General and Comparative Endocrinology 260 (2018) 136–145

Contents lists available at ScienceDirect

General and Comparative Endocrinology

journal homepage: www.elsevier.com/locate/ygcen

Research paper

The repeatability of glucocorticoids: A review and meta-analysis

Conor C. Taff[a,*], Laura A. Schoenle[b], Maren N. Vitousek[a]

[a] Lab of Ornithology and Department of Ecology & Evolutionary Biology, Cornell University, United States
[b] Department of Biological Sciences, Virginia Tech, United States

ARTICLE INFO

Article history:
Received 19 September 2017
Revised 27 November 2017
Accepted 11 January 2018
Available online 31 January 2018

Keywords:
Corticosterone
Cortisol
Fecal glucocorticoid metabolites
Within-individual consistency

ABSTRACT

Glucocorticoids are highly conserved hormones that mediate a suite of responses to changing conditions in vertebrates. Recent work has focused on understanding how selection operates on glucocorticoid secretion in natural populations. Because heritability is rarely estimated and difficult to measure in the wild, many studies report within-individual repeatability as an estimate of stable between individual differences in glucocorticoid secretion. We conducted a systematic review and meta-analysis on estimates of within-individual glucocorticoid repeatability to elucidate general patterns of repeatability, and to test for relationships between covariates and estimates of repeatability. To this end, we collected 203 estimates of within-individual glucocorticoid repeatability drawn from 71 separate studies and 55 species. Overall, we found moderate levels of repeatability (0.29). We also found that repeatability varied by sample type. Long-term measures (e.g., fecal and feather samples) and acute stress-induced plasma glucocorticoids had higher repeatability (long-term: 0.44, stress-induced: 0.38), than baseline glucocorticoid levels (0.18). Repeatability also decreased with increasing time between repeated sampling events. Despite significant overall repeatability, there was substantial heterogeneity in estimates from different studies, suggesting that repeatability of glucocorticoid secretion varies substantially across systems and conditions. We discuss the implications of our results for understanding selection on glucocorticoid traits and suggest that continuing work should focus on evaluating the repeatability of within-individual glucocorticoid reaction norms.

1. Introduction

Glucocorticoids are highly conserved vertebrate hormones that mediate a suite of functional responses to changing conditions over various time scales (Landys et al., 2006; Wingfield et al., 1998). Although responding appropriately to challenging conditions is critical in order to survive and reproduce, studying the evolution of glucocorticoid responses poses an empirical challenge (Bonier and Martin, 2016; Taff and Vitousek, 2016). While it is clear that selection should favor appropriate regulation of glucocorticoids, within-individual variation in secretion may mask any stable between-individual differences. Thus, it is often difficult to detect and interpret selection—or lack thereof—on variation in glucocorticoids (Bonier et al., 2009; Bonier and Martin, 2016). Indeed, glucocorticoids are so closely related to dynamic external conditions that many studies do not report analyses to detect stable between-individual differences in concentrations even when the same individuals are sampled repeatedly (e.g., many studies of primates listed in Table 1 of Cavigelli and Caruso, 2015). Despite the complications of measuring selection on highly flexible traits, a burgeoning body of literature has focused on understanding the way that selection operates on between-individual differences in glucocorticoid expression to produce appropriate phenotypic responses in changing conditions (Bonier et al., 2009; Bonier and Martin, 2016; Hau et al., 2016; Taff and Vitousek, 2016).

Glucocorticoid responses can clearly evolve in response to selection on secretion patterns (Pottinger et al., 1992). In the wild, many studies have looked for correlations between point estimates of glucocorticoid levels and survival or reproductive success; some of these studies demonstrate strong relationships, but patterns across studies are inconsistent (Bonier et al., 2009; Breuner et al., 2008; Sorenson et al., 2017). Further, it is often unclear from these studies whether between-individual differences in glucocorticoids are causal drivers of fitness outcomes, or whether external conditions alter both glucocorticoids and fitness directly (Bonier et al., 2009). Because most studies to date only follow one generation, they cannot estimate whether there is any evolutionary response to selection on glucocorticoid secretion. In captivity, strong selection regimes targeted at glucocorticoid secretion—or correlated traits—can generate lines that differ markedly in secretion patterns

* Corresponding author at: 159 Sapsucker Woods Road, Ithaca, NY 14850, United States.
E-mail address: cct63@cornell.edu (C.C. Taff).

https://doi.org/10.1016/j.ygcen.2018.01.011
0016-6480/

(e.g., Baugh et al., 2012; Evans et al., 2006; Pottinger and Carrick, 1999). However, it is unclear how similar these selection lines are to any selection regimes actually experienced in the wild. Only a few studies have directly measured heritability of glucocorticoid secretion under natural conditions (Jenkins et al., 2014; Stedman et al., 2017); these studies demonstrate low to moderate heritability of glucocorticoid traits.

Given the paucity of field-based estimates of heritability, many recent studies report the repeatability of glucocorticoid traits within individuals in order to demonstrate the existence of stable between-individual differences (e.g., Cockrem et al., 2016; Ouyang et al., 2011). When individuals are sampled at least twice, repeatability can be estimated as the intra-class correlation coefficient (ICC) by calculating the proportion of total variance in a trait that occurs between- rather than within-individuals (Lessells and Boag, 1987; Nakagawa and Schielzeth, 2010). Repeatability is sometimes reported using different methods (correlations, ANOVA, or linear mixed models [LMMs]) and can be reported as 'agreement repeatability'—the similarity of absolute values between repeated measures—or as 'adjusted repeatability'—the similarity between repeated measures after accounting for covariates (recent papers demonstrate that LMM's typically perform better and have several advantages over other estimation approaches; Baugh et al., 2014; Nakagawa and Schielzeth, 2010). Regardless of the methods used, repeatability measures are sometimes presented—implicitly or explicitly—as an upper bound estimate of heritability (sensu, Boake, 1989). Although this assertion holds under certain conditions, there are good reasons to believe that glucocorticoid measures will often violate these assumptions (Dohm, 2002). Still, repeatability provides a tractable, if imperfect, indication of stable between-individual differences in glucocorticoid secretion patterns. While some studies report high repeatabilities for glucocorticoid traits (e.g., Angelier et al., 2010; Narayan et al., 2013; Rogovin and Naidenko, 2011), others fail to detect any repeatability (e.g., Bridge et al., 2009; Pavitt et al., 2016; Tempel and Gutierrez, 2004). In some cases, low repeatability may not reflect a lack of between-individual differences, but rather the effect of heterogeneous external conditions, such as uncontrolled variation in food availability or temperature. Conversely, some reports of high repeatability may reflect persistent environmental differences between individuals, rather than stable phenotypic differences *per se* (i.e., pseudo-repeatability; Niemelä and Dingemanse, 2017). To date, no comprehensive review of the repeatability of glucocorticoid traits has been published, and it is difficult to interpret the generality of available estimates of glucocorticoid repeatability.

Here, we provide a systematic review and meta-analysis on estimates of glucocorticoid repeatability in vertebrates. Several recent papers have included discussions of glucocorticoid repeatability, but these qualitative reviews do not compile a comprehensive set of repeatability estimates or conduct any analyses to uncover general patterns about glucocorticoid repeatability (Cockrem, 2013; Cockrem et al., 2009; Hau et al., 2016; Ouyang et al., 2011). Holtmann et al. (2017) recently included estimates of corticosterone repeatability in a meta-analytical framework, but their analysis included only birds, grouped several hormone types (i.e., glucocorticoids, estrogens, and androgens) and sample substrates together, and was not focused on evaluating the repeatability of glucocorticoid secretion *per se*. In our analysis, we expand on the studies identified by Holtmann et al. (2017) to provide a broad overview of how repeatable glucocorticoid concentrations are in vertebrates. We also use this dataset to assess whether estimates of repeatability are correlated with relevant covariates, including taxon, sample size, baseline versus stress-induced measures, captive versus wild studies, and the sampling interval. Finally, we discuss the limitations of using repeatability as a proxy for evolvability in highly labile traits, and make some suggestions for future studies in this area.

2. Methods

2.1. Literature search

We conducted a meta-analysis on estimates of within-individual glucocorticoid repeatability in vertebrates. We searched for studies to include in our analysis using a combination of approaches. Initially, we consulted a recently published meta-analysis of hormonal, metabolic, and behavioral repeatability in birds (Holtmann et al., 2017); this study included estimates of corticosterone repeatability in birds from a literature search that was conducted in February 2015. We manually checked each entry from that database to confirm suitability for our purposes and extracted additional covariates to be used in our analyses (see below). For some studies included in Holtmann et al. (2017), repeatability estimates were not included in the originally published reports but communicated directly to the authors; we included these effect sizes in our dataset (using the values provided by Holtmann et al. (2017)), but did not confirm repeatability independently.

We added additional studies by conducting a literature search to find studies on birds that were published from 2015 to 2016 as well as studies on other vertebrate taxa published at any time. In addition to this broad search, we conducted both a backwards- and forwards-search of articles that were cited by—or cited—key papers that include discussions of glucocorticoid repeatability (e.g., Cockrem, 2013; Cockrem et al., 2009; Dantzer et al., 2010; Fletcher et al., 2015; Hau et al., 2016; Ouyang et al., 2011). For each of these approaches, we updated our searches to include papers that were published on or before December 31st, 2016.

2.2. Inclusion criteria

To be included in our analyses, studies had to meet six main criteria. First, we only included studies that reported repeatability as an intra-class correlation coefficient (ICC) using an ANOVA based (Lessells and Boag, 1987) or Linear Mixed Model (LMM) Based approach (Nakagawa and Schielzeth, 2010), a Spearman correlation, or a Pearson correlation (r). We were able to include a few studies where repeatability was not reported explicitly by calculating repeatability with data extracted from a scatterplot using WebPlotDigitizer version 3.9 (Ankit Rohatgi, Austin, Texas, USA). In these cases, we calculated repeatability using LMMs and the *rptR* package in R (Stoffel et al., 2017). The LMM approach is constrained to yield positive values; for low repeatabilities (<0.005), we followed Holtmann et al. (2017) in recalculating repeatabilities using an ANOVA based approach because these estimates better fit the normality assumptions of our meta-analytic models. Second, we excluded studies that were conducted on domesticated animals, humans, inbred lines of lab animals, or artificially selected strains (particularly strains selected for high or low glucocorticoid responsiveness). Third, we excluded studies that conducted experimental manipulations that could have influenced glucocorticoid expression, except in cases where a control group was reported separately. Fourth, we excluded studies that calculated repeatability based on samples that did not distinguish between baseline and stress-induced glucocorticoid levels. Fifth, we included studies that were conducted on adult animals only. Finally, we excluded studies that did not provide enough supporting details to be included in our analyses (e.g., number of measurements per individual, season that samples were collected, etc.).

2.3. Data collection

From studies that met the criteria described above, we recorded a number of covariates in addition to repeatability estimates. These covariates included taxon (birds, fish, mammals, amphibians, or reptiles), species, sex (male, female, or mixed/unknown), sample size, number of samples per individual, life history stage (breeding, non-breeding, or mixed), average interval between samples (in days), substrate that the hormone was measured in (blood, feather, feces, water, or urine), average latency between capture and sampling (in minutes), and whether the study was conducted in the wild or in captivity. Estimates from blood, urine and water were categorized as either 'baseline' or 'stress-induced,' those from feathers and feces were categorized as 'long-term,' as these samples are believed to reflect average glucocorticoid levels over a longer time period (hours to weeks, depending on the species/method). In a few cases, studies reported repeatability estimates for both 'stress-induced' and 'stress-response' values (i.e., absolute glucocorticoids vs. change from baseline); we only included estimates based on 'stress-induced' values, because this measure is more widely reported.

We relied on authors' reports to determine what constituted a baseline sample, but typically baseline plasma samples were collected within 3 min of capture in birds and mammals (sometimes longer in reptiles) while stress-induced plasma samples were collected after a standardized interval post-capture (usually 30 min). Finally, we recorded the statistic that was used to report repeatability (ICC or *r*) and whether any covariates were included when calculating repeatability (i.e., agreement vs. adjusted repeatability; Nakagawa and Schielzeth, 2010). Studies sometimes corrected for covariates such as year of sampling, capture number, mass, etc. Additionally, some studies calculated repeatability based on the rank order of glucocorticoid concentrations rather than absolute concentrations (Spearman correlations). Where possible, we included both agreement repeatabilities and adjusted repeatabilities in our dataset, but some studies reported only adjusted repeatabilities and we retained these effect sizes in our analyses. When both an agreement and adjusted repeatability estimate from the same group of animals was added to the dataset, only the agreement estimate was included in our main analyses because agreement estimates were most commonly reported and the type of covariates included was inconsistent across studies.

Many studies reported more than one effect size from the same group of animals when multiple samples were taken; for example, baseline and stress-induced repeatability, multiple stress-induced repeatabilities from a series of samples taken at different time points, or repeatabilities across different time intervals using different subsets of individuals. Some studies also reported the repeatability of estimates from several independent groups; for example, males and females, or separate species that were included in the same publication. Because estimates based on the same animals—or subsets of the same group of animals—are not independent, we included a group identity factor in our dataset that served to group all estimates that were calculated using some subset of the same animals. Individual papers could include more than one group identity when repeatability estimates based on separate groups were reported (e.g., males and females).

A few studies reported effect sizes from nested subsets of individuals taken over similar time scales (e.g., individuals sampled at any time in the breeding season versus only those sampled at the exact same breeding stage; Lanctot et al., 2003; Vitousek et al., In Review); because these cases were rare we included only the effect size calculated with the larger sample size in our main analysis, but address differences between these groups in the discussion. Some studies also reported repeatability estimates for 'area-under-the-curve' (AUC) or 'corrected area-under-the-curve' (AUCc). We added these estimates to our dataset, but did not include them in our main analyses because 1) they combine baseline and stress-induced measures, 2) only a few studies ($n = 6$) reported this statistic, and 3) each of these studies also reported repeatability for baseline and stress-induced values separately. Furthermore, repeatability estimates based on AUC are a problematic trait to interpret because observed repeatability (or lack thereof) can be driven by variation in repeatability at each point in the time series (e.g., high repeatability of baseline samples could drive high AUC repeatability despite low repeatability for stress-series samples or vice versa). Finally, two studies measured glucocorticoid concentrations from identical samples using multiple assay types and report repeatability from each (Frynta et al., 2009; Montiglio et al., 2012); in these cases, we randomly selected a single estimate to include in our analyses.

2.4. Data analysis

Studies included in our dataset varied in sample size, number of samples per individual, and in how repeatability was estimated. Thus, it was important to weight studies appropriately and to convert reported repeatabilities to a comparable statistic. We converted all repeatability estimates to the standardized effect size Fisher's Z along with the corresponding sampling variance for each study (as described in Holtmann et al. (2017) and McGraw and Wong (1996)). Repeatability estimates reported as ICC or *r* statistics to were converted to Z-scores using the following formulas:

$$Z_{ICC} = 0.5 * \ln\frac{1 + (k-1) * ICC}{1 - ICC} \text{ or } Z_r = 0.5 * \ln\frac{(1+r)}{(1-r)}$$

where k is the average number of samples per individual, ICC is the intra-class correlation coefficient, and r is the Pearson or Spearman correlation coefficient. Sampling variance estimates for each effect size were calculated using the formulas:

$$\text{Var}Z_{ICC} = \frac{k}{2 * (n-2)(k-1)} \text{ or } \text{Var}Z_r = \frac{1}{n-3}$$

where n is the total number of individuals included in the estimate. All meta-analytical models were fit using Z-scores and sampling variances, but when plotting and when reporting parameter estimates (both in text and in figures) we back-transformed effect sizes to ICC to make interpretation easier using the formula:

$$ICC = \frac{\exp(2 * Z) - 1}{\exp(2 * Z) + \bar{k} - 1}$$

After back-transformation to ICC, effect sizes ranged from 1 (perfect repeatability) to −1 (values at time n were negatively correlated with samples at time $n + 1$); studies with a value of 0 on this scale had no detectable repeatability. Using these conversion formulas allowed us to include repeatability estimates that were reported in different ways in the same analysis and allowed us to weight studies appropriately for differences in sample size and number of samples taken per individual (Holtmann et al., 2017).

For our main analyses, we fit meta-regression models using the *metafor* package in R with Z-scores for repeatability as the response variable and accounting for study specific sampling variances (Viechtbauer, 2010). In each of these models, group identity was included as a random effect nested within species identity. We initially fit an intercept only model including all effect sizes and no covariates to estimate overall repeatability across all of the studies included in our dataset. We fit similar intercept only models separately for baseline, stress-induced, and integrated samples to illustrate the differences between these sample types. We calculated the heterogeneity statistics τ^2 and I^2 for the intercept only model using the rma function in *metafor*. τ^2 indicates the estimated

between study variance and I^2 is the percent (0–100%) of variation in effect sizes attributable to the between study variance. If I^2 is large (>75%) variation in repeatability might be due to other moderators (Higgin et al., 2003). To test for the effects of moderators on repeatability, we adopted a model selection approach with a set of candidate models compared by AICc scores (as in Foo et al. (2017)). Our set of candidate models included an intercept only (null) model, along with 11 models that contained combinations of one or two moderators each.

Potential moderators included sample type (baseline, stress-induced, or long-term), taxon, season (breeding versus non-breeding), sample size, and interval between sampling. Both sample size and the interval between sample collection were log transformed prior to inclusion. Effect sizes were not evenly distributed across levels of the covariates in our dataset (Table 1), so we could not test the effect of each covariate that was included in our dataset. For example, sample type and sample substrate were almost completely confounded, making it impossible to distinguish the influence of sample type (baseline, stress induced, or long-term) from that of substrate (blood, feces, or feather) on repeatability; thus, we did not explicitly test for differences in repeatability across sample substrates, but note that some of the differences observed between sample types may be attributable to sample substrate. Amphibians were represented in the dataset by only a single species, yet this species has been subject to three independent studies which each reported several effect sizes that were among the highest repeatabilities of any in our dataset. This is also the only species in which glucocorticoid repeatability has been estimated in urine. Given the uncertainty over how this species represents amphibians in general, and the unique sample substrate, we excluded amphibians from the models, but note that the analyses were qualitatively similar when this species was included, except that all repeatability estimates were slightly higher.

In the published effect sizes that we compiled, birds were over-represented, while data on reptiles, amphibians, fish and—to a lesser extent—mammals, were sparse. Given the interest in repeatability in studies of birds, the larger sample size for this taxon, and the application of a more standardized sampling technique (blood-based measures following handling stress), we also conducted a separate analysis similar to that described above, but including only measurements taken from blood in birds. In this analysis, we used a model selection approach to compare the fit of 10 candidate models that included an intercept only (null) model and models that included a combination of the moderators: sample type, the log of sample size, the log of the interval between sample collection, season, and captive versus wild. For both the global dataset and the bird only dataset, we initially fit our candidate models using maximum likelihood to allow comparison by AICc. After comparison with this approach, we refit any models with ΔAICc scores $\leq$4 using restricted maximum likelihood. We report and interpret the parameter estimates from these best-supported models only.

We tested for evidence of publication bias in several ways following approaches used in recent meta-analyses (Foo et al., 2017; Holtmann et al., 2017). Using the overall intercept only model, we ran Egger's regression test, which reports the relationship between standardized residuals and study precision. From this same model we performed a trim-and-fill analysis to test for funnel plot asymmetry using the *trimfill* function in *metafor*. Funnel plot asymmetry can be indicative of heterogeneity among studies' effect sizes or publication bias favoring significant results (Sterne et al., 2011). Next, we plotted the relationship between publication year and effect size, because studies with larger effects may be more likely to be published earlier while those with smaller effects are delayed or never published. Finally, we made funnel plots based on the residuals of the best-supported model for both the global dataset and for the subsequent analysis that was restricted to samples of blood in birds. All analyses were run in R version 3.3.3 (R Core Development Team, 2016). The full dataset that we constructed, including some effect sizes that were excluded from our main analyses, is available in the Supplementary Material (Table S1).

3. Results

Our literature search resulted in a final dataset for analysis that included 156 effect sizes from 67 studies that represented 83 unique groups of animals and 53 species (Table 1; Table S1). Most effect sizes that we included were reported as intra-class correlation coefficients rather than Pearson/Spearman r (ICC = 145, r = 11). Unsurprisingly, the studies that we identified were not evenly distributed across the factors that we recorded. For example, only birds and mammals were well represented, and for some measure types—such as baseline samples—only birds were well represented (Table 1). Still, the dataset represents a substantial body of work to describe the overall repeatability of glucocorticoid levels. In an initial analysis, we used a likelihood ratio test to determine the significance of the random effects included in our models with repeatability Z-score as the response variable (following Foo et al., 2017). When compared to an intercept only model with no random effects, both group identity and species identity significantly improved model fit (likelihood ratio tests for group ID: χ^2_1 = 4,405,603, P < .0001, species: χ^2_1 = 4,405,603, P < .0001). Model fit was further improved when including both random effects in the same model (likelihood ratio test: χ^2_1 = 30.7, P < .0001). All subsequent models included group identity nested within species as random effects.

3.1. Overall glucocorticoid repeatability

Pooling all effect sizes in a single meta-regression model that included group identity and species as random effects, the overall back-transformed ICC estimate was 0.29 [95% confidence interval: 0.25–0.34] (Fig. 1; n = 156 effect sizes from 83 groups). We performed similar intercept only models for subsets of the data that represented long-term, baseline, and stress-induced measures of glucocorticoids (Fig. 2A–C; estimate and confidence interval of back-transformed ICC for long-term measures: n = 31 estimates

Table 1
Summary of effect sizes and studies included in the reduced dataset used for the main analyses.

Taxon	Effect sizes	Species	Sample type			Substrate				Location	
			Base	Stress-induced	Long-term	Blood	Fecal	Water	Feather	Captive	Wild
Birds	119	32	60	40	19	100	15	0	4	23	96
Mammals	22	13	2	6	14	8	14	0	0	4	18
Fish	10	6	3	7	0	6	0	4	0	8	2
Reptiles	5	2	2	3	0	5	0	0	0	1	4
Total	156	53	67	56	33	119	29	4	4	36	120

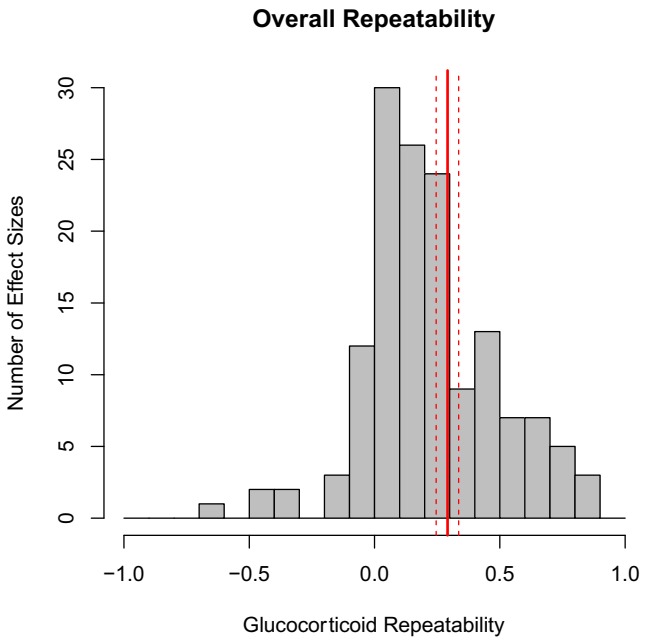

Fig. 1. Overall repeatability of within-individual glucocorticoid measurements included in this study (n = 156 effect sizes from 83 groups). In cases where the same type of effect was calculated multiple times for the same group of animals, only one is included in the histogram. The solid red line indicates the overall repeatability estimate of 0.29 with upper and lower bounds of the 95% confidence interval indicated by dashed red lines (0.25–0.34). These estimates are from an intercept only meta-regression model including all effect sizes and with group identity and species as random effects. Gray bars show a histogram of the raw data without accounting for factors included in the meta-regression model (i.e., sample sizes and random effects).

from 26 groups; 0.44 [0.27–0.57]; baseline: n = 68 estimates from 50 groups; 0.18 [0.11–0.24]; and stress-induced: n = 51 estimates from 39 groups; 0.38 [0.29–0.45]).

The overall intercept only model indicated that there was substantial heterogeneity in effect sizes across studies that was not due to sampling error (τ^2 = 0.177; I^2 = 100%), suggesting that moderators might explain some of this variation. The high heterogeneity for I^2 was driven by the inclusion of a few studies with very large sample sizes and correspondingly precise repeatability estimates; when re-running the same intercept only model excluding studies with sample sizes >200 individuals I^2 was only 85.7%, which is typical for meta-analyses in the fields of ecology and evolution where the mean I^2 is often quite high (Foo et al., 2017; Holtmann et al., 2017; Senior et al., 2016). We next performed a model selection analysis on a set of candidate models that included a variety of moderators that could influence repeatability. In this analysis, only one model had a ΔAICc score ≤4 and this model received 90% of the model weight among our candidate models (Table 2); this model included the effects of sample type and the average interval between repeated sample collection. Measurements from baseline samples had lower repeatability than those from stress-induced or integrated samples. A longer interval between sample collection was also associated with lower repeatability (Fig. 3A and B; Table 3).

3.2. Repeatability of corticosterone measurements from avian blood

When restricting our analysis to measures of corticosterone in bird blood, we found less heterogeneity in effect sizes across studies in the initial intercept only model than in the global model fit earlier (τ^2 = 0.025; I^2 = 61.6%). In our model selection analysis, the best-supported model included only sample type (baseline vs. stress-induced) as a predictor (Table 4). Models that included a second covariate in addition to sample type also received some support (sampling interval, ΔAICc = 0.66, w_i = 0.29; sample size, ΔAICc = 1.94, w_i = 0.15). As in the global analysis, stress-induced samples had higher repeatability than baseline samples (Table 3).

3.3. Tests for publication bias

In our global intercept only model, the trim-and-fill test indicated that no effect sizes were missing and an Egger's test indicated no asymmetry in the funnel plot (n = 156, z = 1.183, P = .24). Separate funnel plots for the two models that received the best support in our model selection analyses also showed no visual evidence of asymmetry (Fig. 4). Finally, there was no evidence for a relationship between year of publication and effect size (Fig. 5; n = 143; estimate = −0.003; P = .27).

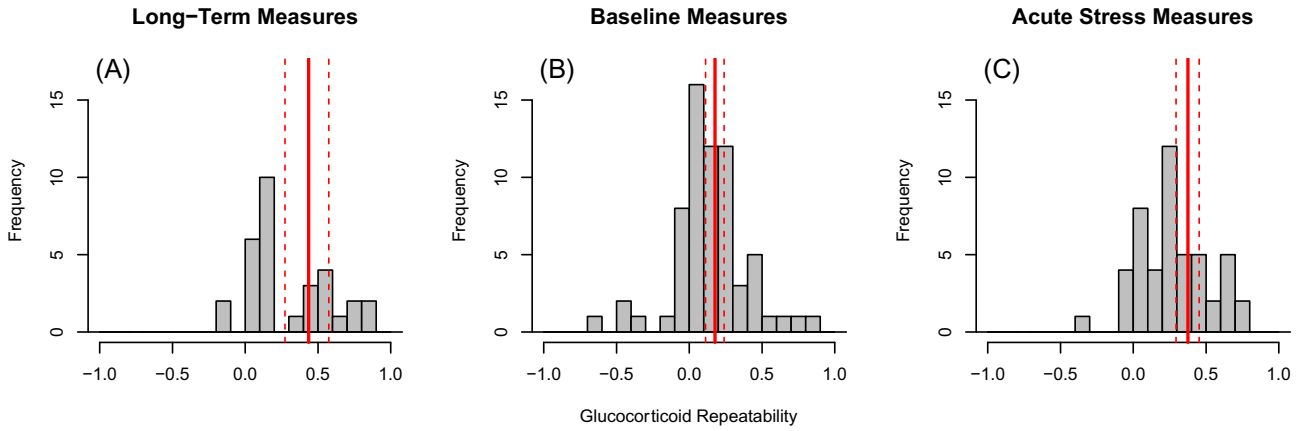

Fig. 2. Histograms showing overall repeatability of glucocorticoid measurements that capture (A) long-term (n = 31 estimates from 26 groups, repeatability = 0.44, CI = 0.27–0.57), (B) baseline (n = 61 estimates from 47 groups, repeatability = 0.18, CI = 0.11–0.24), or (C) acute stress levels (n = 48 estimates from 36 groups, repeatability = 0.38, CI = 0.29–0.45). Solid red lines indicate overall repeatability and dashed red lines indicate the upper and lower bounds of a 95% confidence interval based on three meta-regression models fit separately for each panel with group identity and species included as random effects. Gray bars show a histogram of the raw data without accounting for factors included in the meta-regression model (i.e., sample sizes and random effects). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Candidate models and AICc rankings for meta-regression model set that included all effect sizes with repeatability Z-scores as the response variable. Pseudo r^2 values are calculated with respect to the intercept only model.

Model	Log Likelihood	Pseudo r^2	*k*	*n*	Δ AIC_c	Weight
~Type + log(Interval)	−33.74	0.29	6	156	0	0.90
~Type + log(*n*)	−36.83	0.22	6	156	6.2	0.04
~Type + Captivity	−37.20	0.22	6	156	6.9	0.03
~Type + Taxon	−35.44	0.25	6	156	7.8	0.02
~Type	−39.12	0.18	6	156	8.6	0.01
~Type + Season	−38.30	0.19	7	156	11.3	0.00
~Log(Interval)	−42.06	0.11	4	156	12.3	0.00
~Taxon	−43.27	0.09	6	156	19.1	0.00
~Log(*n*)	−45.43	0.04	4	156	19.1	0.00
~Captivity	−45.60	0.04	4	156	19.4	0.00
~Intercept Only	−47.42	0.00	3	156	21.0	0.00
~Season	−46.65	0.02	5	156	23.7	0.00

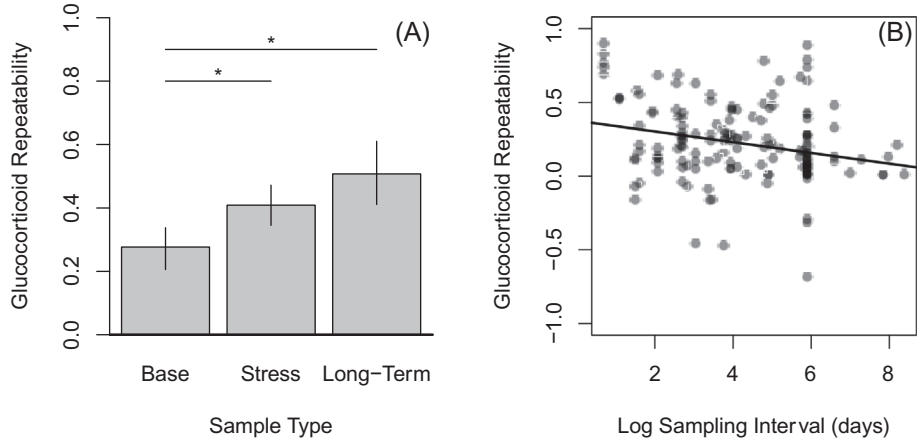

Fig. 3. Results from the best-fit model combining all repeatability estimates with potential covariates (*n* = 156 estimates from 83 groups). (A) Model predicted estimates for repeatability of baseline, stress-induced, and long-term samples with the sampling interval held at a constant (mean) value. (B) Relationship between log sampling interval and glucocorticoid repeatability. Details of the fit model are presented in Table 2.

Table 3
Parameter estimates for the best-supported models from the global analysis (Table 2) and from the analysis restricted only to samples of avian blood (Table 4). In both cases, the intercept represents the estimate for stress-induced samples. Note that parameter estimates in this table are shown as Z-scores but figures illustrate values back-transformed to ICC.

	Parameter	Estimate	SE	CI	Z	P
Global dataset (n = 156 effect sizes, 83 groups, 53 species)						
	Intercept (stress induced)	0.623	0.074	0.48 to 0.77	8.44	<0.0001
	Baseline samples	−0.151	0.037	−0.22 to −0.08	−4.07	<0.0001
	Long-term samples	−0.056	0.072	−0.20 to 0.09	−0.77	0.44
	Log(Interval)	−0.046	0.014	−0.07 to −0.02	−3.40	0.0007
Bird blood only (n = 100 effect sizes, 45 groups, 27 species)						
	Intercept (stress induced)	0.338	0.036	0.268 to 0.408	9.44	<0.0001
	Baseline samples	−0.152	0.038	−0.23 to −0.08	−4.02	<0.0001

Table 4
Candidate models and AICc rankings for meta-regression model set that included only avian blood samples with repeatability Z-scores as the response variable. Pseudo r^2 values are calculated with respect to the intercept only model.

Model	Log Likelihood	Pseudo r^2	*k*	*n*	Δ AIC_c	Weight
~Type	−14.27	0.36	4	100	0	0.40
~Type + log(Interval)	−13.49	0.39	5	100	0.7	0.29
~Type + log(*n*)	−14.13	0.36	5	100	1.9	0.15
~Type + Season	−13.55	0.39	6	100	3.0	0.09
~Type + Captivity	−15.05	0.32	5	100	3.8	0.06
~Intercept Only	−22.13	0.00	3	100	13.6	0.00
~Log(Interval)	−21.37	0.03	4	100	14.2	0.00
~Captivity	−21.67	0.02	4	100	14.8	0.00
~Log(*n*)	−21.79	0.02	4	100	15.0	0.00
~Season	−21.82	0.01	5	100	17.3	0.00

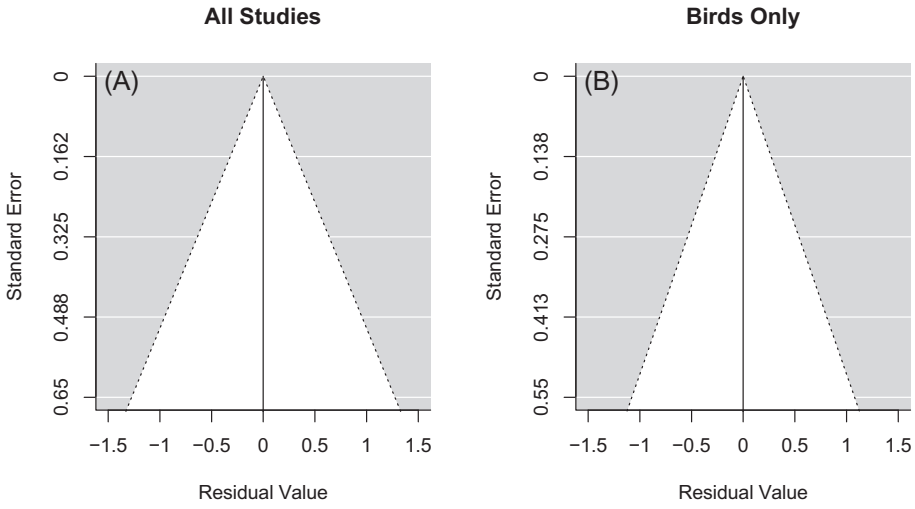

Fig. 4. Funnel plots of the meta-analytic residuals for the best supported overall model (A) and the best-supported model for bird blood only (B). Egger's test found no evidence for asymmetry (n = 156, z = 1.18, P = .24), suggesting that the funnel plots provide no evidence for publication bias.

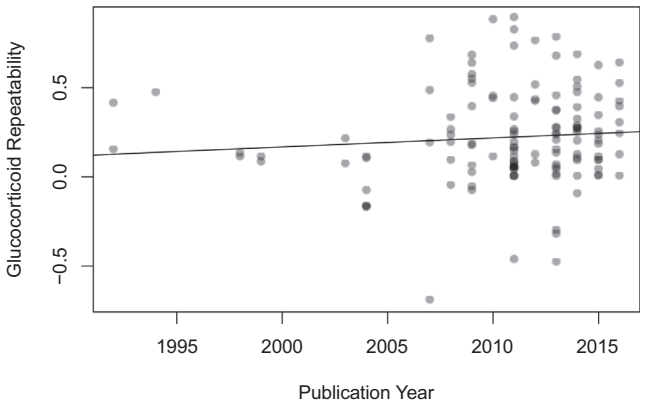

Fig. 5. Glucocorticoid repeatability was not significantly related to publication year (n = 142 estimates from 81 groups; publication year = −0.003, z = −0.5, P = .62).

4. Discussion

At the broadest level, our data support the overall conclusion that glucocorticoid concentrations are moderately repeatable. Despite clear evidence that concentrations across all groups were significantly repeatable, there was substantial heterogeneity in the effect sizes included in our analysis. Some of this variation can be explained by covariates: measures made on baseline samples and at longer intervals were less repeatable than those made on stress-induced or long-term samples and at shorter intervals. Even after accounting for these covariates, however, studies varied widely in estimated repeatability. Thus, it is clear that researchers interested in understanding stable between individual differences in a particular system or under particular conditions will need to design studies to measure repeatability directly in their own system rather than relying on general patterns drawn from previous work. Despite this heterogeneity, our results do indicate that stable between-individual differences in glucocorticoid concentrations often exist. When these differences are heritable, selection on glucocorticoid concentrations could drive the evolution of glucocorticoid phenotypes.

Many recent studies have focused on hormone repeatability as an indication of evolvability, yet repeatability is an imperfect measure of stable between-individual differences in labile traits like glucocorticoid concentrations (Bonier and Martin, 2016). Repeatability of glucocorticoids is sometimes presented as an indication of the upper bound limit of heritability. However, Dohm (2002) critiques this interpretation of repeatability, because it is dependent on several assumptions that may often be violated, especially when investigating labile phenotypic traits. For example, if individuals differ in the glucocorticoid response to uncontrolled changes in environment (e.g., temperature or food availability), repeatability estimates may be low despite intrinsic differences in glucocorticoid phenotype (see further discussion below). Similarly, maternal effects can lead to an underestimate of repeatability when there are negative correlations between maternal and genetic impacts on the phenotype of interest (Dohm, 2002). Given the importance of early environmental effects in shaping the hypothalamic-pituitary-adrenal axis (Weaver et al., 2004), evaluating this assumption may also be important for understanding glucocorticoid repeatability. While repeatability can still provide useful information about glucocorticoids, researchers should be cautious about interpreting a lack of repeatability as evidence for the absence of heritable differences in glucocorticoid secretion patterns. At the same time, researchers should be cautious about interpreting high observed repeatability as strong evidence for heritable differences, because persistent environmental effects could produce high repeatability estimates even in the absence of genetic differences (Niemelä and Dingemanse, 2017). Ideally, more studies would combine repeatability estimates with direct measurements of heritability, but both heritability and repeatability estimates are available from very few populations, especially in the wild (but see Jenkins et al., 2014; Stedman et al., 2017; Vitousek et al., 2014; Vitousek et al., In Review).

Although the relative repeatability of different trait types could provide insight into their function and evolution, we suggest caution in comparing published repeatability estimates across different disciplines for several reasons. First, fields differ in the extent to which studies are designed with the explicit goal of estimating repeatability. For example, understanding behavioral consistency has been a major research goal for more than a decade (Sih et al., 2004), and many studies have been designed for the purpose of examining between-individual stability in behavioral types (e.g., Araya-Ajoy and Dingemanse, 2016). Extensive attention has been given to experimental design and analysis approaches (Dingemanse and Dochtermann, 2012; Dingemanse et al., 2009), and behavioral studies often employ sampling regimes explicitly designed to investigate repeatability within versus across contexts and environments. In contrast, few endocrine studies have been

designed to address repeatability; instead, repeatability estimates typically rely on samples collected for a different primary purpose (e.g., Vitousek and Romero, 2013; Vitousek et al., 2014).

Although hormones are known to vary across contexts and environments, very few studies are designed in such a way that the role of external and internal factors affecting repeatability can be examined. The amount of standardization inherent in common sampling protocols also differs across trait types and fields. For example, metabolic rate is generally measured in highly standardized conditions, following a period of acclimation to the sampling environment (Auer et al., 2016; Holtmann et al., 2017). In contrast, baseline glucocorticoid measurements in wild animals typically include very limited control over the stimuli experienced by an animal prior to capture that may directly influence glucocorticoid concentrations. Interestingly, for stress-induced samples the stimulus control afforded is much greater (i.e., all animals have experienced a similar period held in similar conditions immediately prior to sampling) and our results demonstrate that repeatability is significantly higher. At this point, it is difficult to say whether the difference between baseline and stress induced repeatability is due to biological differences in the degree of between individual stability in these traits or due to methodological differences in the ability to control the pre-sampling stimulus period (or a combination of both). Two recent studies in great tits (*Parus major*) demonstrate this possibility nicely. In the first, individuals were repeatedly captured and sampled from the wild with no control over pre-capture conditions and baseline corticosterone was not repeatable (Baugh et al., 2014). In a subsequent study, wild caught individuals were housed in captivity and sampled repeatedly, which allowed for greater control over pre-sampling stimulus; in this case, significant baseline repeatability was detected (Baugh et al., 2017).

In addition to finding higher repeatability for stress-induced samples, we also found that long-term measures—those that incorporate glucocorticoid secretion over an extended period of time—had relatively high repeatability. This higher repeatability may be driven in part by the fact that these samples reflect both individual differences in stress-induced glucocorticoid levels (due to unknown stressors that occurred prior to sample collection) and a "smoothed" average of baseline glucocorticoid levels over an extended period of time. It is also worth noting, however, that these long-term samples are measured in a different sample substrate (feces or feathers) than baseline and stress-induced glucocorticoid levels (plasma, water, or urine). It is certainly possible that repeatability estimates may differ across sample substrates in addition to varying based on the time period that they represent, but disentangling these sources of variation will be difficult. Repeatability might also vary within the same sample substrate; for example, if the smoothing hypothesis described above is correct, we may expect that fecal glucocorticoids will be more repeatable in species with longer gut passage times, but at present the data are not available to address these possibilities.

In a recent meta-analysis on the repeatability of behavior, metabolism, and hormone measures in birds, Holtmann et al. (2017) estimated an overall repeatability of only 0.15 for all hormone measures. Our results do agree with theirs in detecting higher repeatability for stress-induced measures, but our overall estimate for repeatability of 0.29 is somewhat higher. It is difficult to determine exactly what drives these differences, because the analyses are not directly comparable. For example, Holtmann et al. (2017) include glucocorticoids, androgens, and estrogens in the same models, do not account for substrate type, and focus only on birds. Although they do include moderators in their analyses, each moderator is only included as a univariate predictor. Based on their low repeatability estimate for hormone concentrations, Holtmann et al. (2017) conclude that between-individual variation in hormones is unlikely to act as a mechanism underpinning stable between-individual differences in behavior, and suggest that metabolic rate is a more likely candidate. Although we describe complications associated with comparing repeatability estimates across trait types above, our higher repeatability estimates suggest that it is premature to disregard the role of glucocorticoids in mediating stable behavioral differences, especially for some types of behaviors. For example, the relatively high repeatability of stress-induced glucocorticoid levels (0.38) suggests that individual differences in the behavioral response to challenges may be more likely to be mechanistically linked to glucocorticoids than many of the behaviors associated with daily activity. In fact, our overall estimate of repeatability was only slightly lower than a previously published estimate for the overall repeatability of behavioral traits (0.29 for glucocorticoids, this study; 0.39 for behavior, Bell et al., 2009).

We did not see an effect of life history stage on hormone repeatability. However, the broad generalization used here (breeding versus non-breeding) likely misses much of the subtlety in how glucocorticoid regulation varies across fine scale life history stages. For example, the 'breeding' stage in most bird species included in our analysis encompasses territory establishment, pair formation, nest building and laying, incubation, provisioning, and a pre-molt period on the breeding grounds. Each of these stages has specific demands, and modulation of baseline and stress induced corticosterone is known to play a role in matching these demands (Jacobs and Wingfield, 2000). Indeed, two recent studies that calculated repeatability both for a large pool of samples and from a subset of samples collected only at the exact same life history stage (e.g., during nestling provisioning) found that repeatability differs not only based on interval between sampling, but also based on whether comparisons are made across individuals sampled in the same stages of the breeding season (e.g., incubation vs. provisioning, Lanctot et al., 2003; Vitousek et al., In Review). Future work should aim to compare individuals repeatedly sampled at the same life history stages and to determine whether there are consistent differences in repeatability across different life history stages (Taff and Vitousek, 2016). For example, it may be that repeatability is highest during stages that are particularly energetically demanding, such as nestling provisioning, because these stages force animals to perform close to their individual physiological limits.

Glucocorticoid concentrations are widely recognized to vary across contexts and environments. Studies of hormone repeatability sometimes report adjusted repeatabilities to account for internal and external factors that can affect circulating hormone concentrations (e.g., season, weather, life history stage). However, this approach is sufficient to control for context only if individuals respond similarly to changing conditions. This assumption has rarely been tested, but some recent evidence indicates that there can be individual by environment interactions in glucocorticoid secretion (e.g., Lendvai et al., 2014). Rather than simply controlling for covariates, understanding hormone repeatability across contexts will require adopting the within-individual reaction norm approach that has been used successfully in studies of animal personality (Dingemanse and Dochtermann, 2012; Dingemanse et al., 2009). Several recent reviews have advocated the use of a reaction norm approach for studying labile endocrine phenotypes (Hau et al., 2016; Hau and Goymann, 2015; Lema, 2014; Taff and Vitousek, 2016; Wada and Sewall, 2014). In fact, given the importance of glucocorticoids in mediating responses to changing conditions (Wingfield, 2003), we might predict that differences in glucocorticoid reaction norms (the pattern of within-individual changes across contexts) would be more repeatable than absolute concentrations of glucocorticoids. Very few studies to date have explicitly assessed the repeatability of reaction norms (but see Fürtbauer et al., 2015; Lendvai et al., 2014), but this is a promising

approach for future work. At present, it will be logistically challenging to collect the samples needed to adequately assess reaction norms for endocrine traits, but detailed recommendations are available from the behavioral repeatability literature that describe optimal experimental design and sampling regimes along with statistical methods to characterize the repeatability of reaction norms (Araya-Ajoy et al., 2015; van de Pol, 2012). The continued development of techniques that allow for repeated, non-invasive sampling of endocrine levels in the same individuals across conditions will be essential to understanding variation in these traits (e.g., Bauch et al., 2013).

Despite the logistical challenges of collecting samples, and shortcomings of using repeatability as a measure of stable between-individual differences in labile traits, we suggest that studies should report repeatability when multiple samples are collected from each individual. These estimates should also include information on how repeatability was calculated and what environmental conditions differed between sampling periods. Best practices currently recommend the use of LMM approaches to report repeatability, because they avoid several potential weaknesses associated with traditional ANOVA based approaches; correlation based approaches should be avoided (Nakagawa and Schielzeth, 2010). In our dataset, most estimates were based on ANOVAs with only a small number of correlation or LMM effects included. Given the lack of variation in the dataset, we could not evaluate the effect of statistical choice empirically, but there are strong theoretical arguments for preferring the LMM approach (Nakagawa and Schielzeth, 2010). At present, there are pronounced cultural differences between sub-fields in how and when repeatability is reported that make taxonomic comparisons difficult. For example, Table 1 in Cavigelli and Caruso (2015) summarizes at least 31 studies of primates that have collected and measured glucocorticoids from an average of ~30 fecal samples per individual, yet none of these studies explicitly report repeatability (note, however, that some have called for increased focus on within-individual variation in glucorticoids in primates as well; Beehner and Bergman, 2017). In contrast, recent studies of birds often report repeatability as a way to describe stable between-individual differences in glucocorticoid phenotypes (e.g., Cockrem et al., 2016; Ouyang et al., 2011; Vitousek et al., In Review). These differences limit the generality of comparative questions about differences in repeatability across taxa, sample types, and measurement periods. Regardless of these differences, our analysis clearly demonstrates that overall glucocorticoid secretion is repeatable across a wide range of species, but also that repeatability varies enormously between particular studies and conditions.

Data accessibility

The dataset used in these analyses and a list of references from which the dataset was collected are available in the electronic Supplementary material.

Author contributions

C.T., L.S., and M.V. conceived the ideas and designed the methodology. C.T. collected the data, analyzed the data, and led the writing of the manuscript with input from L.S. and M.V. All authors gave final approval for publication.

Acknowledgments

We thank the Vitousek Lab for discussion and feedback on this manuscript. CCT was funded by a postdoctoral fellowship from the Cornell Lab of Ornithology. LAS was supported by an EPA STAR Fellowship (FP-917686). Support was also provided by NSF IOS-1457251 to MNV. The EPA has not formally reviewed this publication, the EPA does not endorse products or services described here, and any views expressed here belong to the authors alone.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ygcen.2018.01.011.

References

Angelier, F., Wingfield, J.C., Weimerskirch, H., Chastel, O., 2010. Hormonal correlates of individual quality in a long-lived bird: a test of the 'corticosterone-fitness hypothesis'. Biol. Lett. 6, 846–849.
Araya-Ajoy, Y.G., Dingemanse, N.J., 2016. Repeatability, heritability, and age-dependence in the aggressiveness reaction norms of a wild passerine bird. J. Anim. Ecol. 86, 227–238.
Araya-Ajoy, Y.G., Mathot, K.J., Dingemanse, N.J., 2015. An approach to estimate short-term, long-term, and reaction norm repeatability. Methods Ecol. Evol. 6, 1462–1473.
Auer, S.K., Bassar, R.D., Salin, K., Metcalfe, N.B., 2016. Repeatability of metabolic rate is lower for animals living under field versus laboratory conditions. J. Exp. Biol. 219, 631–634.
Bauch, C., Becker, P.H., Verhulst, S., 2013. Telomere length reflects phenotypic quality and costs of reproduction in a long-lived seabird. Proc. R. Soc. B 280, 20122540.
Baugh, A.T., Oers, K.V., Dingemanse, N.J., Hau, M., 2014. Baseline and stress-induced glucocorticoid concentrations are not repeatable but covary within individual great tits (Parus major). Gen. Comp. Endocrinol. 208, 154–163.
Baugh, A.T., Schaper, S.V., Hau, M., Cockrem, J.F., de Goede, P., van Oers, K., 2012. Corticosterone responses differ between lines of great tits (*Parus major*) selected for divergent personalities. Gen. Comp. Endocrinol. 175, 488–494.
Baugh, A.T., Senft, R.A., Firke, M., Lauder, A., Schroeder, J., Meddle, S.L., van Oers, K., Hau, M., 2017. Risk-averse personalities have a systemically potentiated neuroendocrine stress axis: a multilevel experiment in Parus major. Horm. Behav. 93, 99–108.
Beehner, J.C., Bergman, T.J., 2017. The next step for stress research in primates: to identify relationships between glucocorticoid secretion and fitness. Horm. Behav. 91, 68–83.
Bell, A.M., Hankison, S.J., Laskowski, K.L., 2009. The repeatability of behaviour: a meta-analysis. Anim. Behav. 77, 771–783.
Boake, C.R.B., 1989. Repeatability: its role in evolutionary studies of mating behavior. Evol. Ecol. 3, 173–182.
Bonier, F., Martin, P., Moore, I., Wingfield, J., 2009. Do baseline glucocorticoids predict fitness? TREE 24, 634–642.
Bonier, F., Martin, P.R., 2016. How can we estimate natural selection on endocrine traits? Lessons from evolutionary biology. Proc. R. Soc. B 283, 20161887.
Breuner, C.W., Patterson, S.H., Hahn, T.P., 2008. In search of relationships between the acute adrenocortical response and fitness. Gen. Comp. Endocrinol. 157, 288–295.
Bridge, E.S., Schoech, S.J., Bowman, R., Wingfield, J.C., 2009. Temporal predictability in food availability: effects upon the reproductive axis in Scrub-Jays. J. Exp. Zool. 311, 35–44.
Cavigelli, S.A., Caruso, M.J., 2015. Sex, social status and physiological stress in primates: the importance of social and glucocorticoid dynamics. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 370, 20140103.
Cockrem, J.F., 2013. Individual variation in glucocorticoid stress responses in animals. Gen. Comp. Endocrinol. 181, 45–58.
Cockrem, J.F., Barrett, D.P., Candy, E.J., Potter, M.A., 2009. Corticosterone responses in birds: individual variation and repeatability in Adelie penguins (Pygoscelisadeliae) and other species, and the use of power analysis to determine sample sizes. Gen. Comp. Endocrinol. 163, 158–168.
Cockrem, J.F., Candy, E.J., Barrett, D.P., Agnew, P., Potter, M.A., 2016. Individual variation and repeatability of corticosterone responses of little penguins (Eudyptula minor) sampled in two successive years at Oamaru, New Zealand. Gen. Comp. Endocrinol. 244, 86–92.
Dantzer, B., McAdam, A.G., Palme, R., Fletcher, Q.E., Boutin, S., Humphries, M.M., Boonstra, R., 2010. Fecal cortisol metabolite levels in free-ranging North American red squirrels: assay validation and the effects of reproductive condition. Gen. Comp. Endocrinol. 167, 279–286.
Dingemanse, N., Dochtermann, N., 2012. Quantifying individual variation in behaviour: mixed-effect modelling approaches. J. Anim. Ecol. 82, 39–54.
Dingemanse, N., Kazem, A., Reale, D., Wright, J., 2009. Behavioural reaction norms: animal personality meets individual plasticity. TREE 25, 81–89.
Dohm, M., 2002. Repeatability estimates do not always set an upper limit to heritability. Funct. Ecol. 16, 273–280.
Evans, M.R., Roberts, M.L., Buchanan, K.L., Goldsmith, A.R., 2006. Heritability of corticosterone response and changes in life history traits during selection in the zebra finch. J. Evol. Biol. 19, 343–352.

Fletcher, Q.E., Dantzer, B., Boonstra, R., 2015. The impact of reproduction on the stress axis of free-living male northern red backed voles (Myodes rutilus). Gen. Comp. Endocrinol. 224, 136–147.
Foo, Y.Z., Nakagawa, S., Rhodes, G., Simmons, L.W., 2017. The effects of sex hormones on immune function: a meta-analysis. Biol. Rev. 92, 551–571.
Frynta, D., Nováková, M., Kutalová, H., Palme, R., Sedláček, F., 2009. Apparatus for collection of fecal samples from undisturbed spiny mice (Acomys cahirinus) living in a complex social group. J. Am. Assoc. Lab. Anim. Sci. 48, 196–201.
Fürtbauer, I., Pond, A., Heistermann, M., King, A.J., 2015. Personality, plasticity and predation: linking endocrine and behavioural reaction norms in stickleback fish. Funct. Ecol. 29, 931–940.
Hau, M., Casagrande, S., Ouyang, J.Q., Baugh, A., 2016. Glucocorticoid-mediated phenotypes in vertebrates: multilevel variation and evolution. Adv. Stud. Behav., 41–115
Hau, M., Goymann, W., 2015. Endocrine mechanisms, behavioral phenotypes and plasticity: known relationships and open questions. Front. Zool. 12, 1–15.
Higgin, J., Thompson, S., Deeks, J., Altman, D., 2003. Measuring inconsistency in meta-analysis. Br. Med. J. 327, 557–560.
Holtmann, B., Lagisz, M., Nakagawa, S., 2017. Metabolic rates, and not hormone levels, are a likely mediator of between-individual differences in behaviour: a meta-analysis. Funct. Ecol. 31, 685–696.
Jacobs, J.D., Wingfield, J.C., 2000. Endocrine control of life-cycle stages: a constraint on response to the environment? Condor 102, 35–51.
Jenkins, B., Vitousek, M., Hubbard, J., Safran, R., 2014. An experimental analysis of the heritbability of variation in glucocorticoid concentrations in a wild avian population. Proc. R. Soc. B 281, 20141302.
Lanctot, R.B., Hatch, S.A., Gill, V.A., Eens, M., 2003. Are corticosterone levels a good indicator of food availability and reproductive performance in a kittiwake colony? Horm. Behav. 43, 489–502.
Landys, M.M., Ramenofsky, M., Wingfield, J.C., 2006. Actions of glucocorticoids at a seasonal baseline as compared to stress-related levels in the regulation of periodic life processes. Gen. Comp. Endocrinol. 148, 132–149.
Lema, S.C., 2014. Hormones and phenotypic plasticity in an ecological context: linking physiological mechanisms to evolutionary processes. Int. Comp. Biol. 54, 850–863.
Lendvai, Á., Ouyang, J., Schoenle, L., Fasanello, V., Haussmann, M., Bonier, F., Moore, I., 2014. Experimental food restriction reveals individual differences in corticosterone reaction norms with no oxidative costs. PLoS One 9, e110564.
Lessells, C.M., Boag, P.T., 1987. Unrepeatable repeatabilities: a common mistake. Auk 104, 116–121.
McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1, 30–46.
Montiglio, P.O., Pelletier, F., Palme, R., Garant, D., Réale, D., Boonstra, R., 2012. Noninvasive monitoring of fecal cortisol metabolites in the Eastern Chipmunk (Tamias striatus): validation and comparison of two enzyme immunoassays. Physiol. Biochem. Zool. 85, 183–193.
Nakagawa, S., Schielzeth, H., 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol. Rev. 85, 935–956.
Narayan, E.J., Cockrem, J.F., Hero, J.-M., 2013. Repeatability of baseline corticosterone and short-term corticosterone stress responses, and their correlation with testosterone and body condition in a terrestrial breeding anuran (Platymantis vitiana). Comp. Biochem. Physiol. 165, 304–312.
Niemelä, P.T., Dingemanse, N.J., 2017. Individual versus pseudo-repeatability in behaviour: lessons from translocation experiments in a wild insect. J. Anim. Ecol. 86, 1033–1043.
Ouyang, J.Q., Hau, M., Bonier, F., 2011. Within seasons and among years: when are corticosterone levels repeatable? Horm. Behav. 60, 559–564.
Pavitt, A.T., Pemberton, J.M., Kruuk, L.E.B., Walling, C.A., 2016. Testosterone and cortisol concentrations vary with reproductive status in wild female red deer. Ecol. Evol. 6, 1163–1172.
Pottinger, T.G., Carrick, T.R., 1999. Modification of plasma cort response to stress in rainbow trout by selective breeding. Gen. Comp. Endocrinol. 116, 122–132.
Pottinger, T.G., Pickering, A.D., Hurley, M.A., 1992. Consistency in the stress response of individuals of two strains of rainbow troug, Oncorhynchus mykiss. Aquaculture 103, 275–289.
R Core Development Team, 2016. R: Alanguage and environment for statistical computing. R Foundation forStatistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URLhttp://www.R-project.org.
Rogovin, K.A., Naidenko, S.V., 2011. Noninvasive assessment of stress in bank voles (Myodes glareolus, Cricetidae, Rodentia) by means of enzyme-linked immunosorbent assay (ELISA). Bio. Bull. 37, 959–964.
Senior, A.M., Grueber, C.E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E.S., Nakagawa, S., 2016. Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. Ecology 97, 3293–3299.
Sih, A., Bell, A.M., Johnson, J.C., Ziemba, R.E., 2004. Behavioral syndromes: an integrative overview. Q. Rev. Biol. 79, 241–277.
Sorenson, G.H., Dey, C.J., Madliger, C.L., Love, O.P., 2017. Effectiveness of baseline corticosterone as a monitoring tool for fitness: a meta-analysis in seabirds. Oecologia 183, 353–365.
Stedman, J.M., Hallinger, K.K., Winkler, D.W., Vitousek, M.N., 2017. Heritable variation in circulating glucocorticoids and endocrine flexibility in a free-living songbird. J. Evol. Biol. 30, 1724–1735.
Sterne, J.A., Sutton, A.J., Ioannidis, J.P., Terrin, N., Jones, D.R., Lau, J., Carpenter, J., Rucker, G., Harbord, R.M., Schmid, C.H., Tetzlaff, J., Deeks, J.J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D.G., Moher, D., Higgins, J.P., 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 343, 1–8.
Stoffel, M.A., Nakagawa, S., Schielzeth, H., 2017. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol. Evol. 8, 1639–1644.
Taff, C.C., Vitousek, M.N., 2016. Endocrine flexibility: optimizing phenotypes in a dynamic world? TREE 31, 476–488.
Tempel, D., Gutierrez, R.J., 2004. Factors related to fecal corticosterone levels in California spotted owls: implications for assessing chronic stress. Conserv. Biol., 538–547
van de Pol, M., 2012. Quantifying individual variation in reaction norms: how study design affects the accuracy, precision and power of random regression models. Methods Ecol. Evol. 3, 268–280.
Viechtbauer, W., 2010. Conducting meta-analysesin R with the metafor package. Journal of statistical software. 36, 1–48.
Vitousek, M., Romero, M., 2013. Stress responsiveness predicts individual variation in mate selectivity. Gen. Comp. Endocrinol. 187, 32–38.
Vitousek, M.N., Jenkins, B.R., Safran, R.J., 2014. Stress and success: Individual differences in the glucocorticoid stress response predict behavior and reproductive success under high predation risk. Horm. Behav. 66, 812–819.
Vitousek, M.N., Taff, C.C., Hallinger, K.K., Zimmer, C., Winkler, D.W., In Review. Glucocorticoids and Fitness: Threshold Effects and Trade-offs Influence Optimal Endocrine Flexibility.
Wada, H., Sewall, K.B., 2014. Introduction to the symposium—uniting evolutionary and physiological approaches to understanding phenotypic plasticity. Int. Comp. Biol. 54, 774–782.
Weaver, I.C., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., Meaney, M.J., 2004. Epigenetic programming by maternal behavior. Nat. Neurosci. 7, 847–854.
Wingfield, J.C., 2003. Control of behavioural strategies for capricious environments. Anim. Behav. 66, 807–816.
Wingfield, J.C., Maney, D.L., Breuner, C.W., Jacobs, J.D., Lynn, S., Ramenofsky, M., Richardson, R.D., 1998. Ecological bases of hormone-behavior interactions: the "emergency life history stage". Am. Zool. 38, 191–206.