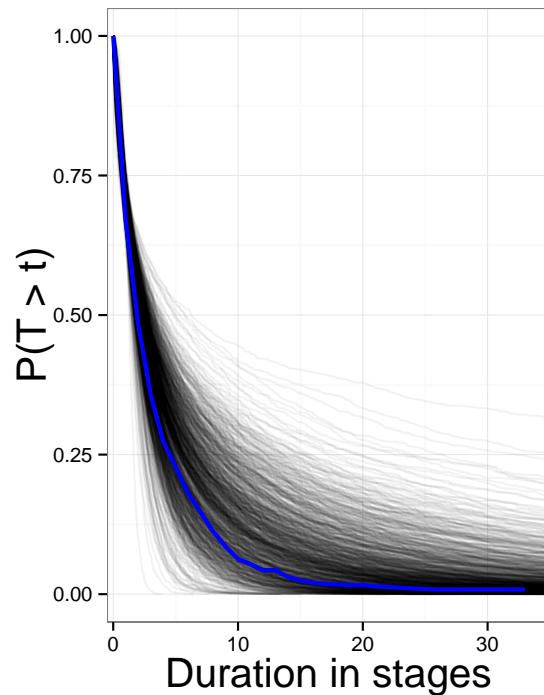


Figure 1: Comparison of empirical estimates of  $S(t)$  versus estimates from 1000 posterior predictive data sets.  $S(t)$  corresponds to  $P(T > t)$  as it is the probability that a given genus observed at age  $t$  will continue to live. This is equivalent to the probability that  $t$  is less than the genus' ultimate duration  $T$ .

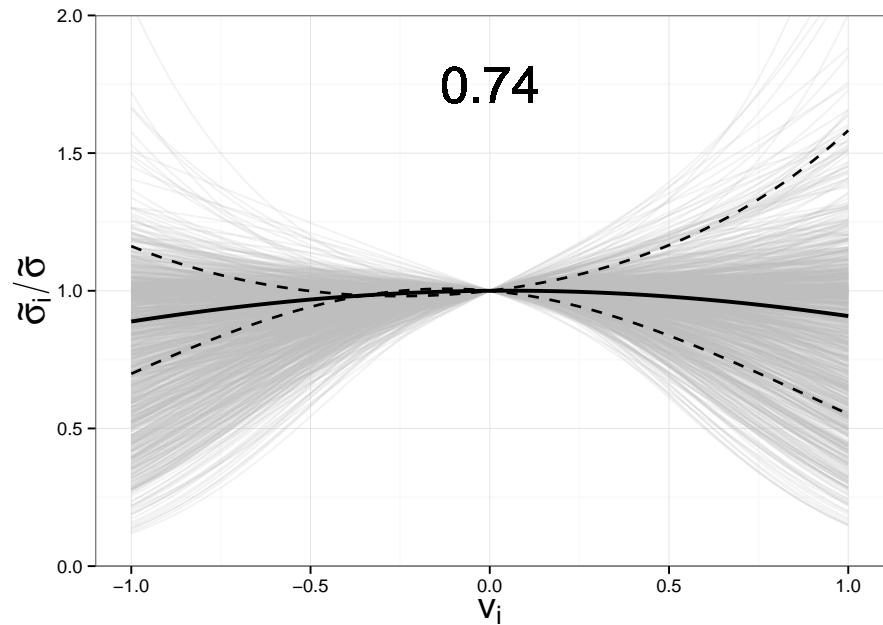


Figure 2: The overall expected relationship  $f(v_i)$  between environmental affinity  $v_i$  and a multiplier of extinction risk (Eq. 3). Each grey line corresponds to a single draw from the posterior predictive distribution, the solid black line corresponds to the median of the posterior predictive distribution, and the dashed black lines correspond to the median relationship plus or minus one standard deviation. The overall shape of  $f(v_i)$  is concave down with an optimum of close 0, which corresponds to affinity approximately equal to the expectation based on background environmental occurrence rates.

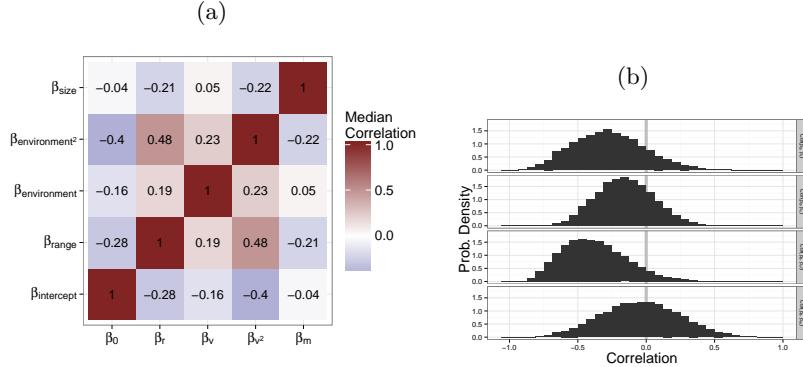


Figure 3: **A:** Heatmap for the median estimates of the terms of the correlation matrix  $\Omega$  between cohort-level covariate effects. Both the exponential (left) and Weibull (right) models are presented. The off-diagonal terms are the correlation between the estimates of the cohort-level estimates of the effects of covariates, along with intercept/baseline extinction risk. **B:** Marginal posterior distributions of the correlations between intercept terms/baseline extinction risk and the effects of each of the covariates. These are presented for both the exponential (left) and Weibull (right) models.

parameter	mean	standard deviation
$\mu_i$	-1.52	0.16
$\mu_r$	-1.39	0.13
$\mu_v$	-0.04	0.16
$\mu_{v^2}$	0.30	0.45
$\mu_m$	-0.07	0.08
$\tau_i$	0.77	0.14
$\tau_r$	0.40	0.13
$\tau_v$	1.05	0.23
$\tau_{v^2}$	1.87	0.64
$\tau_m$	0.24	0.13

Table 1: Group-level estimates of the intercept terms the effects of biological traits on brachiopod generic survival from equations 1 and 2, presented as means and standard deviations.  $\mu$  values are the location parameters of the effects, while  $\tau$  values are the scale terms describing the variation between cohorts. The subscripts correspond to the following:  $i$  intercept,  $r$  geographic range,  $v$  environmental affinity,  $v^2$  environmental affinity squared,  $m$  body size.

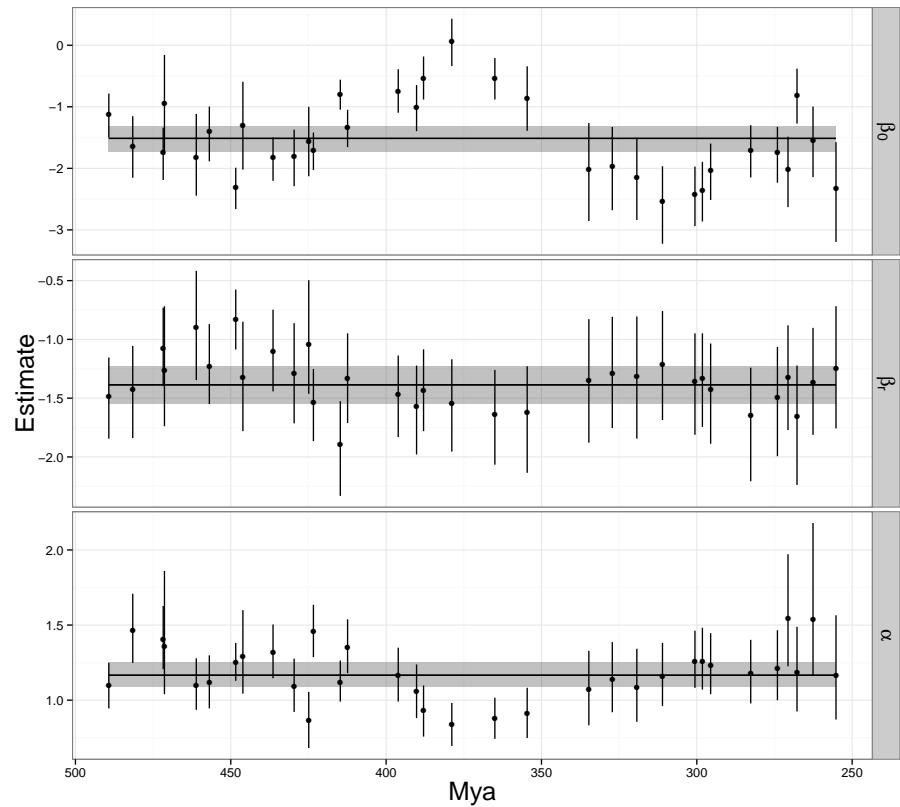


Figure 4: Comparison of cohort-specific estimates of  $\beta_0$  (first row), cohort-specific estimates of the effect of geographic range on extinction risk  $\beta_r$  (second row), and cohort-specific estimates of the Weibull shape parameter  $\alpha$  where values greater than 1 correspond to accelerating extinction with age, and those below 1 to decelerating extinction with age. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction risk along with 80% credible intervals.

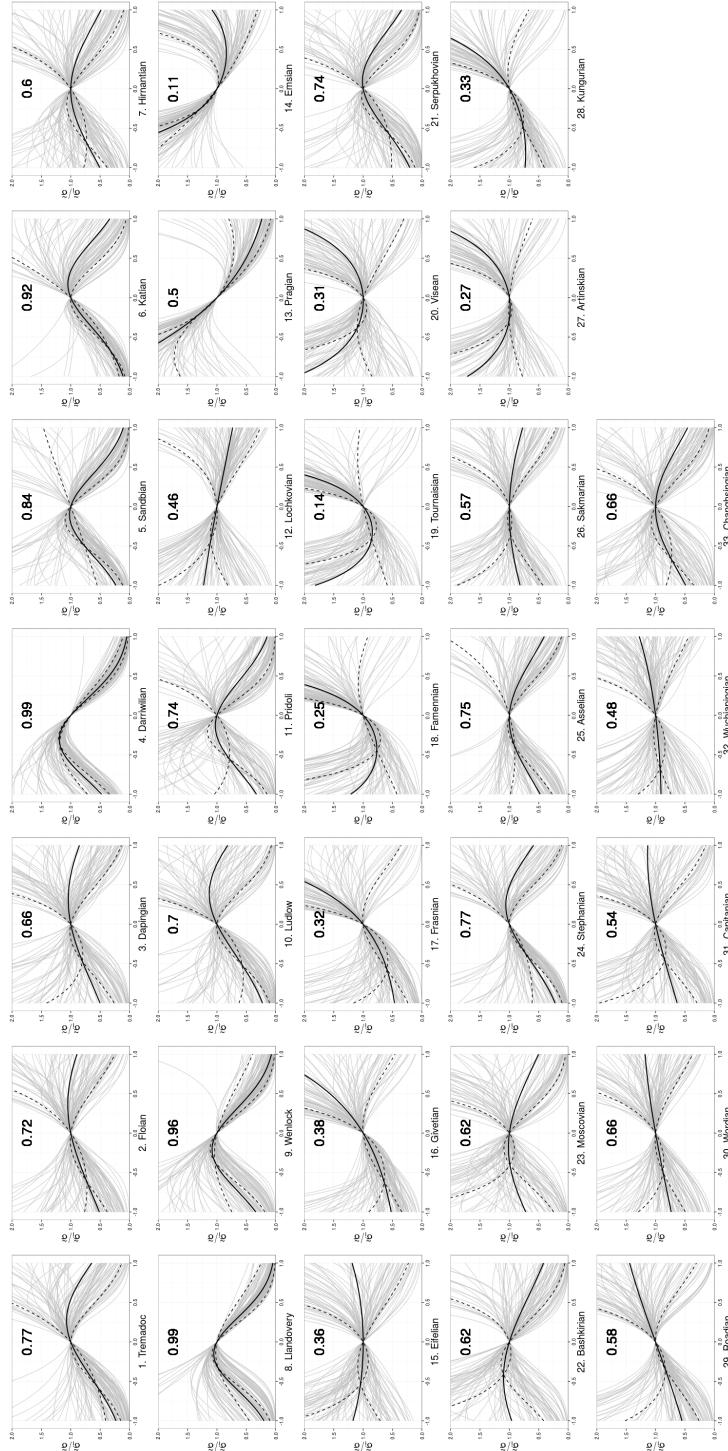


Figure 5: Comparison of the cohort-specific estimates of  $f(v_i)$  (Eq. 3) for the 33 analyzed origination cohorts. The stage of origination is labeled on the x-axis of each panel. The oldest stage is in the upper left, while the youngest is in the lower left. The number in each panel corresponds to the posterior probability that  $f(v_i)$  is concave down. Those that are highlighted in red have less than 51% posterior predictive probability that  $f(v_i)$  is concave down. Each grey line corresponds to a single draw from the posterior predictive distribution, the solid black line corresponds to the median of the posterior predictive distribution, and the dashed black lines correspond to the median relationship plus or minus one standard deviation. Note that all estimates must pass through  $y = 1$  when  $x = 0$  (Eq. 3).