4  Supporting Text 4–Accounting for environmental effects and measurement error

We cannot perfectly observe $D$ or $V$. Instead, our observations of the "true" $D$ and $V$ are subject to distortion by the confounding effects environmental variations and measurement error. If we represent environmental effects by the factor $Z$ and measurement error by $\epsilon$, we can relate our measured $D*$ and $V*$ to the "true" underlying $D$ and $V$ as follows:

$$v_i^* = v_i + z_{v,i} + \epsilon_{v,i} + \epsilon_v$$
$$d_i^* = d_i + z_{d,i} + \epsilon_{d,i} + \epsilon_d$$

(1)

with $E(\epsilon_{v,i}) = E(\epsilon_{d,i}) = E(z_{v,i}) = E(z_{d,i}) = 0$ and $Var(\epsilon_{v,i}) = \sigma_{v,v}^2$, $Var(\epsilon_{d,i}) = \sigma_{d,d}^2$, $Var(z_{v,i}) = \sigma_{v,v}^2$, and $Var(z_{d,i}) = \sigma_{d,d}^2$. In this application, there can be no measurement error in $Z$, as $Z$ is a categorical variable coding for a discrete organizational unit (the plate or device on which the measurements were obtained) predefined before the start of the study.

In regression Eq. 1 in S3 Text, we seek to estimate the relationship $d_i = \alpha + \epsilon_{b,i}$. Combining this with (1) we find that the relationship between observed variables $D*$ and $V*$ is

$$d_i^* = \alpha v_i^* - z_{d,i} + \alpha z_{v,i} + \epsilon_{b,i} - \epsilon_{d,i} + \alpha \epsilon_{v,i} - \epsilon_d + \alpha \epsilon_v$$

(2)

The correlation between $v_i$, $Z$, and $\epsilon_{v,i}$ is an unfortunate example measurement error’s effect on regression. We wish to estimate $\alpha$ in the presence of environmental factors $Z$ using the multiple regression model

$$D_i = \beta V_i + \beta Z X_i + \epsilon_i;$$

(3)

However, the influence of $\alpha$ on $V$, $Z$ and $\epsilon$ in Eq. 2 introduces structural problems in such estimation. Multiple regression in the presence of measurement error is well studied (Buonaccorsi. Measurement Error: Models, Methods, and Applications) (E. S. Gillespie. The influence of measurement errors in multiple linear regression.), and predicts that $\beta_V$ can be a biased estimator of $\alpha$. The nature of this bias will depend on the correlation between $Z$ and $V$, as well as the relative magnitudes of $\sigma_{v,v}^2$, $\sigma_{d,d}^2$, $\sigma_{z,v}^2$ and $\sigma_{z,d}^2$. To avoid extended analytic excursions, in the following sections we follow two approaches to identify any bias in our estimation of $\alpha$ using regression.

1. **Bias in multiple regression estimates I : Estimating $\alpha$ within batches**

First, we study the relationship between our estimates of $\alpha$ across the whole data set to estimates obtained in subsets of environmental factors (e.g temperature) that influence the $V$ and $D$ of all animals housed on that scanner. As previously shown elsewhere, regression estimates $\alpha$ at a value significantly below 1. Without accounting for batch effects, the single regression Eq. 1 in S3 Text fails to identify a significant deviation from $\alpha = 1$, highlighting the importance of accounting for batch effects. To see whether $B_V$ in the multiple regression is a biased estimator of $\alpha$, we considered the relationship between $D$ and $V$ in each batch separately, using the single regression model Eq. 1 in S3 Text. In Table A in S4 Text, we find that when grouped with each batch-estimated $\beta_y$, the all-batch estimate of $\beta_y$ is the median value. This suggests that the structure of our multiple regression approach (Eq. 2) does not add substantial bias to our estimate of $\alpha$ on top of what is present in a single variate regression model of the same data.

<table>
<thead>
<tr>
<th>Data used</th>
<th>Model</th>
<th>N</th>
<th>$\beta_V$</th>
<th>95% CI</th>
<th>p-value (a &lt;1)</th>
<th>Intercept</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>$d_i = \beta V_i + \beta Z X_i$</td>
<td>1380</td>
<td>0.86</td>
<td>(0.82-0.91)</td>
<td>.&lt;.0001</td>
<td>4.83</td>
<td>(4.23-5.43)</td>
</tr>
<tr>
<td>All</td>
<td>$d_i \beta V_i$</td>
<td>1380</td>
<td>0.98</td>
<td>(0.93-1.02)</td>
<td>0.1432</td>
<td>3.49</td>
<td>(2.93-4.05)</td>
</tr>
<tr>
<td>Batch 1</td>
<td>$d_i \beta V_i$</td>
<td>312</td>
<td>0.82</td>
<td>(0.72-0.92)</td>
<td>0.0006</td>
<td>6.00</td>
<td>(4.73-7.26)</td>
</tr>
<tr>
<td>Batch 2</td>
<td>$d_i \beta V_i$</td>
<td>430</td>
<td>0.78</td>
<td>(0.70-0.87)</td>
<td>.&lt;.0001</td>
<td>6.33</td>
<td>(5.24-7.42)</td>
</tr>
<tr>
<td>Batch 3</td>
<td>$d_i \beta V_i$</td>
<td>391</td>
<td>0.97</td>
<td>(0.89-1.05)</td>
<td>0.2666</td>
<td>2.49</td>
<td>(1.63-3.35)</td>
</tr>
<tr>
<td>Batch 4</td>
<td>$d_i \beta V_i$</td>
<td>247</td>
<td>0.89</td>
<td>(0.79-0.98)</td>
<td>0.0054</td>
<td>4.56</td>
<td>(3.36-5.77)</td>
</tr>
</tbody>
</table>

Table A in S4 Text: Comparing $\alpha$ estimates within and across batches
The values of $\beta_D$ in Table A in S4 Text suggests that our multiple regression approach does not introduce additional bias compared to single regression model estimates. The estimate of $\beta_v$ obtained via multiple regression is centered among batch-specific $\beta_v$ estimates.

It still remains the case that these single regression model estimates may themselves be biased due to the measurement error term $\epsilon_{v,i}$ in Eq. (1). Measurement error in single regression produces attenuation in parameter estimates that case might make $|\beta_D| \leq |\alpha|$. This would impede our ability to distinguish between causal models (1), (2), and (3) by preventing us from distinguishing between $\alpha = 1$ or $\alpha < 1$. Fortunately in this case analytic theory is more straightforward than for multiple regression. Attenuation will take the form $|\beta_V| = |\hat{\kappa}|$ with

$$\hat{\kappa} = \frac{\hat{\sigma}_V^2}{\hat{\sigma}_D^2 + \hat{\sigma}_{Derr}^2}$$

Here, $\hat{\sigma}_V^2$ and $\hat{\sigma}_{Verr}^2$ are the variances of $V$ and the measurement error in $V$ in eq. (1). By substituting in our empiric estimate $\beta_V = 0.8$ and asserting $\alpha = 1$, we can use eq. (4) to predict the relative magnitude of measurement error required to produce a bias large enough to generate the observed deviation from $\beta_v = 1$. We find that $\hat{\sigma}_{Verr}^2 = 0.25\hat{\sigma}_V^2$, so that a measurement error with a standard deviation equal to half of $V$ would be required produce the observed bias. Unfortunately, the nature of our time-to-event data is that repeated measurements are impossible, as each individual’s cessation of vigorous movement and death can only be observed once. It’s therefore impossible to directly estimate the true empiric ratio of $\hat{\sigma}_{Verr}^2/\hat{\sigma}_V^2$. However, we can inspect various aspects of our data to estimate the contribution of measurement error to our analysis.

In the wild-type data set shown in main text Fig 2, $\hat{\sigma}_V = 2.2$. Numerical simulation using a Weibull distribution $V$ and additive Gaussian noise for measurement error, we found that the underlying biological noise and measurement error would combine to show $\hat{\sigma}_V = 2.2$ when $\sigma_V = 1.99$ and $\sigma_{Verr} = 1.00$, as shown below.

Fig A in S4 Text: Additive error with the standard deviation such that $\hat{\sigma}_{Verr}^2 = 0.25\hat{\sigma}_V^2$ is shown in the PDF inset. This error has the effect of transforming the "true" distribution (black) into an "observed" distribution (red).

Fig A in S4 Text provides an example in which measurements of an individuals’ vigorous movement span would need to have an average absolute error of 0.8 days. This is roughly two-fold higher than the typical error of our automated analysis pipeline, and subsequent by-hand annotation of our data sets reduces the error to less than 0.1 days error on average. Yet, reducing error below 0.8 days on average is a taxing standard, and so we performed additional diagnostics to see whether the observed $\sigma_V$ or other population level statistics correlated with the observed $\beta_V$. 

2
Fig B in S4 Text: The relationship between observed $B_v$ and population summary statistics across replicate experiments

We found that the $\beta_v$ was inversely correlated with $\sigma_V$, but also with the $E(V)$. Were $V$ to vary across replicates by a scaling relationship, we’d expect proportional changes in $\sigma_V$ and $E(V)$. So, we calculated the coefficient of variation to understand the "mean-normalized" with our $\beta_v$ estimate. We found that "less noisy" experiments in fact showed lower $\beta_v$, not supporting the hypothesis that measurement-error is attenuating our parameter estimates. We then considered the relationship between $D$ and $V$ and $\beta_V$. Larger intervals $v_i$ $d_i$ correlated with smaller $B_v$, but so did the standard error $\hat{\sigma}_{D-V}^2 = var(d_i - v_i)$. So, we calculated the coefficient of variation for $D - V$, and found no relationship with $B_v$ estimates. In conclusion, though we identified relationships between $\beta_V$ and various population statistics including $\sigma_V$, accounting for the understood mean-variance relationship between distributions of vigorous movement span, increased variation is related to increased (not attenuated) $\beta_v$ estimates. Therefore, though we cannot fully exclude the role of measurement error, and eagerly await improvements in measurement accuracy that will definitively resolve the issue, we find no evidence to substantiate theoretic concerns that measurement errors make $\beta_v$ deviate from $\alpha$ in our analysis.
4.2 Simulating the effect of measurement error and environmental effects

To demonstrate our ability to distinguish between causal models (2) and (3) in the presence of environmental noise, we performed a series of simulations.

Variables shared in both simulations

\[ V^\dagger \sim \text{Weibull}(\beta = 12, \alpha = 5.75) \]
\[ R \sim \text{Gamma}(\beta = 70, \alpha = 1/70) \]
\[ Z \in \{1, 2, 3, \ldots, N\} \]
\[ \beta_Z \sim \text{Gamma}(\beta = 50, \alpha = 1/50) \]

Simulation of Causal Model (2)

\[ X^\dagger \sim \text{Weibull}(\beta = 3, \alpha = 2.5) \]
\[ v_i = r_i * v_i^\dagger \]
\[ x_i = x_i * x_i^\dagger \]
\[ d_i = v_i + x_i \]
\[ v_i^\ast = v_i * (\beta_Z Z_i) \]
\[ d_i^\ast = d_i * (\beta_Z Z_i) \]

Simulation of Causal Model (3)

\[ D^\dagger \sim \text{Weibull}(\beta = 15, \alpha = 7.2) \]
\[ v_i = r_i v_i^\dagger \]
\[ d_i = r_i d_i^\dagger \]
\[ v_i^\ast = v_i * \beta_Z Z_i \]
\[ d_i^\ast = d_i * \beta_Z Z_i \]

The "true" underlying sources of biological variability \( V^\dagger, X^\dagger, \) and \( D^\dagger \) are modeled as Weibull distributions with parameters chosen to approximate our empiric data. The causal effect of \( R \) is modeled by multiplying each individual's "true" \( v_i \) and \( x_i \) (model (2)) or \( d_i \) (model (3)) by a value drawn from a gamma distribution with a mean of 1. \( R \) thereby introduces a correlation between \( V X \), and \( D \) in model (2), and between \( V \) and \( D \) in model (3). Batch effects, meant to mimic the influence of environmental factors, were introduced by randomly assigning each individual \( i \) membership to a batch \( Z \). Each batch produces a characteristic effect shared among individuals in that batch. The "observed" \( V^\ast \) and \( D^\ast \) analyzed in the simulation are obtained by multiplying \( v_i \) and \( d_i \) by the batch effect \( \beta_Z Z_i \) characteristic of batch \( Z_i \). Note that here we choose to model noise and environmental variation as multiplicative factors in contrast to previous sections, as multiplicative noise is likely to more closely match the physical effect of temperature and other factors on event times (Stroustrup et al 2016.)

First, we performed a simulation in which paired \( v_i \) and \( d_i \) were generated as samples from the \( V \) and \( D \) as defined above for each model. The regression model Eq. 1 in S3 Text was fit to estimate \( \beta_v \), allowing us to evaluate the performance of our statistical approach in the absence of batch effects. Then, we performed the same simulation but this time comparing \( v_i^\ast \) and \( d_i^\ast \) generated as samples from \( V^\ast \) and \( D^\ast \) defined above for each model. The regression model \( 3 \) was fit to estimate \( \beta_v \), allowing us to evaluate the performance of our statistical approach in the presence of environmental confounders. In both sets of simulations, 1000 replicates were performed each considering a population of 1000 simulated individuals. In addition to estimates of \( \beta_v \), we obtained bootstrapped p-values for the test \( H_0 : \beta_v < 1 \). We summarize these results in Table B in S4 Text providing the average \( \beta_v \) obtained across all replicates as well as the fraction of p-values obtained that fell below a \( p = .05 \) threshold.

<table>
<thead>
<tr>
<th>Data Generating Model</th>
<th>Batch effects added?</th>
<th>( \beta_v )</th>
<th>Fraction of replicates with significant ( \alpha &lt; 1 ) at ( p &lt; .05 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model (2)</td>
<td>No</td>
<td>0.999</td>
<td>.044</td>
</tr>
<tr>
<td>Model (2)</td>
<td>Yes</td>
<td>1.000</td>
<td>.056</td>
</tr>
<tr>
<td>Model (3)</td>
<td>No</td>
<td>.520</td>
<td>1.000</td>
</tr>
<tr>
<td>Model (3)</td>
<td>Yes</td>
<td>.519</td>
<td>0.985</td>
</tr>
</tbody>
</table>

Table B in S4 Text: Differentiating simulated causal models.
These results show that regardless of environmental effects, we can differentiate between causal models (1) and (2). Causal model (2) generates data matching the expected relationship of $D = \alpha V + \epsilon$ with $\alpha = 1$. Our false positive rate of our bootstrapping test for $\beta_v \neq 1$ matches the expected rate of 5%. Causal model (3) generates data matching the expected relationship $D = \alpha V + \epsilon$ with $\alpha < 1$. The with a false negative rate at this effect magnitude appears at or below 2.5%.

### 4.3 Power Simulation

To estimate our power to identify $\alpha \neq 1$ and the sample sizes required to detect the deviations we observe, we performed a simulation. Using the data presented in Fig 2 of the main text which describes 2346 wildtype individuals, we sub-sampled the data set, selecting an equal number of individuals from each batch (device). Across 1000 replicates at each sample size, we then measured the relationship between $\beta_v$ when estimated using a single regression model Eq. 1 in S3 Text or the full multiple regression model 2. In each replicate, we calculated the boot-strapped p-value for $\beta_v \neq 1$ for the single and multiple regression models, and estimated the proportion of replicates for which this p-value was significant at $p < .05$.

![Graph](image)

**Fig C in S4 Text:** Subsampling our empiric data to estimate parameter bias and statistical power as a function of sample size. Multiple regression to account for batch effects (black) and the naive single regression (red).

We found that failure to account for batch effects biased the estimation of $\beta_v$ toward 1 at all population sizes. However, given the small overall magnitude of effect, large populations were required to obtain marginal statistical power, with a power of 0.60 being reached only when 1000 individuals were considered.