https://doi.org/10.70401/Geromedicine.2026.0044

Supplementary information

Old worms, new tricks: Dynamical instability explains late-life rejuvenation in *C. elegans*

Djakim Latumalea, Adrian Molière, Peter Fedichev, Collin Y. Ewald, Jan Gruber

Supplementary Note 1: Predictive comparison of the Langevin and Gompertz models

We compared the survival predictions of the Langevin model with those of an age-dependent Gompertz hazard. The main predictive test used the auxin day-21 cohort as a held-out intervention cohort: neither model was fitted to these post-day-21 observations. The analysis therefore asks whether the observed post-intervention survival trajectory is compatible with the conditional survival distribution predicted by each model, and separately which model assigns greater predictive probability to the observed outcomes.

Models

Langevin model. The Langevin model follows

$$\dot{z} = \alpha z + g z^2 + \sqrt{2D}\,\eta(t),$$

with $z(0) = 0$, a reflecting boundary at $z = 0$, and death at the first passage of z through $z_{\max} = \alpha/g$. Parameters were taken from the corresponding training conditions. For the auxin day-21 cohort, trajectories follow the DMSO parameters until day 21 and the auxin day-10 parameters afterward, with the latent state continuous across the intervention. Survival probabilities were estimated from 10^6 simulated first-passage trajectories using Euler–Maruyama integration with $\Delta t = 0.1$ days.

Gompertz model. The comparison model has hazard

$$M(t) = M_0 e^{\alpha_G t}.$$

The DMSO day-10 and auxin day-10 cohorts were fitted separately by maximum likelihood using the grouped survival likelihood. For the auxin day-21 prediction, the model follows the fitted DMSO Gompertz hazard until day 21. At day 21, the hazard remains continuous at the mortality level already reached, while its subsequent exponential growth rate changes to the α_G estimated from the auxin day-10 cohort. No parameters are fitted to the auxin day-21 cohort.

Post-day-21 predictive assessment

The primary analysis conditions on animals that were alive and still under observation immediately after day 21. Deaths or censoring at day 21 are not included, leaving $n = 59$ animals. For each model, the relevant null prediction is its fixed conditional survival distribution

$$S(t \mid T > 21) = \frac{S(t)}{S(21)}.$$

Conditional survival calibration. We assessed whether the observed post-day-21 survival trajectory was compatible with each fixed model prediction using a model-based parametric bootstrap with 10,000 replicates. Because deaths were recorded at inspection times rather than as exact death times, observed and predicted survival were compared on the common inspection grid.

As a summary of remaining survival, we calculated conditional restricted mean survival time (RMST) from day 21 through a specified endpoint τ. The reported difference is

$$\Delta\mathrm{RMST} = \mathrm{RMST}_{\mathrm{observed}} - \mathrm{RMST}_{\mathrm{model}},$$

so positive values indicate longer observed remaining survival than predicted. A one-sided bootstrap test quantified how often data generated from the model produced an RMST excess at least as large as observed. We also used an omnibus statistic based on the integrated squared distance between the complete observed and predicted conditional survival curves. This curve-distance test detects discrepancies in either direction. Bootstrap p values use a plus-one correction, giving a minimum attainable value of $1/(10{,}000 + 1) \approx 0.0001$.

Paired predictive scores. For a death, the predictive probability is the conditional probability assigned to the observed death interval; for censoring, it is the corresponding conditional survival probability. The log predictive score, ℓ_i, is the natural logarithm of this probability. For animal i we define

$$\Delta\ell_i = \ell_{i,\text{Langevin}} - \ell_{i,\text{Gompertz}},$$

so positive values favor Langevin. Uncertainty in the mean paired difference was quantified using a paired animal-level percentile bootstrap with 20,000 replicates. We additionally report a two-sided normal-approximation p value for the null hypothesis that the mean animal-level log-score difference is zero, calculated from the observed mean difference and its standard error. The primary comparison uses all recorded post-day-21 follow-up; follow-up through days 35 and 39 was examined as endpoint sensitivity analyses.

All tests condition on the fitted model predictions and treat animals as independent. They therefore quantify sampling variation in the held-out post-day-21 cohort. The analyses are exploratory and reported p values are unadjusted.

Results

Observed and predicted survival curves for the two training cohorts and the held-out auxin day-21 cohort are shown in Figure S1, with the corresponding model scores and post-day-21 predictive comparisons summarized in Table S2. Among the 59 animals remaining under observation after day 21, conditional RMST through day 39 was 7.90 days. The Gompertz model predicted only 2.65 days, corresponding to an observed-minus-predicted difference of 5.25 days. In 10,000 bootstrap samples generated from the fixed Gompertz prediction, none produced an RMST excess as large as the one observed; with the plus-one correction this corresponds to $p \approx 0.0001$. The omnibus survival-curve test likewise showed a strong discrepancy between the observed conditional survival trajectory and the Gompertz prediction ($p < 0.001$). Thus, the post-day-21 observations were highly inconsistent with the survival distribution predicted by the fitted Gompertz model.

The Langevin model predicted a conditional RMST through day 39 of 8.89 days, compared with the observed value of 7.90 days, giving an observed-minus- predicted difference of -0.99 days. The omnibus survival-curve discrepancy did not reach the 0.05 significance threshold ($p = 0.067$). The corresponding one-sided RMST test, which specifically tests for observed survival exceeding the model prediction, gave $p = 0.908$. These results do not establish statistical equivalence or prove exact calibration; rather, the post-day-21 data did not show the same clear evidence of predictive inadequacy for the Langevin model that was observed for Gompertz.

The conclusion was similar when the endpoint was restricted to day 35. The observed conditional RMST exceeded the Gompertz prediction by 5.15 days, with both the one-sided RMST test and the omnibus curve-distance test giving $p \approx 0.0001$. For Langevin, the observed-minus-predicted RMST difference was -0.31 days, with $p = 0.520$ for the omnibus curve-distance test.

The direct paired predictive comparison also favored Langevin. Over full post-day-21 follow-up, the mean Langevin-minus-Gompertz log-score difference was 4.04 nats per animal (95% paired-

bootstrap CI, 2.60–5.59). The advantage persisted when follow-up was restricted to day 35, with a mean difference of 3.95 nats per animal (95% CI, 2.58–5.40). Follow-up through day 39 gave the same numerical result as full follow-up because all deaths contributing to the comparison had occurred by day 39.

Interpretation

The held-out post-day-21 analysis provides two complementary assessments of predictive performance. First, the absolute calibration tests ask whether the observed survival trajectory is plausible under each model's fixed post-intervention prediction. The Gompertz model substantially underpredicted remaining survival, and the observed trajectory was highly unlikely under its predicted conditional survival distribution. In contrast, the corresponding omnibus test did not reject the Langevin prediction at the 0.05 level. Second, the paired log-score analysis directly compares the probability assigned by the two models to the same observed outcomes and showed a clear predictive advantage for Langevin.

These results support the Langevin model as a better predictor of the held-out post-intervention survival data than the specific age-dependent Gompertz model considered here. They do not by themselves establish that the underlying biological process is Markovian or exclude all forms of explicit age dependence. The latent state is unobserved, survival is recorded at discrete inspection times. The inference is therefore specifically about the adequacy of the fixed survival predictions for the held-out post-day-21 cohort.

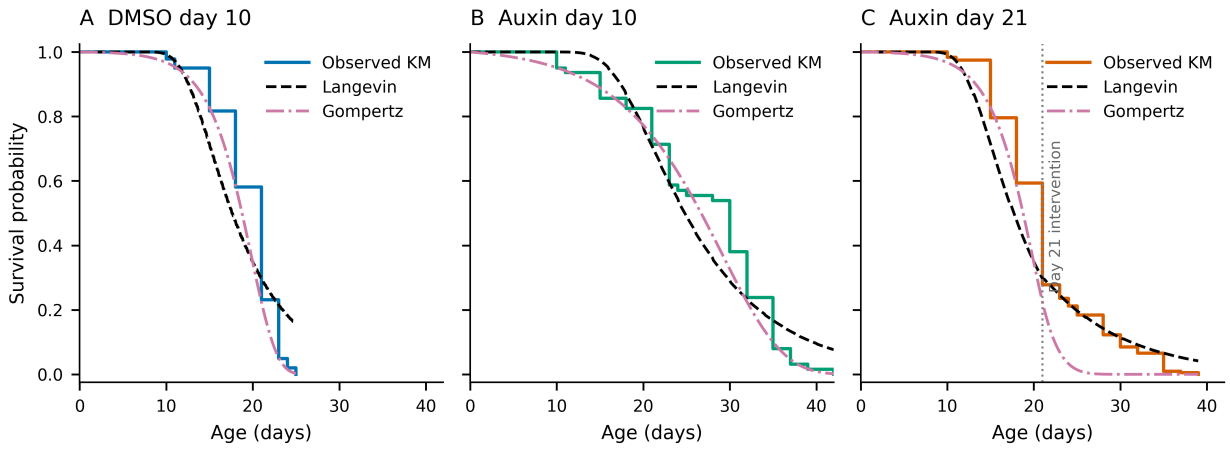


Figure S1: Observed Kaplan–Meier survival for the DMSO day-10 (A), auxin day-10 (B), and auxin day-21 (C) cohorts, together with the Langevin and Gompertz survival predictions. Panels A and B show the training conditions. Panel C is a held-out two-phase prediction in which neither model is fitted to the auxin day-21 cohort; the dotted vertical line marks the intervention at day 21. The curves show unconditional survival from cohort origin, whereas the statistical tests condition on survival and observation after day 21.

A. Whole-cohort model scores

Condition	Model	N	log L	k	Nominal AIC
DMSO day 10	Langevin	232	-404.61	3	815.22
DMSO day 10	Gompertz	232	-342.64	2	689.28
Auxin day 10	Langevin	81	-212.62	2	429.24
Auxin day 10	Gompertz	81	-164.52	2	333.04
Auxin day 21	Langevin	247	-459.86	0	919.72
Auxin day 21	Gompertz	247	-688.39	0	1376.79

B. Conditional post-day-21 survival

Model	End day	Observed RMST	Predicted RMST	Delta RMST	p (greater)	p (curve)
Gompertz	35	7.80	2.65	+5.15	1.00e-04	1.00e-04
Langevin	35	7.80	8.10	-0.31	0.690	0.520
Gompertz	39	7.90	2.65	+5.25	1.00e-04	1.00e-04
Langevin	39	7.90	8.89	-0.99	0.908	0.067

C. Paired predictive comparison (Langevin minus Gompertz)

Follow-up	N	Mean difference	95% bootstrap CI	p
Full follow-up	59	+4.04	[2.60, 5.59]	1.84e-07
Through day 35	59	+3.95	[2.58, 5.40]	6.12e-08
Through day 39	59	+4.04	[2.60, 5.59]	1.84e-07

Table S2: Model predictions and post-intervention survival. **A:** Whole-cohort grouped log-likelihoods use the unmodified model probabilities. k is the number of parameters estimated from each condition; the auxin day-21 cohort is held out ($k = 0$). Nominal AIC values are descriptive. **B:** Conditional RMST is measured among the 59 animals alive and under observa-tion after day 21. ΔRMST is observed minus predicted. The one-sided test assesses excess observed RMST, and the omnibus test uses the integrated squared distance between observed and predicted conditional survival curves. Bootstrap tests use 10,000 replicates. **C:** Log-score differences are reported in nats per animal; positive values favor Langevin. Confidence intervals use 20,000 paired animal-level bootstrap replicates. Full follow-up and day 39 are identical because all contributing deaths occurred by day 39. P values for the paired predictive comparison are two-sided normal-approximation tests of a zero mean animal-level log-score difference.

Supplementary Note 2: Sensitivity of survival to the instability rate α and feedback strength g

To examine how survival responds when only one parameter changes, we perturbed α and g one at a time around the fitted DMSO baseline and compared the resulting simulated survival.

Model and baseline

Survival was simulated from the Langevin equation of the main text (Eq. 1),

$$\frac{dz}{dt} = \alpha z + g z^2 + \sqrt{2D}\,\eta(t), \tag{1}$$

by Euler–Maruyama integration with fixed step $\Delta t = 0.1$ and 2×10^4 paths per parameter set, with $z(0) = 0$, a reflecting boundary at $z = 0$, and death defined as the first passage of z through $z_{\max} = \alpha/g$. The baseline is the DMSO fit (main text, Table I): $\alpha_0 = 0.155$, $g_0 = 2.86 \times 10^{-3}$, $D = 1.27$ ($z_{\max,0} = 54.3$). D was held fixed throughout. We report the median simulated lifespan T_{50} and, as a measure of the spread of lifespans, the interquartile range $\mathrm{IQR} = T_{25} - T_{75}$, where T_p is the time at which the simulated Kaplan-Meier curve falls to $S = p$.

Design

Each parameter was multiplied by 0.5, 0.75, 1.25 and 1.5 while the other was held at its baseline value. Because $z_{\max} = \alpha/g$, the failure threshold moves with α when α is varied and inversely with g when g is varied.

Results

Decreasing either parameter extends median survival and increasing either shortens it (Figure S2, Table S3). For equal relative changes, α has the somewhat larger effect on the median: over $\pm 25\%$, T_{50} ranges from 15.8 to 19.9 d when α is varied and from 16.3 to 19.4 d when g is varied (14.4–22.9 d and 15.2–22.0 d over $\pm 50\%$).

The two parameters differ more clearly in how they change the shape of the survival curve. Varying g shifts the curve in time with nearly unchanged spread: the interquartile range stays between 7.3 and 7.9 d across the whole g sweep. Varying α also stretches or compresses the curve: the interquartile range falls from 13.1 to 5.2 d as α increases from $0.5\,\alpha_0$ to $1.5\,\alpha_0$, and lower α produces a longer late-life tail.

The reduction in α alone raises T_{50} from 17.6 to 20.8 d, the reduction in g alone to 20.4 d, and both together to 24.8 d. They differ in shape as in sweeps: the reduction in g alone leaves the interquartile range unchanged (7.7 d), whereas the reduction in α alone widens it to 10.6 d (11.2 d with both).

Conclusion

Median survival responds to both parameters, somewhat more strongly to α per unit relative change, and at the fitted intervention the two reductions contribute comparably. They differ in how they reshape the survival curve: changing g mainly shifts in time, whereas changing α also rescales its time axis, altering the spread of lifespans and the late-life tail.

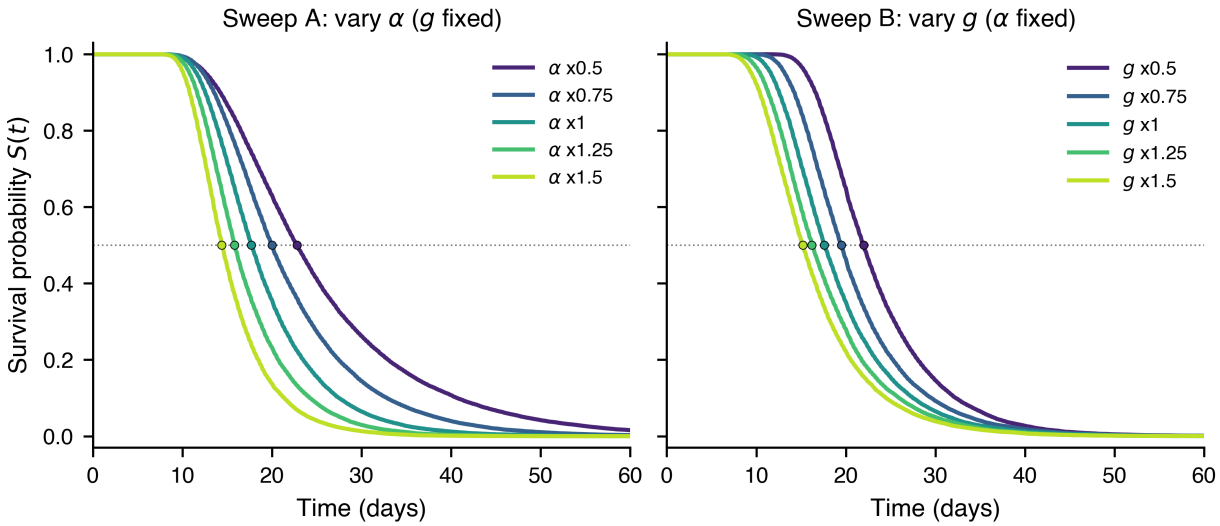


Figure S2: Simulated Kaplan-Meier survival curves under one-at-a-time perturba-tion of α (left) or g (right) by $\pm 25\%$ and $\pm 50\%$ around the DMSO baseline ($\times 1$). Circles mark each curve's median survival time T_{50}. Changing g mainly shifts the curves in time; changing α also changes their spread.

Perturbation	Factor	α (d^{-1})	g (10^{-3})	$z_{\max}$	T_{50} (d)	IQR (d)
α varies	0.5	0.0777	2.86	27.2	22.9	13.1
α varies	0.75	0.116	2.86	40.8	19.9	9.7
α varies	1	0.155	2.86	54.3	17.6	7.7
α varies	1.25	0.194	2.86	67.9	15.8	6.2
α varies	1.5	0.233	2.86	81.5	14.4	5.2
g varies	0.5	0.155	1.43	108.7	22.0	7.9
g varies	0.75	0.155	2.14	72.4	19.4	7.7
g varies	1	0.155	2.86	54.3	17.6	7.7
g varies	1.25	0.155	3.57	43.5	16.3	7.4
g varies	1.5	0.155	4.29	36.2	15.2	7.3

Table S3: One-at-a-time sensitivity of survival to α and g. Each parameter scaled by the given factor around the DMSO baseline ($\alpha_0 = 0.155$, $g_0 = 2.86 \times 10^{-3}$, $z_{\max,0} = 54.3$) with the other held fixed. $D = 1.27$ throughout. T_{50} is the median of 2×10^4 simulated first-passage times (Euler–Maruyama, $\Delta t = 0.1$); IQR is $T_{25} - T_{75}$, the interquartile range of the same simulated lifespans.

Condition	RMSE	KS
DMSO day 10	0.1158	0.165
Auxin day 10	0.0806	0.1817
Auxin day 21	0.0488	0.1271

Table S1. Goodness of fit of the simulated survival curves. Root-mean-square error (RMSE) and Kolmogorov-Smirnov distance (KS, the maximum absolute difference) between each simulated Kaplan-Meier curve and its corresponding experimental Kaplan-Meier curve, evaluated at the experimental curve's observed time points, for DMSO controls and cohorts receiving auxin from day 10 or day 21. Both statistics are in units of survival probability (0-1); lower values indicate closer agreement. Experimental data are from Venz *et al*. (2021).