Simulating physiological flexibility in the acute glucocorticoid response to stressors reveals limitations of current empirical approaches

Conor Taff [Corresp. 1, 2]

[1] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, United States

[2] Lab of Ornithology, Cornell University, Ithaca, NY, United States

Corresponding Author: Conor Taff
Email address: cct63@cornell.edu

Wild animals often experience unpredictable challenges that demand rapid and flexible responses. The glucocorticoid mediated stress response is one of the major systems that allows vertebrates to rapidly adjust their physiology and behavior. Given its role in responding to challenges, evolutionary physiologists have focused on the consequences of between-individual and, more recently, within-individual variation in the acute glucocorticoid response. However, empirical studies of physiological flexibility are severely limited by the logistical challenges of measuring the same animal multiple times. Data simulation is a powerful approach when empirical data are limited, but has not been adopted to date in studies of physiological flexibility. In this paper, I develop a simulation that can generate realistic acute glucocorticoid response data with user specified characteristics. Simulated animals can be sampled continuously through an acute response and across as many separate responses as desired, while varying key parameters. Using the simulation, I develop several scenarios that address key questions in physiological flexibility. These scenarios demonstrate the conditions under which a single glucocorticoid trait can be accurately assessed with typical experimental designs, the consequences of covariation between different components of the acute stress response, and the way that context specific differences in variability of acute responses can influence the power to detect relationships between the strength of the acute stress response and fitness. I also describe how to use the simulation tools to aid in the design and evaluation of empirical studies of physiological flexibility.

1   Simulating physiological flexibility in the acute glucocorticoid response to stressors reveals

2   limitations of current empirical approaches

5   Conor Claverie Taff[1,2]

7   [1] Department of Ecology & Evolutionary Biology, Ithaca, NY, USA

8   [2] Lab of Ornithology, Ithaca, NY, USA

10   Corresponding Author:

11   Conor Taff

12   Ithaca, NY 14850

13   Email address: cct63@cornell.edu

16 **Abstract**

18 Wild animals often experience unpredictable challenges that demand rapid and flexible

19 responses. The glucocorticoid mediated stress response is one of the major systems that allows

20 vertebrates to rapidly adjust their physiology and behavior. Given its role in responding to

21 challenges, evolutionary physiologists have focused on the consequences of between-individual

22 and, more recently, within-individual variation in the acute glucocorticoid response. However,

23 empirical studies of physiological flexibility are severely limited by the logistical challenges of

24 measuring the same animal multiple times. Data simulation is a powerful approach when

25 empirical data are limited, but has not been adopted to date in studies of physiological flexibility.

26 In this paper, I develop a simulation that can generate realistic acute glucocorticoid response data

27 with user specified characteristics. Simulated animals can be sampled continuously through an

28 acute response and across as many separate responses as desired, while varying key parameters.

29 Using the simulation, I develop several scenarios that address key questions in physiological

30 flexibility. These scenarios demonstrate the conditions under which a single glucocorticoid trait

31 can be accurately assessed with typical experimental designs, the consequences of covariation

32 between different components of the acute stress response, and the way that context specific

33 differences in variability of acute responses can influence the power to detect relationships

34 between the strength of the acute stress response and fitness. I also describe how to use the

35 simulation tools to aid in the design and evaluation of empirical studies of physiological

36 flexibility.

38 *Keywords: acute stress response, physiological flexibility, glucocorticoids, evolutionary*

39 *endocrinology*

41 **Introduction**

43 Animals live in a dynamic environment in which they regularly encounter unpredictable

44 challenges. Successfully navigating these challenges often requires the ability to rapidly adjust

45 behavior and physiology to match current conditions. For vertebrates, the glucocorticoid

46 mediated stress response plays a major role in coordinating these changes when stressors are

47 encountered (Sapolsky et al., 2000; Wingfield et al., 1998) and similar rapid response systems

48 mediate changes in other taxa (Taborsky et al., 2021). Because of the central role that this

49 response plays in coping with challenges, a great deal of research effort over the past 15 years

50 has focused on understanding whether between-individual differences in the magnitude of this

51 response predict coping ability and, ultimately, fitness (Breuner et al., 2008; Schoenle et al.,

52 2020).

54 More recently, a series of conceptual papers have asked whether the degree of within-individual

55 variation in glucocorticoid modulation (i.e., endocrine flexibility) across different contexts or in

56 response to different stressors might also be an important predictor of performance (Hau et al.,

57 2016; Lema & Kitano, 2013; Taff & Vitousek, 2016; Wada & Sewall, 2014). Perhaps the major

58 limit to empirical progress, especially for within-individual variation, is the logistical difficulty

59 of accurately characterizing the functional shape of the acute physiological stress response for an

60 individual during a single acute response and across multiple acute responses occurring under

61   different conditions. Often these measures are strictly limited by the number of samples that can

62   safely be taken from an animal during a single capture and the number of repeated captures that

63   are possible (but see Koolhaas et al., 2011). Given these limitations, data simulation is a

64   powerful tool that could complement empirical work in this area, but that has not yet been

65   applied to studies of endocrine flexibility.

66

67   Several recent papers have suggested that physiologists interested in endocrine flexibility should

68   adopt a within-individual reaction norm approach (e.g., Hau et al., 2016; Taff & Vitousek, 2016).

69   This approach has been widely adopted in studies of behavioral flexibility where statistical

70   methods and empirical progress have developed synergistically (e.g., Araya-Ajoy et al., 2015;

71   Dingemanse et al., 2010; Westneat et al., 2015). This field has also benefited from simulation

72   studies to evaluate optimal study design (Pol, 2012) and packages that can create artificial

73   datasets with desired patterns of between, within, and residual variance to evaluate the

74   consequences of different patterns of variation on the ability to detect effects (see SQuID

75   package, Allegue et al., 2017). While these approaches are powerful, they have proven difficult

76   to apply directly to endocrine flexibility data for two reasons. First, simulation studies suggest

77   that successfully modeling within-individual variation in flexible traits using an hierarchical

78   modeling framework often requires a level of repeated sampling that is possible for many

79   behaviors (especially when collected autonomously), but that is currently not possible for most

80   studies of endocrine flexibility, because it would require sampling of many separate

81   glucocorticoid responses per individual. Second, many behavioral papers focus on somewhat

82   discrete measures (e.g., aggression score or activity level), whereas for acute glucocorticoid

83   responses, the functional shape of the response itself may be the important trait. Fully describing

84 the functional shape of a single acute glucocorticoid increase may require many samples in close

85 succession, but for small vertebrates logistical and ethical constraints mean that it is rarely

86 possible to take more than a few samples during the course of a single acute response.

87

88 The function valued trait (FVT) framework is an alternative approach that explicitly considers

89 the functional shape of a biological response (Gomulkiewicz et al., 2018; Kingsolver et al., 2015;

90 Stinchcombe et al., 2012). While FVT approaches have been suggested for studies of endocrine

91 flexibility, I am not aware of any papers that have applied this framework to empirical data on

92 acute glucocorticoid responses, probably because sufficient data are not available. Conceptually,

93 however, this approach is a better match to the acute glucocorticoid response, because the shape

94 of a response curve is explicitly considered as the phenotypic trait of interest. In some cases, it

95 may make sense to estimate particular parameters of the curve (e.g., maximum rate of increase

96 and maximum value reached) and then treat those parameters as phenotypic values for

97 downstream analysis, although statistical methods also exist to analyze the shape of the entire

98 curve directly without the need to extract discrete parameters (Kingsolver et al., 2015). This

99 approach has been used to study a variety of phenotypes where values can be measured

100 continuously or pooled across many individuals from the same group to accurately estimate the

101 shape of a curve (see Table 1 in Stinchcombe et al., 2012). Applying the technique to endocrine

102 flexibility at the within-individual level faces the same empirical challenges described for within-

103 individual reaction norms above, such as the need for repeated sampling of individuals and high

104 temporal resolution of samples within individual physiological responses. Note that FVT and

105 within-individual reaction norms approaches are not necessarily incompatible, but they have

106 largely developed separately.

107

108 The recognition that characterizing the functional shape of an acute stress response is

109 challenging goes back to the earliest studies conducted in wild animals. Early studies often

110 employed various control groups and sampled individual animals at a variety of time points over

111 a long period in order to describe the full response curve for a particular group (e.g., a species or

112 a breeding stage, Wingfield et al., 1992). These validations were considered essential to

113 characterize key parameters of the acute response for each group being studied (i.e., baseline,

114 rate of increase, maximum level, time of peak, and area under the curve; John Wingfield,

115 *personal communication*). Indeed, there is a long and rich history of empirical work

116 characterizing differences in each aspect of the acute glucocorticoid response (Breuner et al.,

117 1999; e.g., J. F. Cockrem & Silverin, 2002; Love et al., 2003; Wingfield et al., 1992) and

118 empirical data has contributed to a variety of conceptual models of acute glucocorticoid

119 regulation (Romero et al., 2009; e.g., Wingfield et al., 1998). More recently, mathematical

120 models have been used to generate predictions about how flexibility and variation in different

121 parts of the response system might respond to particular conditions (Grindstaff et al., 2022; e.g.,

122 Luttbeg et al., 2021; Taborsky et al., 2021). However, the challenge of estimating these

123 parameters and of decomposing within- and between-individual variance becomes much more

124 difficult when trying to fully describe the response for an individual animal rather than for a

125 group, because glucocorticoids can often only be measured at two or three time points and only a

126 small number of times per animal (e.g., Vitousek et al., 2018). Because these studies require an

127 estimate for each individual, the solutions used by older studies that added additional animals to

128 allow for sampling at more time points are not available.

129

130   For individual based studies, the most common approach to this problem is to standardize

131   measurements as much as possible by measuring animals at the same time of the day during the

132   same context, and by taking blood samples at standard times (often <3 and 30 minutes after

133   capture) to characterize baseline and stress-induced glucocorticoids. This standardization allows

134   for comparison between individuals, but in some cases it may also completely obscure the ability

135   to detect variation in certain characteristics of the acute response curve. For example, if the speed

136   (rate of initial increase or time required to reach maximum) and scope (maximum value) of the

137   acute response vary independently, samples taken at only two time points cannot accurately

138   capture variation in either parameter (Taff et al., In Press). Indeed, several discussions in recent

139   years about methods such as the '3 minute rule' and the relative merits of 'area under the curve'

140   versus time point measures of glucocorticoids are fundamentally related to a recognition of the

141   importance of understanding variation in the functional shape of stress responses and whether

142   different components of that shape covary within individuals (e.g., J. F. Cockrem & Silverin,

143   2002; Small et al., 2017). While a great deal of empirical work has focused on characterizing the

144   rate of initial increase (J. Cockrem, 2005; J. Cockrem et al., 2006), there has been relatively little

145   work on understanding individual differences in the time required to reach maximal levels, and

146   this attribute of speed is harder to estimate (Taff et al., In Press).

147

148   One of the characteristics of both the within-individual reaction norm and FVT literature is that

149   empirical work has proceeded in very close coordination with simulation and statistical method

150   development. In contrast, studies of endocrine flexibility often point to these methods, but don't

151   address the ways that the particular logistical challenges of hormone measurement, such as the

152   difficulty of repeatedly collecting samples within- or between-individuals, might necessitate

153   different empirical approaches. I believe this is one reason that there are currently more

154   conceptual papers arguing for a reaction norm approach to endocrine variation than there are

155   empirical papers actually applying the approach (but see, Fürtbauer et al., 2015; Houslay et al.,

156   2022; Lendvai et al., 2014; Taff et al., In Press). While many of the tools developed in these

157   related fields are transferable, studies of physiological flexibility would benefit from a focus on

158   analysis development and testing that explicitly incorporates the particular details and challenges

159   of these questions. One way to accomplish these goals is to integrate empirical work with

160   simulations, but to my knowledge no studies of physiological flexibility have developed

161   simulations of the acute stress response that address the issues discussed above.

162

163   Data simulation is a powerful approach for several reasons. Because true parameter values (e.g.,

164   maximum glucocorticoid level) are known, it is possible to evaluate how well different study

165   designs and analytical choices perform in recovering true patterns and how sensitive those

166   designs are to different assumptions. Thus, simulation can tell us whether the study designs we

167   use can *in principle* detect the patterns we predict given realistic effect sizes. Simulated data can

168   also identify conditions under which current study designs will perform well or poorly. For

169   example, if simulations suggest that the baseline paired with stress-induced paradigm only works

170   well when the speed and scope of responses are positively correlated, then empirical work could

171   seek to determine the degree of correlation for a particular study system as justification for the

172   approach. This ability to highlight key assumptions and create data sets with known properties

173   has the potential to both provide insight into physiological flexibility directly and to guide

174   empirical work by improving study design and identifying key areas for subsequent sampling. In

175   the rest of this paper, I develop a simple simulation of acute physiological stress responses and

176   then briefly illustrate several possible applications of the simulation.

177

178   The goal of this simulation is to provide a flexible tool that can produce realistic datasets of

179   physiological flexibility for a variety of different systems and scenarios. As such, there are many

180   possible applications and here I briefly highlight a few possibilities. Initially, I demonstrate that

181   the simulation can produce datasets that are qualitatively similar to empirical data on acute

182   glucocorticoid responses. Next, I explore three specific scenarios with the simulation that

183   highlight challenges to empirical progress in this field. First, I ask how well maximum

184   glucocorticoid measures can be estimated with a single time point measure in populations that

185   differ systematically in features of the response. Second, I explore how different patterns of

186   covariation between maximum glucocorticoids and one measure of speed (the time to reach

187   maximum) influence the ability to reliably measure either component of the response. Third, I

188   ask how different amounts of between- and within-individual variation in maximum

189   glucocorticoid levels impact the ability to detect known relationships between glucocorticoids

190   and fitness. These questions are all important for understanding physiologically flexibility and

191   are difficult to fully address empirically, but this is by no means an exhaustive list of the

192   questions or possible parameter permutations that could be explored with simulation. Finally, I

193   demonstrate how the simulation tools can be used to evaluate the performance of different

194   experimental designs given a set of logistical limitations (e.g., number of samples that can be

195   collected) and discuss how this procedure could be used as a planning tool to increase the power

196   and reproducibility of future empirical work.

197 **Materials & Methods**

198

199 *Description of the simulation*

200

201 I developed a set of functions in R version 4.0.2 (R Core Team, 2020) to generate acute

202 physiological response curves. This simulation makes no assumptions about the mechanistic

203 process that results in the shape of a glucocorticoid response. Rather, parameters are sampled to

204 generate curves that are similar in shape and degree of variation to empirically observed

205 responses (Figure 1). This simulation is designed to create data sets with realistic structure that

206 can be used to better design and plan studies of physiological flexibility, to evaluate power of

207 current study designs, and to evaluate the sensitivity of sampling regimes to any number of

208 modifications to the shape of glucocorticoid response curves (e.g., changing covariation patterns

209 between different features of the response). I explore a small number of scenarios in the next

210 section, but I expect that many other scenarios can be addressed with these tools. For illustration

211 purposes, I refer to simulated glucocorticoid responses, but the simulation applies equally well to

212 any physiological mediator of a rapid response. The code for the complete simulation along with

213 full documentation of each function and argument can be accessed in a GitHub repository with a

214 current version permanently archived on Zenodo (https://doi.org/10.5281/zenodo.6784207). The

215 package can be installed directly within the R environment using the following command.

216 devtools::install_github("cct663/simcoRt")

217

218 The simulation is constructed as two main functions with several minor functions for

219 downstream analysis. Detailed descriptions of the arguments to each function are included with

220   the package documentation. Briefly, function cort_sim1 samples the parameters shown in Figure

221   1 from an arbitrary number of animals. These parameters are sampled from a multivariate normal

222   distribution with user specified mean, variance, and covariance for each parameter. By default,

223   maximum glucocorticoid values are sampled from a normal distribution on the log scale and then

224   exponentiated to determine absolute values. This results in a right skewed maximal

225   glucocorticoid distribution on the absolute scale that is typical of many empirical datasets, but

226   users can easily specify any parameters to be sampled from a normal or log normal distribution

227   as required to match the characteristics of a particular study system. For the purposes of the

228   simulation, I consider the sampled parameter values to be the 'true,' unobserved, phenotype of

229   the animal (setting aside the question of whether or not a 'true' physiological phenotype exists).

230

231   A second function, cort_sim2, starts with a population of animals generated from cort_sim1 and

232   samples observed acute glucocorticoid responses an arbitrary number of times for each animal.

233   Two sources of variation in the observed relative to true parameter values can be specified. First,

234   within-individual variation in expression is represented by specifying what amount of variation

235   in the observation of each parameter is determined by the true value and what amount is

236   determined by an additional randomly sampled response, based on the population parameters

237   (this additional sampling maintains the user specified covariance structure of the population).

238   After sampling the parameters, values are interpolated for each one minute time point and a

239   localized regression is fit to create a smoothed curve that represents the observed glucocorticoid

240   response. From this expressed response, individual data points are then collected at user

241   specified times that would reflect an empirical study design (e.g., 1, 30, and 60 minutes).

242   Additional noise can be added to these data points to represent measurement error (e.g., assay

243   error). This simulated dataset can then be treated as the input for any desired analyses and

244   statistical approaches, while maintaining the ability to compare results to the 'true' values used

245   in the simulation.

246

247   The function also generates a simple simulated performance (e.g., fitness) measure, based on the

248   underlying true parameter values sampled for each individual in the population (e.g., their

249   baseline and maximum glucocorticoid value). The single fitness measure per animal is

250   determined by allowing the user to specify the relative degree to which unmeasured traits plus

251   each true parameter contribute to fitness outcomes. Data reflecting the true phenotypic values,

252   the repeated expression of acute responses, and the observed time points can then be used in

253   downstream analyses with any standard statistical approaches or software. For example, a user

254   could perform an analysis to ask whether a known relationship between fitness and a particular

255   true parameter is recovered in a study that includes only measures taken at particular time points.

256   An additional convenience function summarizes the output of a simulation run in a multi-panel

257   plot (Figure 2).

258

259   Finally, given recent interest in estimating the repeatability of glucocorticoid regulation (J. F.

260   Cockrem, 2013; Hau et al., 2016; Taff et al., 2018), I also included a function that takes input

261   from cort_sim2 and calculates the observed repeatability of several measures using package rptR

262   (Stoffel et al., 2017). Full details are included in the package documentation, but this function

263   returns repeatability for each individual time point specified in the down sampled data set,

264   profile repeatability (Reed et al., 2019), and repeatability for area under the curve calculated as

265   both increase (AUC_I) and ground (AUC_G) approaches (Pruessner et al., 2003). For each AUC

266   measure, the function returns repeatability for the full time course, for an estimate using only the

267   observed values in the down sampled data set, and for the full data set constrained to the time

268   period encompassing the observed data points. Simple plots illustrating repeated samples from

269   the same individuals are also returned by default. I do not develop an example of repeatability in

270   this manuscript, but the functions here could be used to determine the impact of different study

271   design choices on repeatability estimates.

272

273   *Methods for each simulated scenario*

274

275   Each of the scenarios described here uses the basic simulation functions described above with

276   input parameters adjusted to address the question of interest. A complete set of reproducible code

277   to create all of the examples presented here is available on GitHub and permanenty archived at

278   Zenodo (https://doi.org/10.5281/zenodo.6784203).

279

280   *Scenario 1: Simulating empirically parameterized data*

281

282   In order for simulation to be useful, we should be able to create artificial datasets that have

283   similar characteristics to empirical data for different systems. Simulating realistic data provides a

284   starting point for evaluating different study designs and the consequences of changes in different

285   assumptions or parameters. Simulating realistic data is also useful because it can aid in study

286   design or be used as a basis for pre-registered reports that demonstrate the feasibility of a

287   planned study before data are ever collected. Simulated data can be created and entered in a

288   complete analysis pipeline, with empirical data substituted later. In addition to helping to design

289   better studies, this approach has the advantage of increasing the transparency and reliability for

290   studies of physiological flexibility, by making analysis choices and predictions clear before data

291   are collected.

292

293   To demonstrate this utility, I attempted to create simulated datasets that recreate the

294   characteristics of the empirical data presented in Koolhaas et al. (2010). As part of that study, a

295   series of corticosterone measurements were collected during and after an acute stressor from 14

296   laboratory rats *Rattus norvegicus* using permanently implanted jugular vein canulae. I extracted

297   data from Figure 6 in Koolhaas et al. (2010) using WebPlotDigitizer

298   (https://automeris.io/WebPlotDigitizer/) and then simulated data using the functions described

299   above starting with the input values calculated directly from the empirical data. I then compared

300   patterns of variation and population level response curves for the empirical and simulated data.

301

302   *Scenario 2: Accurately measuring a single glucocorticoid trait*

303

304   Single time point measures of glucocorticoids are often interpreted as representing meaningful

305   variation between individuals. For example, variation in the level of glucocorticoids after 30

306   minutes of standardized restraint is typically interpreted as variation in the magnitude of the

307   stress response (Taff et al., 2019). However, this interpretation rests on assumptions that are

308   rarely explicitly tested with empirical data. For example, the time chosen to take a stress-induced

309   sample is often assumed to be either at the species peak or during a plateau period after the

310   species peak. In some early studies, great care was taken to determine an average population

311   level peak time (e.g., Wingfield et al., 1992), but many studies adopt the widely used 'standard'

312   time of 30 minutes post capture without extensive validation (compiled in Vitousek et al., 2019).

313

314   While there is a general assumption that sampling later than the peak is acceptable (and perhaps

315   preferable) because animals will be sampled during a relatively stable high plateau, there is little

316   empirical data to evaluate this assertion or to determine how much under or overshooting the

317   species peak timing might influence inferences. Furthermore, even when the average peak timing

318   is well established, differences in the amount of between-individual variation in the time to reach

319   the peak or in peak values are common across species and even in different life history stages

320   within species (Wingfield et al., 1992). Some studies focus instead on the rate of initial increase

321   in glucocorticoids with a sample taken at 10 or 15 minutes after disturbance (e.g., J. Cockrem,

322   2005; J. Cockrem et al., 2006; Love et al., 2003), often well before peak levels are reached. The

323   combination of different patterns of within- and between-individual variation with the exact

324   time(s) chosen for sampling could have consequences for the accuracy of point estimates taken at

325   any single time point, but these questions cannot be addressed directly with empirical datasets

326   where the true underlying values of each individual are unknown.

327

328   I simulated a simple scenario to explore the consequences of variation in each of these

329   parameters on the accuracy of estimating between individual differences in maximally expressed

330   glucocorticoids during an acute response. For purposes of this illustration, I consider a single

331   study design in which animals are sampled at 30 minutes. Using this design as a starting point, I

332   systematically vary i) the timing of the population average peak (15, 30, or 45 minutes), ii) the

333   amount of variation in maximum glucocorticoid levels reached (SD of 1 to 12 ng/μl), iii) and the

334   amount of variation in the number of minutes taken to reach peak levels (SD of 1 to 20 minutes).

335   All other variables in the simulation are constrained to be invariant between individuals in the

336   population (e.g., all individuals have identical baseline glucocorticoids in this case), though I

337   consider cases in which multiple aspects of the rapid response are correlated with each other in

338   the next scenario. I included moderate within-individual variability and a small amount of assay

339   error across all iterations. For each combination of parameters, I simulated 200 animals and

340   estimated the R^2 value from a regression of the observed estimates of glucocorticoid levels at 30

341   minutes to the true known values. This simulation is likely a best case scenario for detection

342   because it eliminates many sources of variation or noise that would be present in real data, but it

343   illustrates the effect of variation in these three key parameters even when the exact same

344   sampling design is employed.

345

346   *Scenario 3: Exploring covariance between response components*

347

348   In reality, fully characterizing the acute glucocorticoid response requires more than identifying

349   just the maximum value reached. Individuals may differ in baseline levels, rate of initial

350   increase, the speed of reaching the maximum level, time spent at maximum, and the speed of

351   return to baseline. Moreover, each of these components of the endocrine response could be

352   positively or negatively correlated with each other within and between individuals. In these

353   cases, measurements taken at particular time points contain information about multiple aspects of

354   the response and without additional information it may be difficult to know what trait is being

355   measured. The fact that each of these traits might be important and that they might covary has

356   been discussed in a general sense (e.g., Baugh et al., 2013), but simulations are uniquely

.

357 powerful for exploring under exactly what conditions time point measure of glucocorticoids can

358 or cannot be used as indicators of these traits.

359

360 To illustrate this point, I explored the consequences of variation in the correlation between and

361 relative amount of variation in just two aspects of the acute stress response: the maximum

362 glucocorticoid level reached and the time required to reach the maximum level. For simplicity, I

363 refer to the 'speed' of the response, but note that other aspects, such as the rate of initial increase,

364 could also be considered as variation in the speed of response. When considering these two traits,

365 a population of animals could plausibly display one of three patterns. Individuals that reach their

366 maximum value faster might also reach higher values (simulation correlation = -0.6).

367 Alternatively, the speed and maximum values might vary independently (correlation = 0).

368 Finally, individuals that are faster responders might max out at lower glucocorticoid values

369 (correlation = 0.6). While many researchers in this field might have intuitions about which of

370 these scenarios is most likely to prevail, there is very little empirical data available to actually

371 determine which is most common. Moreover, regardless of the specifics for this particular

372 correlation, the general pattern and considerations presented here will apply in similar ways to

373 correlations between other aspects of the acute stress response.

374

375 Using these three simulated populations as a starting point, I asked how well glucocorticoid

376 values measured at one timepoint reflected true trait values. For each population I set an average

377 population level speed of 30 minutes with other values in the simulation set at their default value.

378 For every time point from 0 to 35 minutes I fit two simple linear regressions of the measured

379 value on the true speed and maximum value and extracted the R^2 value from the model. I

380   repeated this simulation for all populations 50 times with 100 individuals sampled from the

381   population each time. Finally, I repeated the entire set of simulations with each combination of

382   low and high between-individual variation in the speed or maximum values (variation in speed:

383   low = 2 minute SD, high = 12 minute SD; variation in maximum: low = 1ng/μl SD, high = 10ng/

384   μl SD).

385

386   *Scenario 4: Detecting links between fitness and responses*

387

388   A common goal of recent studies is to establish whether variation in glucocorticoids is associated

389   with fitness or some proxy for fitness (Schoenle et al., 2020). While there has been a great deal

390   of discussion about the extent to which these relationships might differ with life history

391   characteristics or between breeding stages, there has been relatively little consideration of the

392   way that methodological limitations might limit the ability to detect these relationships even

393   when they exist.

394

395   Here, I imagine a simple scenario in which the 'true' maximum glucocorticoid level during an

396   acute response explains 80% of the variation in fitness (clearly this is unrealistically high, but it

397   is chosen for illustration only). I next construct a study in which researchers measure 50

398   individuals using a typical stress-induced (30 minute) sampling protocol. For simplicity, I set the

399   other parameters in the simulation at their default values. Keeping the study design constant, I

400   ask whether the glucocorticoid-fitness relationship can be recovered for two hypothetical

401   populations that have low or high between-individual variation in maximum glucocorticoid

402   levels. Using these populations, I asked asked how the ability to detect glucocorticoid-fitness

403 relationships changed with different amounts of within-individual variation in acute response

404 expression. For each combination of parameters, I simulated 50 populations and fit a simple

405 linear regression model with observed glucocorticoid levels at 30 minutes as a predictor of

406 fitness to ask whether the true glucocorticoid-fitness relationship was recovered.

407

408 *Scenario 5: Designing optimal sampling strategies*

409

410 One of the major benefits of simulating glucocorticoid response curves will be the ability to

411 design optimal sampling strategies before data are collected. A simulation can be constrained to

412 match any real world limitations (e.g., maximum number of samples possible per individual) and

413 then explored to determine how to best allocate sampling resources. The specifics of this task

414 will vary considerably with the study system and question being addressed, but here I illustrate

415 one possible application. Consider an experiment in which the acute glucocorticoid response of a

416 treatment group and control group are compared after some experimental manipulation. The

417 details of the manipulation are unimportant here, but suppose that the prediction is that this

418 manipulation should result in a difference in the speed of the corticosterone response between

419 our two groups, such that the treatment group will reach it's maximum glucocorticoid value

420 faster than the control group, but will not differ in the maximum value itself. I have implemented

421 this difference by simulating two populations in which the treatment group has a steeper initial

422 slope and also reaches the maximum value faster. Any number of possible hypotheses for a

423 particular study system could be specified following a similar approach.

424

425   Using these simulated groups, I asked how well different study designs could detect the

426   differences. Here we can impose any logistical constraints relevant to the study system. As an

427   example, in this case we can only sample a maximum of 20 individuals per group, we can only

428   sample each individual once post-treatment, and during that single sampling event we can take a

429   blood sample at a maximum of three different time points, resulting in a total of 120 data points.

430   Given these constraints, I compare several sampling designs. First, I explored 'standard'

431   sampling strategies that are typical of empirical studies in this field that sample each individual

432   animal at either i) 1 minute, 30 minutes, and 60 minutes, ii) 1 minute, 15 minutes, and 30

433   minutes, or iii) 1 minute, 15 minutes, and 60 minutes. Second, I explored two alternative

434   sampling approaches that are not typically used in empirical work but might better capture the

435   entire functional shape of response. These included, iv) a study in which three sampling times

436   between 1 and 60 minutes are randomly chosen for every animal, and v) a study in which three

437   sampling times are randomly chosen for each animal, but weighted more heavily around the

438   range of times when maximum levels are expected to be reached for the population.

439

440   For illustration purposes I sampled directly from the 'true' response curves in this example so

441   that there is no additional measurement error added. To evaluate these schemes I compare

442   estimates of the acute response curve for each group to the 'true' known curves. Note that a more

443   complete analysis of a sampling schemes performance should include many more iterations and

444   full statistical comparisons, but the details here will be highly dependent on the study system and

445   goals, so I provide this simple example to illustrate the approach rather than to make any more

446   widely applicable conclusions.

447

448 **Results**

449

450 *Scenario 1: Simulating empirically parameterized data*

451

452 The simulation functions were able to produce a new synthetic dataset that has similar variation

453 and patterns to the empirical data in Koolhaas et al. (2010) (Figure 3A). The average population

454 wide response curve shape also closely matched the empirical data (Figure 3B). In this case, the

455 plotted simulation data include the same number of animals sampled at the same time points as

456 the empirical data, but these sampling points and total sample size can easily be changed as

457 desired. The parameterized simulation can now be used to test the sensitivity of any number of

458 experimental designs before additional data is collected, such as different sample sizes or

459 sampling time points.

460

461 *Scenario 2: Accurately measuring a single glucocorticoid trait*

462

463 The amount of between-individual variation in maximum glucocorticoid values has a profound

464 effect on the ability to detect true maximal levels with samples taken at 30 minutes (Figure 4). A

465 single sample taken at 30 minutes was highly correlated with true maximal glucocorticoid levels

466 when between-individual variation was high, but with low between-individual variation in

467 maximum a single sample was uninformative. I simulated a wide range of variation that may

468 include unrealistically high and low values, but the result highlights the importance of

469 considering the amount of variation expected in a study population. There is a weaker, but still

470 substantial impact of variation in the time taken to reach maximum values on the accuracy of

471   estimates in this simulation. Greater variation in the speed (time to reach maximum) of the

472   response reduces the accuracy of estimates of maximal values. Finally, the timing of sampling

473   relative to the average population peak timing also influences accuracy. Measuring after the

474   average peak time results in the most accurate estimates across a range of parameter values,

475   while measuring before the average peak time produces the least accurate measures, particularly

476   when there is also high variation in the time to reach maximum values between individuals.

477

478   *Scenario 3: Exploring covariance between response components*

479

480   Simulated populations had either a negative correlation between the time required to reach

481   maximum values and the maximum glucocorticoid level reached (Figure 5A), no correlation

482   (Figure 5B), or a negative correlation (Figure 5C). Using these populations, the time that samples

483   were taken at, relative amount of variation in speed and maximum, and degree of correlation

484   between the speed and maximum all had substantial impacts on the ability to infer true trait

485   values from single time point glucocorticoid measures (figure 6). Neither speed nor maximum

486   traits could be assessed accurately when between-individual variation in both traits was low

487   (figure 6A). Overall, accurately assessing variation in speed was much harder—if not

488   impossible—with single measures.

489

490   It was only possible to accurately estimate speed when high between-individual variation in

491   speed was coupled with low variation in maximal values, but this situation may be rare in natural

492   populations. When speed was tightly correlated with maximum (figure 6D) it was sometimes

493   possible to attain reasonable estimates of speed (figure 6C-D), but when speed was not correlated

494   with maximum, single measures were not good indicators of variation in speed (figure 6A, C-D).

495   Finally, measuring variation in maximum values was much easier under many conditions (figure

496   6C-D), but the accuracy of assessment of maximum values was also negatively impacted by

497   variation in speed and the degree of this impact differed depending on the correlation between

498   the two traits (figure 6).

499

500   *Scenario 4: Detecting links between fitness and responses*

501

502   Several patterns can be identified by examining the results of this simulation. First, as specified

503   by the simulation parameters, the correlation between the true maximum glucocorticoid value

504   and fitness does not differ for populations simulated with high or low between-individual

505   variation (figure 7A-B). In all cases, however, the observed correlation is lower than the true

506   correlation and always lowest in the population with low between-individual variation. The

507   ubiquity of this pattern is a product of the simulation structure, because adding within-individual

508   variation effectively adds noise to the true correlation. It is important to note that in the real

509   world, it is unlikely that this pattern would be so universal, because unmeasured variables could

510   influence both fitness and glucocorticoids. For example, if habitat quality directly alters fitness

511   and glucocorticoids, the observed correlation could be stronger than the 'true' correlation. Thus,

512   interpretation of these results should be made cautiously in light of the simplicity of the

513   simulation compared to real world conditions. Nevertheless, general patterns illustrated by the

514   simulation are likely to pertain across a wide range of conditions.

515

516   It was easiest to detect statistical evidence for known glucocorticoid-fitness relationships when

517   within-individual variation was low (figure 7A). Under these conditions, the true pattern was

518   recovered in nearly all simulated populations with high between-individual variation and in

519   approximately half of the populations with low between-individual variation. It becomes harder

520   to detect these true relationships within-individual variation is high (figure 7B) are high, but even

521   in these more challenging situations the relationship can be detected the majority of the time if

522   between-individual variation in maximum levels is high. When within-individual variation is

523   high and between-individual variation is low, it is nearly impossible to detect glucocorticoid-

524   fitness relationships.

525

526   *Scenario 5: Designing optimal sampling strategies*

527

528   In this simulated scenario, standard sampling schemes cannot fully describe the difference

529   between treatment groups (Figure 8; Figure 9A-C). Sampling at 1, 30, and 60 minutes

530   completely fails to distinguish the difference between groups, despite the fact that the treatment

531   group reaches it's maximum value on average 12 minutes (~40%) faster than the control group.

532   Many empirical studies include an earlier sample at 15 minutes to capture the rate of initial

533   increase Silverin (1998). Shifting one of the sampling time points in this simulation to 15

534   minutes does recover a clear treatment difference in initial increase, but cannot capture the large

535   difference in the time required to reach maximum values between the groups 9B-C).

536

537   In contrast, both the random sampling and weighted sampling schemes detect differences in the

538   overall shape of the acute response (Figure 9 D-E). In this particular scenario, there is no clear

539   difference between these two approaches and both perform well in describing both the difference

540   in initial increase and the difference in the time required to reach maximal levels in each group.

541

542   **Discussion**

543

544   I demonstrate that simple simulation tools can produce datasets that closely match empirically

545   observed acute glucocorticoid responses. Once developed, simulations can then be used to

546   explore a wide range of hypothetical scenarios and to guide empirical studies. I explore a few

547   possible scenarios with a limited range of parameters here. Nevertheless, even the simple

548   demonstrations included in this paper suggest several ways that simulation could help move

549   studies of physiological flexibility forward. One of the main benefits of simulating datasets is

550   identifying unmeasured properties and assumptions of currently available data that can become

551   targets for future empirical work.

552

553   For example, I demonstrate that the ability to accurately measure a single glucocorticoid trait of

554   interest (maximum glucocorticoid level) is strongly impacted by the amount of between-

555   individual variation in the population and more subtly influenced by between-individual

556   variation in speed (the time required to reach maximum levels) and by the timing of sample

557   collection relative to the average population peak timing. In one sense, this result is unsurprising

558   because it is intuitive that large differences in maximum glucocorticoids should be easier to

559   detect, but there are important consequences of this fact for interpreting studies that seek to link

560   between-individual variation in the magnitude of the stress response with other traits. The

561   magnitude of the acute stress response often varies substantially across life history stages

562   (Wingfield et al., 1992). Even if study designs are identical it will be easier to accurately

563   measure individual glucocorticoid traits under some conditions with the same level of sampling.

564   This simple scenario shows that the same sampling regime will perform better or worse

565   depending on the combination of glucocorticoid regulation parameters in the population being

566   studied for statistical, rather than biological, reasons.

567

568   The pattern of within-individual covariation between separate aspects of the acute glucocorticoid

569   response (e.g., speed and maximum values) can also have important effects on the interpretation

570   of empirical results. I found that covariation between speed and magnitude of the acute stress

571   response and the relative amount of variation in each of these traits have profound effects on the

572   ability to accurately measure either single component. This result occurs because measures taken

573   at any given time point reflect a mixture of speed and magnitude. Empirical work specifically

574   designed to assess covariation and variance at different times could help to understand what

575   conclusions we can reasonably draw from available data. This simulation also clearly

576   demonstrates that measuring timing components of the acute glucocorticoid response (e.g., the

577   time required to reach maximum) is much more challenging over a wide range of conditions than

578   measuring maximal levels. Although a great deal of empirical work has focused on estimating

579   differences in the initial rate of increase in glucocorticoids (J. Cockrem, 2005; e.g., J. F.

580   Cockrem, 2007; Wingfield et al., 1992), much less is known about between-individual variation

581   in the time required to reach maximal levels.

582

583   Beyond the specifics of this particular example, what these results demonstrate is that

584   understanding what aspect of the glucocorticoid response is being measured by any particular

585   study design depends on extensive knowledge of the overall shape and amount of variation in

586   different aspects of the acute stress response. There is still a great need for observational

587   sampling to characterize within- and between-individual characteristics of glucocorticoid

588   responses, particularly when new species or contexts are being studied. An increased focus on

589   descriptive data collection can subsequently contribute to more targeted and powerful hypothesis

590   tests and better parameterized simulations.

591

592   I also showed that the amount of within- and between-individual variation in glucocorticoid traits

593   directly impacts the ability to detect statistical support for known relationships between the trait

594   of interest and fitness. The fact that low between-individual variation and high within-individual

595   variation in maximum glucocorticoids make it harder to detect true glucocorticoid-fitness

596   relationships across a wide range of conditions has important consequences for interpreting

597   empirical results. Many studies have demonstrated different relationships (or lack thereof)

598   between glucocorticoids and fitness at different life history stages (Bonier et al., 2009; Vitousek

599   et al., 2018), but it is also well known that the absolute amount of between individual variation in

600   glucocorticoid traits varies considerably at different stages (Wingfield et al., 1992). My

601   simulation demonstrates that the power to detect true relationships will differ drastically across

602   these conditions even with identical study designs and samples sizes; thus, apparent differences

603   in glucocorticoid-fitness relationships across seasons or stages can easily arise as statistical

604   artefacts when between-individual variation in hormones also differs across contexts. Great care

605   may be needed to conclusively differentiate true differences in these relationships from statistical

606   artefacts. Of course, it may be common for physiological traits have stronger direct impacts on

607   fitness in some contexts than others (Bonier et al., 2009), but simulation guided studies can help

608   to ensure that the statistical approach to detect differences is equally powerful in different

609   contexts.

610

611   Finally, I show that many standard sampling schemes perform poorly in some situations and for

612   certain questions. A few clear takeaways can be derived from these results. First, while strict

613   standardization of the timing of samples has some clear advantages, it also comes with costs and

614   likely makes it nearly impossible to detect certain types of variation between groups or

615   individuals. It should be clear that no amount of additional sampling would allow that approach

616   to detect the difference in time to reach maximum glucocorticoid levels developed in the

617   scenario here. Second, while it may be very difficult to accurately estimate the full shape of the

618   acute stress response for *individuals*, alternative sampling schemes with random or weighted

619   sample timing make it possible to describe these shapes accurately for groups (e.g., treatments,

620   species, different contexts) even without extraordinarily large sample sizes. A similar argument

621   about the power of randomly timed sampling has been put forward in the function valued trait

622   literature (Gomulkiewicz et al., 2018), but this type of sampling scheme is rarely used in

623   evolutionary endocrinology research (but see, Vitousek et al., In Press). It is perhaps

624   unsurprising that empirical papers that have emphasized the importance of different time courses

625   (rather than only maximum) of the stress response often focus on between group comparisons or

626   investigate variation in the exact sampling time between individuals (e.g., Baugh et al., 2013;

627   Small et al., 2017). These alternative sampling schemes also come with drawbacks (e.g., the

628   inability to directly compare individuals sampled at different times) and I don't suggest that they

629   are universally better options than standardization; rather, different schemes will perform better

630   or worse given the specific scenario and the exact hypothesis being tested.

631

632   The simulation of different sampling schemes in particular addresses a single specific scenario,

633   but a similar scenario could be designed for any number of studies and any number of

634   predictions about how the speed, scope, or other attributes of the glucocorticoid response are

635   expected to change with a treatment or between different groups or species. Clearly, when

636   estimating the timing of peak glucocorticoids, a simple baseline plus induced sampling scheme is

637   sub optimal, but this scheme may perform well in other situations where the maximum value is

638   the target and there is relatively little variation in response time. Creating simulations like this

639   before studies are conducted has the potential to increase the efficient use of researchers time and

640   funds, but also forces researchers to think explicitly about quantitative predictions ahead of time.

641   These simulations could be included as part of a study pre-registration, grant application, or

642   registered report to demonstrate exactly what data collection and analysis approaches are planned

643   and to justify those decisions. Across a wide range of disciplines there has been an increasing

644   push for pre-registration, reproducible research, and transparent research practices (O'Dea et al.,

645   2021). Simulation provides an opportunity for evolutionary endocrinologists to embrace these

646   best practices by improving the quality of study design, allowing for more quantitative

647   hypotheses and predictions, and providing a clear justification for experimental choices.

648

649   While there has been increasing interest in understanding within- and between-individual

650   variation in the acute glucocorticoid response in recent years (Hau et al., 2016; Lema & Kitano,

651   2013; Taff & Vitousek, 2016; Wada & Sewall, 2014), the methods and data available to tackle

652   these questions have changed relatively little. Many sophisticated statistical tools are now

653   available and clear arguments have been made about the need to apply these approaches to

654   endocrine traits, but relatively few empirical studies have effectively used these tools (but see,

655   Fürtbauer et al., 2015; Houslay et al., 2022). Arguably, the biggest roadblock at the moment is

656   the limited availability of empirical data needed to test hypotheses. Although there is a rich

657   history of empirical work focused on understanding the acute glucocorticoid response (J. F.

658   Cockrem & Silverin, 2002; Romero, 2002; Vitousek et al., 2019; e.g., Wingfield et al., 1992),

659   many current questions require a level of repeated sampling within- and between-individuals that

660   is difficult to achieve. Simulation offers one way forward, by allowing for more efficiently

661   designed studies, by allowing researchers to identify when the question of interest can *in*

662   *principle* be answered with a given study design, and by providing a tool to translate recent

663   theoretical advances (Grindstaff et al., 2022; e.g., Luttbeg et al., 2021; Taborsky et al., 2021) into

664   tractable, hypothesis driven empirical studies. Ideally, conceptual papers, empirical work, theory,

665   and simulation will proceed together to make progress in this field. The tools presented here only

666   scratch the surface of the ways that data simulation can be applied to address pressing questions

667   in evolutionary endocrinology.

668

669   **Acknowledgments**

670

671   I would like to thank Michaela Hau and her lab for discussions about the ideas presented in this

672   project. John Wingfield also provided an insightful overview of the history of study design in

673   field-based studies of the acute stress responses and pointed me to key references. Finally, I

674   thank Maren Vitousek and her lab members for feedback and discussion on early versions of this

675   project.

676

677 **References**

678

679 Allegue, H., Araya-Ajoy, Y. G., Dingemanse, N. J., Dochtermann, N. A., Garamszegi, L. Z.,

680 Nakagawa, S., Reale, D., Schielzeth, H., & Westneat, D. F. (2017). Statistical quantification of

681 individual differences (SQuID): An educational and statistical tool for understanding multilevel

682 phenotypic data in linear mixed models. *Methods in Ecology and Evolution*, *8*(2), 257–267.

683 Araya-Ajoy, Y. G., Mathot, K. J., & Dingemanse, N. J. (2015). An approach to estimate short-

684 term, long-term and reaction norm repeatability. *Methods in Ecology and Evolution*, *6*(12),

685 1462–1473.

686 Baugh, A. T., Oers, K. van, Naguib, M., & Hau, M. (2013). Initial reactivity and magnitude of

687 the acute stress response associated with personality in wild great tits (parus major). *General and*

688 *Comparative Endocrinology*, *189*, 96–104.

689 Bonier, F., Moore, I. T., Martin, P. R., & Robertson, R. J. (2009). The relationship between

690 fitness and baseline glucocorticoids in a passerine bird. *General and Comparative*

691 *Endocrinology*, *163*(1-2), 208–213.

692 Breuner, C. W., Patterson, S. H., & Hahn, T. P. (2008). In search of relationships between the

693 acute adrenocortical response and fitness. *General and Comparative Endocrinology*, *157*(3),

694 288–295.

695 Breuner, C. W., Wingfield, J. C., & Romero, L. M. (1999). Diel rhythms of basal and stress-

696 induced corticosterone in a wild, seasonal vertebrate, gambel's white-crowned sparrow. *Journal*

697 *of Experimental Zoology*, *284*(3), 334–342.

698 Cockrem, J. (2005). Conservation and behavioral neuroendocrinology. *Hormones and Behavior*,

699 *48*(4), 492–501.

700   Cockrem, J. F. (2007). Stress, corticosterone responses and avian personalities. *Journal of*

701   *Ornithology*, *148*(2), 169–178.

702   Cockrem, J. F. (2013). Individual variation in glucocorticoid stress responses in animals.

703   *General and Comparative Endocrinology*, *181*, 45–58.

704   Cockrem, J. F., & Silverin, B. (2002). Variation within and between birds in corticosterone

705   responses of great tits (parus major). *General and Comparative Endocrinology*, *125*(2), 197–206.

706   Cockrem, J., Potter, M., & Candy, E. (2006). Corticosterone in relation to body mass in adelie

707   penguins (pygoscelis adeliae) affected by unusual sea ice conditions at ross island, antarctica.

708   *General and Comparative Endocrinology*, *149*(3), 244–252.

709   Dingemanse, N. J., Kazem, A. J., Réale, D., & Wright, J. (2010). Behavioural reaction norms:

710   Animal personality meets individual plasticity. *Trends in Ecology & Evolution*, *25*(2), 81–89.

711   Fürtbauer, I., Pond, A., Heistermann, M., & King, A. J. (2015). Personality, plasticity and

712   predation: Linking endocrine and behavioural reaction norms in stickleback fish. *Functional*

713   *Ecology*, *29*(7), 931–940.

714   Gomulkiewicz, R., Kingsolver, J. G., Carter, P. A., & Heckman, N. (2018). Variation and

715   evolution of function-valued traits. *Annual Review of Ecology, Evolution, and Systematics*, *49*,

716   139–164.

717   Grindstaff, J. L., Beaty, L. E., Ambardar, M., & Luttbeg, B. (2022). Integrating theoretical and

718   empirical approaches for a robust understanding of endocrine flexibility. *Journal of*

719   *Experimental Biology*, *225*(Suppl_1), jeb243408.

720   Hau, M., Casagrande, S., Ouyang, J. Q., & Baugh, A. T. (2016). Glucocorticoid-mediated

721   phenotypes in vertebrates: Multilevel variation and evolution. *Advances in the Study of Behavior*,

722   *48*, 41–115.

723 Houslay, T. M., Earley, R. L., White, S. J., Lammers, W., Grimmer, A. J., Travers, L. M.,

724 Johnson, E. L., Young, A. J., & Wilson, A. (2022). Genetic integration of behavioural and

725 endocrine components of the stress response. *Elife*, *11*, e67126.

726 Huber, N., Mahr, K., Tóth, Z., Szarka, E. Z., Çınar, Y. U., Salmón, P., & Lendvai, Á. Z. (2021).

727 The stressed bird in the hand: Influence of sampling design on the physiological stress response

728 in a free-living songbird. *Physiology & Behavior*, *238*, 113488.

729 Kingsolver, J., Diamond, S., & Gomulkiewicz, R. (2015). Curvethinking: Understanding

730 reaction norms and developmental trajectories as traits. *Integrative Organismal Biology*.

731 *Hoboken (NJ): Wiley Blackwell*, 39–54.

732 Koolhaas, J. M., Bartolomucci, A., Buwalda, B., Boer, S. F. de, Flügge, G., Korte, S. M.,

733 Meerlo, P., Murison, R., Olivier, B., Palanza, P.others. (2011). Stress revisited: A critical

734 evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews*, *35*(5), 1291–1301.

735 Koolhaas, J. M., De Boer, S., Coppens, C., & Buwalda, B. (2010). Neuroendocrinology of

736 coping styles: Towards understanding the biology of individual variation. *Frontiers in*

737 *Neuroendocrinology*, *31*(3), 307–321.

738 Lema, S. C., & Kitano, J. (2013). Hormones and phenotypic plasticity: Implications for the

739 evolution of integrated adaptive phenotypes. *Current Zoology*, *59*(4), 506–525.

740 Lendvai, A. Z., Ouyang, J. Q., Schoenle, L. A., Fasanello, V., Haussmann, M. F., Bonier, F., &

741 Moore, I. T. (2014). Experimental food restriction reveals individual differences in

742 corticosterone reaction norms with no oxidative costs. *PLoS One*, *9*(11), e110564.

743 Love, O. P., Bird, D. M., & Shutt, L. J. (2003). Plasma corticosterone in american kestrel

744 siblings: Effects of age, hatching order, and hatching asynchrony. *Hormones and Behavior*,

745 *43*(4), 480–488.

746   Luttbeg, B., Beaty, L. E., Ambardar, M., & Grindstaff, J. L. (2021). Mathematical modeling

747   reveals how the speed of endocrine regulation should affect baseline and stress-induced

748   glucocorticoid levels. *Hormones and Behavior*, *136*, 105059.

749   O'Dea, R. E., Parker, T. H., Chee, Y. E., Culina, A., Drobniak, S. M., Duncan, D. H., Fidler, F.,

750   Gould, E., Ihle, M., Kelly, C. D.others. (2021). Towards open, reliable, and transparent ecology

751   and evolutionary biology. *BMC Biology*, *19*(1), 1–5.

752   Pol, M. van de. (2012). Quantifying individual variation in reaction norms: How study design

753   affects the accuracy, precision and power of random regression models. *Methods in Ecology and*

754   *Evolution*, *3*(2), 268–280.

755   Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas

756   for computation of the area under the curve represent measures of total hormone concentration

757   versus time-dependent change. *Psychoneuroendocrinology*, *28*(7), 916–931.

758   R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation

759   for Statistical Computing. https://www.R-project.org/

760   Reed, J. M., Harris, D. R., & Romero, L. M. (2019). Profile repeatability: A new method for

761   evaluating repeatability of individual hormone response profiles. *General and Comparative*

762   *Endocrinology*, *270*, 1–9.

763   Romero, L. M. (2002). Seasonal changes in plasma glucocorticoid concentrations in free-living

764   vertebrates. *General and Comparative Endocrinology*, *128*(1), 1–24.

765   Romero, L. M., Dickens, M. J., & Cyr, N. E. (2009). The reactive scope model—a new model

766   integrating homeostasis, allostasis, and stress. *Hormones and Behavior*, *55*(3), 375–389.

767 Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How do glucocorticoids influence

768 stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions.

769 *Endocrine Reviews*, *21*(1), 55–89.

770 Schoenle, L. A., Zimmer, C., Miller, E. T., & Vitousek, M. N. (2020). Does variation in

771 glucocorticoid concentrations predict fitness? A phylogenetic meta-analysis. *General and*

772 *Comparative Endocrinology*, 113611.

773 Silverin, B. (1998). Behavioural and hormonal responses of the pied flycatcher to environmental

774 stressors. *Animal Behaviour*, *55*(6), 1411–1420.

775 Small, T. W., Bebus, S. E., Bridge, E. S., Elderbrock, E. K., Ferguson, S. M., Jones, B. C., &

776 Schoech, S. J. (2017). Stress-responsiveness influences baseline glucocorticoid levels: Revisiting

777 the under 3 min sampling rule. *General and Comparative Endocrinology*, *247*, 152–165.

778 Stinchcombe, J. R., Kirkpatrick, M., Group, F. T. W.others. (2012). Genetics and evolution of

779 function-valued traits: Understanding environmentally responsive phenotypes. *Trends in Ecology*

780 *& Evolution*, *27*(11), 637–647.

781 Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: Repeatability estimation and

782 variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and*

783 *Evolution*, *8*(11), 1639–1644.

784 Taborsky, B., English, S., Fawcett, T. W., Kuijper, B., Leimar, O., McNamara, J. M.,

785 Ruuskanen, S., & Sandi, C. (2021). Towards an evolutionary theory of stress responses. *Trends*

786 *in Ecology & Evolution*.

787 Taff, C. C., Schoenle, L. A., & Vitousek, M. N. (2018). The repeatability of glucocorticoids: A

788 review and meta-analysis. *General and Comparative Endocrinology*, *260*, 136–145.

789   Taff, C. C., & Vitousek, M. N. (2016). Endocrine flexibility: Optimizing phenotypes in a

790   dynamic world? *Trends in Ecology & Evolution*, *31*(6), 476–488.

791   Taff, C. C., Wingfield, J. C., & Vitousek, M. N. (In Press). The relative speed of the

792   glucocorticoid stress response varies independently of scope and is predicted by environmental

793   variability and longevity across birds. *Hormones & Behavior*.

794   Taff, C. C., Zimmer, C., & Vitousek, M. N. (2019). Achromatic plumage brightness predicts

795   stress resilience and social interactions in tree swallows (tachycineta bicolor). *Behavioral*

796   *Ecology*, *30*(3), 733–745.

797   Vitousek, M. N., Houtz, J., Pipkin, M., Chang van Oordt, D., Hallinger, K., Uehling, J. J., &

798   Taff, C. C. (In Press). Natural and experimental cold exposure in adulthood increase the

799   sensitivity to future stressors in a free-living songbird. *Functional Ecology*.

800   Vitousek, M. N., Johnson, M. A., Downs, C. J., Miller, E. T., Martin, L. B., Francis, C. D.,

801   Donald, J. W., Fuxjager, M. J., Goymann, W., Hau, M.others. (2019). Macroevolutionary

802   patterning in glucocorticoids suggests different selective pressures shape baseline and stress-

803   induced levels. *The American Naturalist*, *193*(6), 866–880.

804   Vitousek, M. N., Taff, C. C., Hallinger, K. K., Zimmer, C., & Winkler, D. W. (2018). Hormones

805   and fitness: Evidence for trade-offs in glucocorticoid regulation across contexts. *Frontiers in*

806   *Ecology and Evolution*, *6*, 42.

807   Wada, H., & Sewall, K. B. (2014). Introduction to the symposium—uniting evolutionary and

808   physiological approaches to understanding phenotypic plasticity. *American Zoologist*, *54*(5),

809   774–782.

810   Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual

811   within-individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743.

812   Wingfield, J. C., Maney, D. L., Breuner, C. W., Jacobs, J. D., Lynn, S., Ramenofsky, M., &

813   Richardson, R. D. (1998). Ecological bases of hormone—behavior interactions: The "emergency

814   life history stage." *American Zoologist*, *38*(1), 191–206.

815   Wingfield, J. C., Vleck, C. M., & Moore, M. C. (1992). Seasonal changes of the adrenocortical

816   response to stress in birds of the sonoran desert. *Journal of Experimental Zoology*, *264*(4), 419–

817   428.

Figure 1

Conceptual illustration of the structure of the simulation.

Conceptual illustration of the structure of the simulation. For each simulated animal, seven parameters are sampled from a multivariate normal distribution. Together, these seven parameters define the turning points in an acute response curve. The mean and standard deviation for each parameter can be set along with the degree of covariation between each pair of parameters. Note that the simulation can easily be simplified as desired by setting some parameter mean or standard deviations to zero.
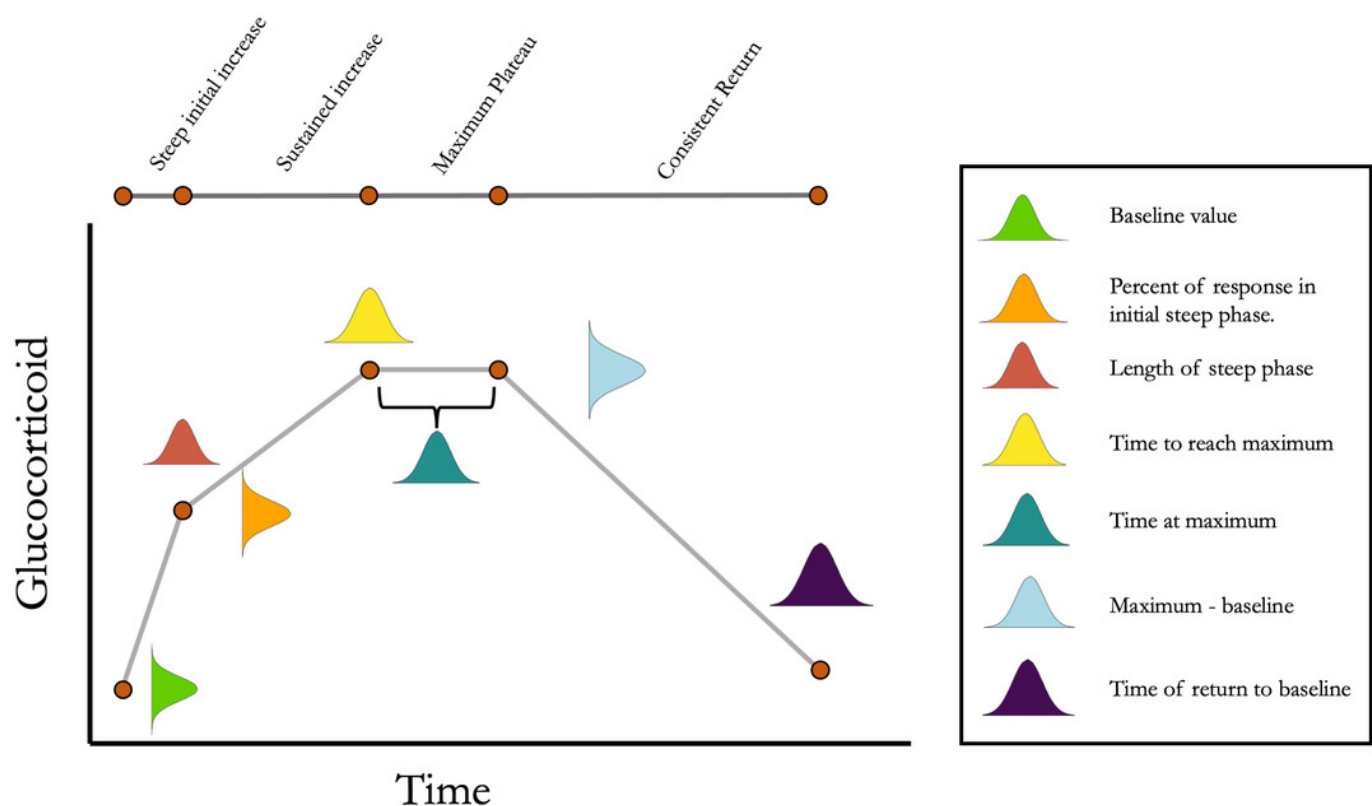
Figure 2

Example of simulation output with default settings.

Example of simulation output with default settings. Panel A shows the downsampled data set for this run with samples collected at 1, 15, and 30 minutes in this case. Panel B shows the full observed response curve for each animal. Panel C shows the rank order of glucocorticoid level at each time point for each animal. In each panel, the vertical dashed lines represent the three time points that might have been measured in a typical empirical study. Note that individuals in the top panels do not match perfectly because measurement error is added to the downsampled dataset in panel A.
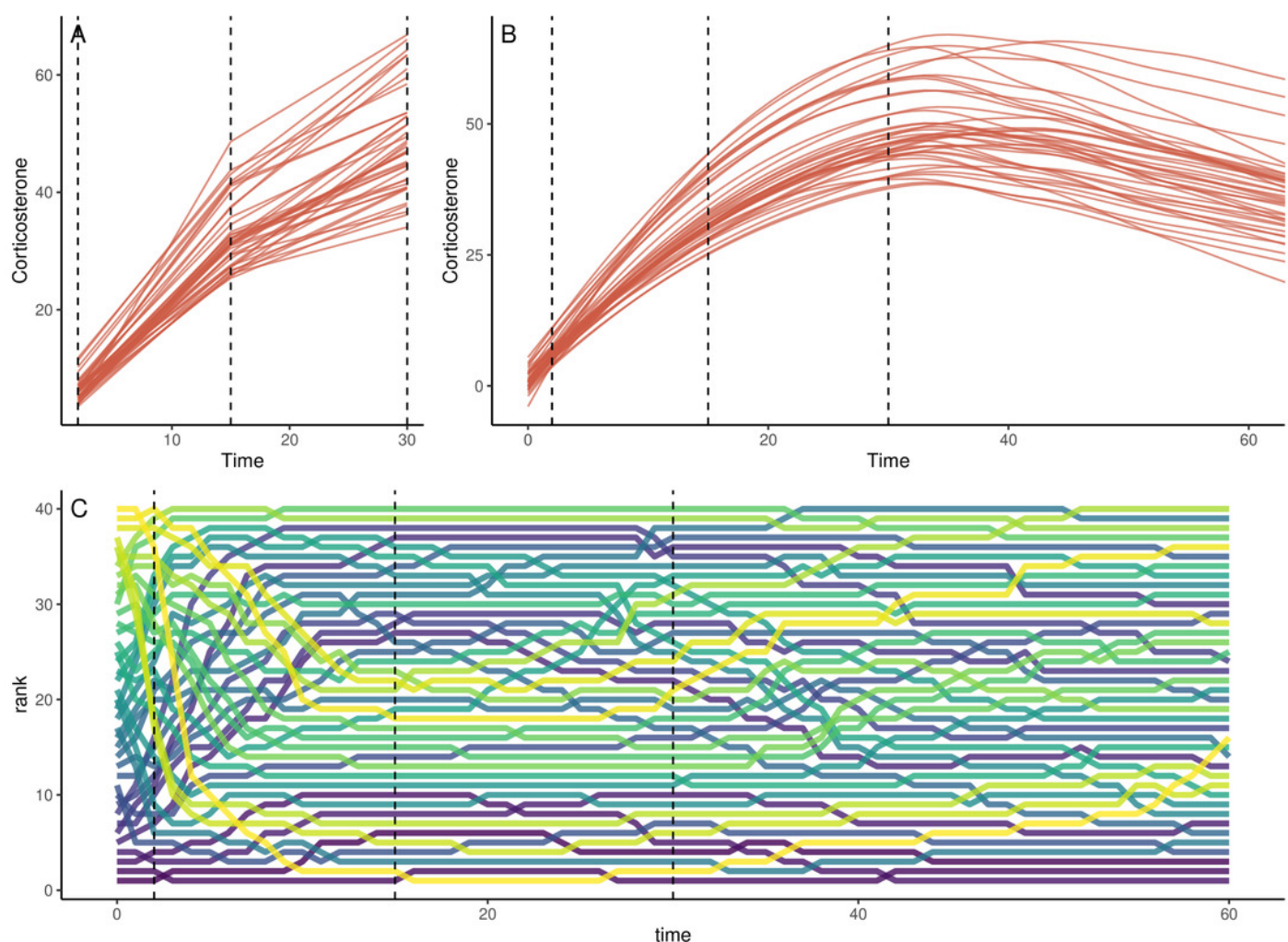
Figure 3

Simulated dataset matching empirical data.

Panel A shows the acute corticosterone response for measured (blue) or simulated (orange) rats measured at five time points. Panel B shows the mean and standard error of the two datasets. Empirical data are extracted from Koolhaas et al.2010 Figure 6 using the WebPlotDigitizer tool.
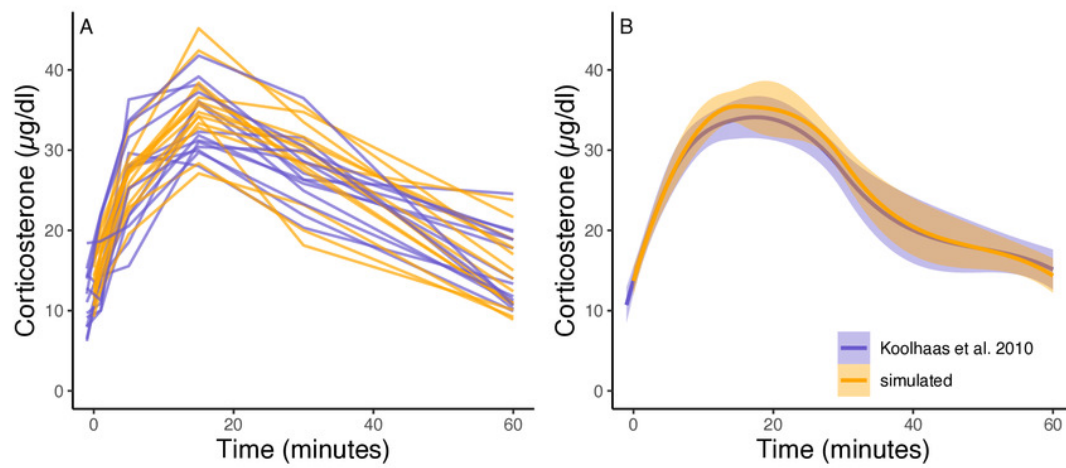
Figure 4

Results of simulation runs with different amounts of between-individual variation in the time to reach maximum glucocorticoid levels and in the maximum level reached.

Results of simulation runs with different amounts of between-individual variation in the time to reach maximum glucocorticoid levels and in the maximum level reached. Simulations are run with samples taken at 30 minutes on populations with an average peak time of 15 minutes (left), 30 minutes (center), or 45 minutes (right). Each grid cell is the R^2 value from the regression of observed glucocorticoids at 30 minutes to true maximum levels in a simulation of 200 individuals.
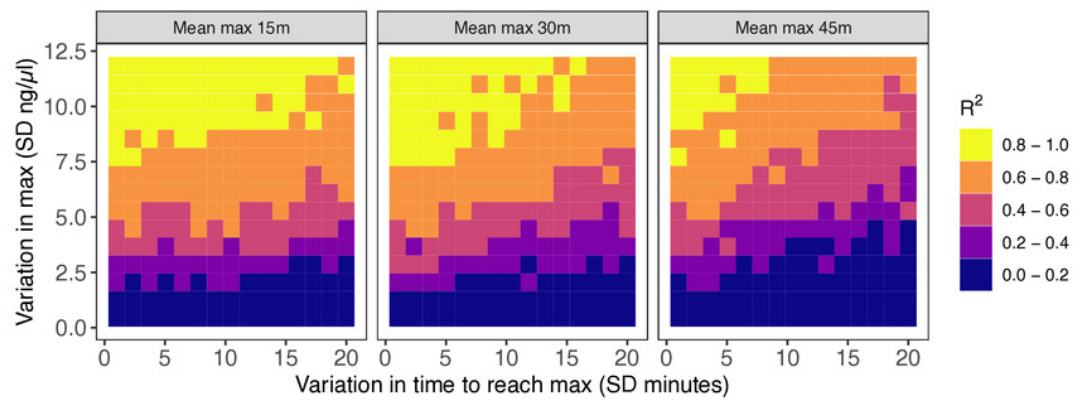
Figure 5

Varying correlation between speed and scope of response.

Simulated glucocorticoid responses in which the maximum value and response speed are positively correlated (A), uncorrelated (B), or negatively correlated (C). Orange curves show the full response for each individual. Blue points show the maximum value and time to reach maximum for each individual. Blue lines are simple linear regressions of speed and maximum value for each group. For clarity, only the first 40 individuals in each simulated dataset are plotted.
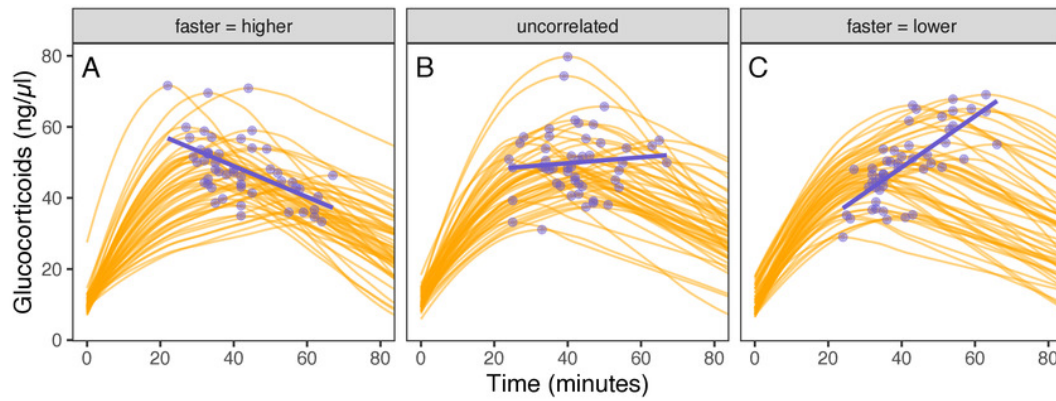
Figure 6

Simulating different amounts of variation in speed and scope of response.

Relationship between single time point measures of glucocorticoids and the true value of either maximum level (solid lines) or the speed of the glucocorticoid response (dashed lines). Panels show results when the overall variation in maximum values and speed are both low (A), when one is low while the other is high (B and C), and when both are high (D). In each panel, three different simulation scenarios illustrate the patterns when speed and maximum value are positively correlated (purple), uncorrelated (teal), or negatively correlated (yellow). Faded lines show the results from each of 50 separate simulation runs and thick lines are the averages across all runs.
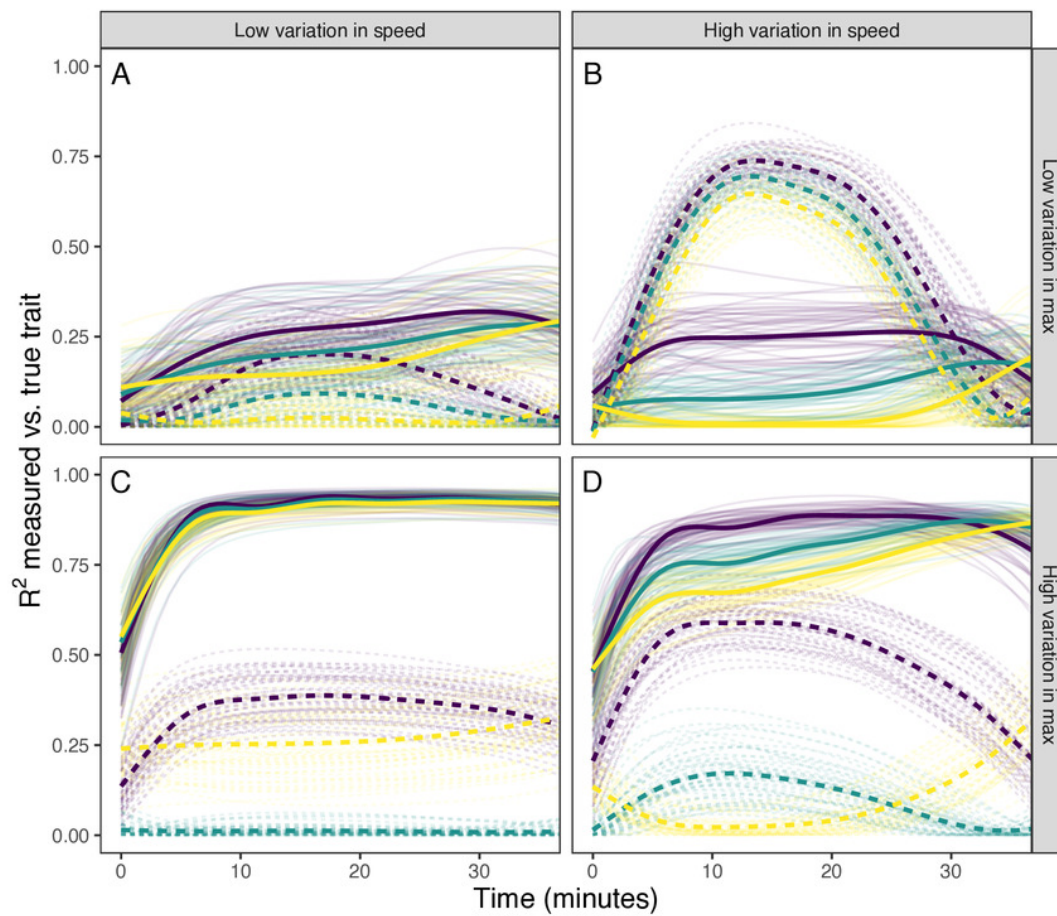
Figure 7

Variation in response and ability to detect fitness correlations.

Relationship between observed maximum glucocorticoid values and fitness for simulated populations that have low between-individual variation in maximum glucocorticoids (blue) or high between-individual variation in maximum glucocorticoids (red). Each point is the result of a separate simulation of 50 individuals using the settings described in the text. Filled circles are simulations in which observed glucocorticoid values at 30 minutes were significantly correlated with fitness and crosses are simulations in which the relationship was not significant. Panels illustrate conditions with low within-individual variation (A) versus high within-individual variation (B). For each simulation, the correlation between true maximum glucocorticoids fitness is plotted on the y-axis and the correlation with observed values is plotted on the x-axis.
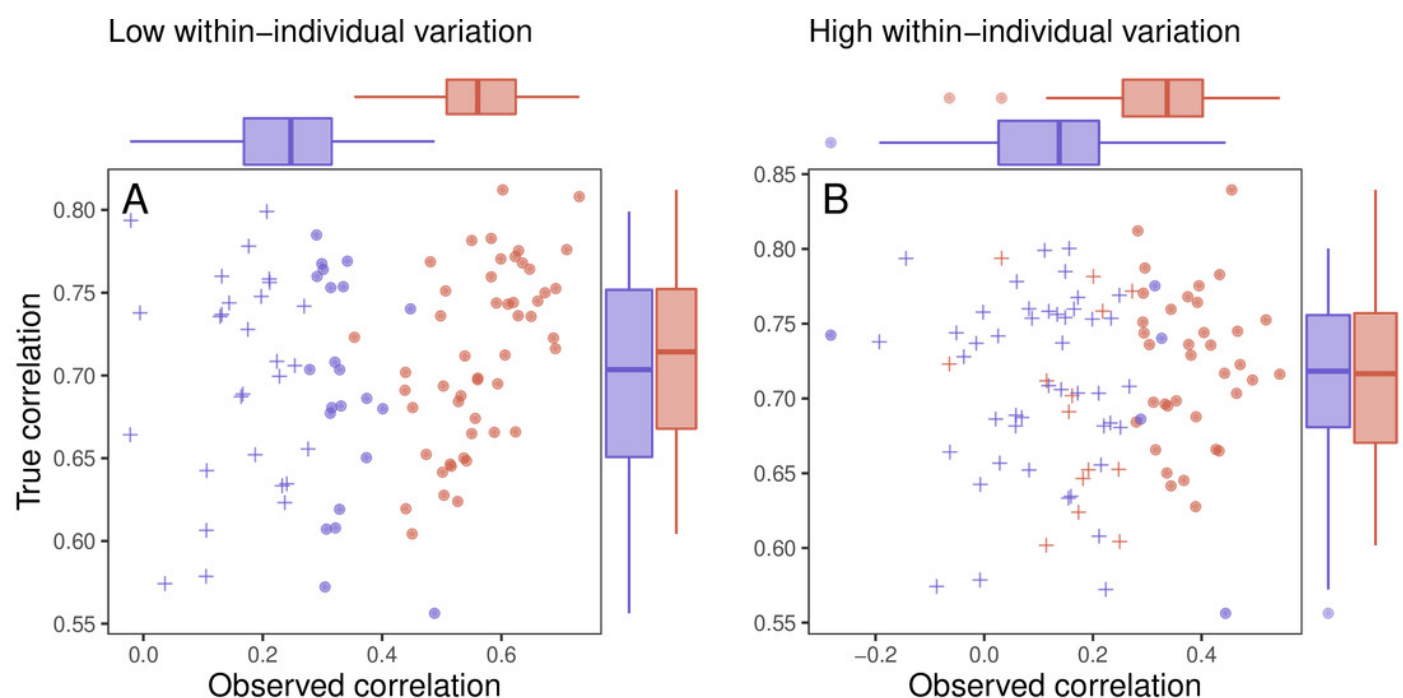
Figure 8

Simulated data for a hypothetical control (blue) and treatment (orange) group.

Faded thin lines show the acute response for each individual simulated (20 per group) and thick lines show the average response curve for each group.
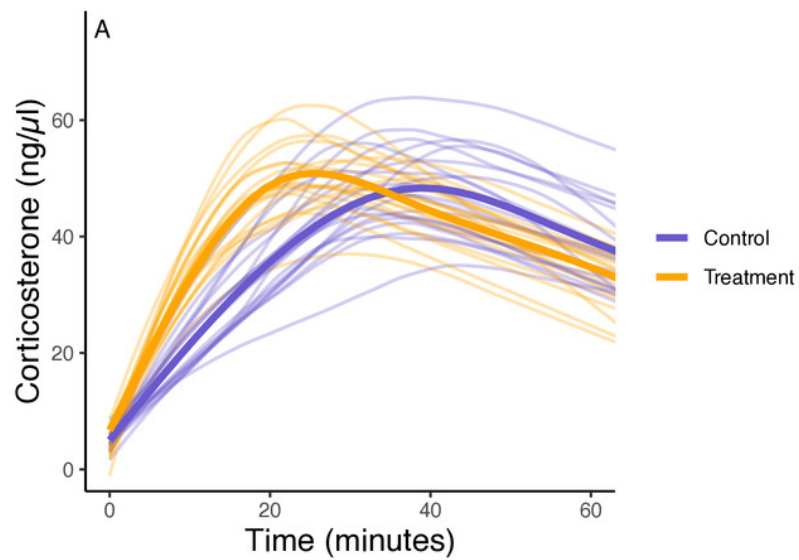
Figure 9

Three possible sampling schemes to compare two groups.

Five possible sampling schemes to compare two groups. For standard sampling (A-C), every individual is sampled at exactly 3 time points that include baseline along with samples at 15, 30, or 60 minutes. For random sampling (D) each individual is sampled at three random points between 1 and 60 minutes. For weighted sampling (E) three sampling times are chosen for each individual from a normal distribution with mean of 32 and sd of 9 minutes. In all panels, solid lines are the true group averages, dashed black lines are the estimates based on samples, shaded intervals are the 95% estimates from a simple generalized additive model, and points are individual samples collected.