# Do egg hormones have fitness consequences in wild birds? A systematic review and meta-analysis

Lucia Mentesana[1,2,3*], Michaela Hau[1,4], Pietro B. D'Amelio[5], Nicolas M. Adreani[2,6], Alfredo Sánchez-Tójar[3]

[1] Max Planck Institute for Biological Intelligence, Germany.
[2] Biology Department, Facultad de Ciencias, Universidad de la República, Uruguay.
[3] Evolutionary Biology Department, Faculty for Biology, Bielefeld University, Germany.
[4] Department of Biology, University of Konstanz.
[5] Department of Biology, Reed College, 3203 SE Woodstock Blvd, Portland, OR, 97202, USA.
[6] Max Planck Institute for Animal Behaviour, Germany.

Dr. Lucia Mentesana: luciamentesana@gmail.com (https://orcid.org/0000-0003-4090-2570)
Dr. Michaela Hau: michaela.hau@bi.mpg.de (https://orcid.org/0000-0002-3836-4083)
Dr. Pietro B. D'Amelio: pie.damelio@googlemail.com (https://orcid.org/0000-0002-4095-6088)
Dr. Nicolas M. Adreani: mn.adreani@gmail.com (https://orcid.org/0000-0002-5043-0389)
Dr. Alfredo Sánchez-Tójar: alfredo.tojar@gmail.com (https://orcid.org/0000-0002-2886-0649)

Keywords (max 10): avian physiology, offspring size, offspring survival, hatching success, clutch size, life-history, steroids, androstenedione, testosterone, 5-α dihydrotestosterone.

Type of article: Reviews and Syntheses.

* Author for correspondence: Dr. Lucia Mentesana
Address: Igua 4225, Montevideo (11400), Uruguay. Telephone: (+598) 2525-8617.

## Abstract

Egg-laying species are key models for understanding the adaptive significance of maternal effects, with egg hormones being identified as an important underlying mechanism. However, even thirty years after their discovery, the evolutionary consequences of hormone-mediated maternal effects remain unclear. Using an evidence synthesis approach, we tested the extent to which an increase of prenatal maternal hormone deposition into eggs relates to fitness in wild birds (19 species, 443 effect sizes, 57 studies). Egg androgens, glucocorticoids, and thyroid hormones showed an overall near-zero mean effect for both maternal and offspring fitness proxies. However, heterogeneity was high, suggesting that egg hormone effects on fitness will vary depending on context. Hormone type, offspring sex and age did not explain much of the observed variance, nor did methodological factors such as type of study or experimental design. Our heterogeneity investigation showed that differences in effect sizes were mostly driven by within-study variability and phylogenetic relationships. Our study provides the most comprehensive estimate to date of the relationship between egg hormones and fitness in vertebrates. Additionally, it identifies significant biases (e.g., taxonomic, publication) and knowledge gaps that are key to address in future research to fully understand the evolutionary relevance of maternal egg hormones.

## 1. Introduction

Maternal effects are widespread across nature, occurring in plants, invertebrates, and vertebrates. They occur when mothers shape their offspring's phenotype by influencing the environment they provide, independently of the genes they pass on (Mousseau & Fox 1998; Reinhold 2002; Uller 2008, 2012; Wolf & Wade 2016). Maternal effects can contribute to half the phenotypic variation within populations compared to additive genetic effects, influencing offspring morphology, physiology, and phenology, among other traits (meta-analyses by Moore *et al.* 2019; Sánchez-Tójar *et al.* 2020; Yin *et al.* 2019). Mothers can exert these effects both before and after birth, but prenatal maternal effects are generally expected to have larger and more irreversible impacts on offspring than postnatal effects, with consequences on offspring survival and reproductive success (Groothuis *et al.* 2005a; Marshall & Uller 2007; Mousseau & Fox 1998; Uller 2008; Wolf *et al.* 2011).

Egg-laying species are key models for understanding the adaptive significance of maternal effects as well as the underlying mechanisms. Unlike in most mammals, embryos develop outside the mother's body, limiting the time for direct maternal influence on the offspring. In egg-laying species, mothers can influence the offspring's prenatal environment by depositing compounds into the egg (e.g., Lovern & Wade 2001; McCormick 1999; Mentesana *et al.* 2019; Valcu *et al.* 2019; Weiss *et al.* 2011). In 1993, Schwabl first reported maternal hormones in bird eggs. These findings then were extended to other egg-laying vertebrates like fishes, reptiles, and amphibians, highlighting egg hormones as mediators of maternal effects (e.g., Groothuis *et al.* 2005a). Since then, many studies have suggested that hormones, especially steroid hormones present in the yolk, shape offspring growth, behavior (e.g., begging, aggression, boldness, activity, food competitiveness), and physiological systems (e.g., immune system, HPA axis, oxidative condition; (von Engelhardt & Groothuis 2011; Gil 2008; Groothuis *et al.* 2005a; Lovern & Wade 2001; Mentesana *et al.* 2021a). Yet, whether such phenotypic changes translate into fitness benefits for the mother, the father, or the offspring remains under debate (reviewed by Bebbington & Groothuis 2021), especially for wild populations (e.g., Tschirren *et al.* 2014).

Egg hormone-mediated maternal effects are often considered an adaptive mechanism, where mothers are assumed to adjust hormone concentrations in response to environmental cues to shape offspring phenotypes and optimize maternal and/or offspring fitness in variable but predictable environments (reviewed by Groothuis *et al.* 2005a). However, maternal egg hormone concentrations may also lack adaptive value and instead reflect unavoidable physiological processes, such as changes in maternal hormone concentrations that passively diffuse into the egg during formation, leading to varying hormone allocation (reviewed by Gil 2008; Groothuis &

Schwabl 2008). Hence, even 30 years after the discovery of maternal hormones in eggs, the evolutionary consequences of hormone-mediated maternal effects remain unknown.

There are several reasons why the evolutionary consequences of maternal egg hormones remain unclear. First, the influence of maternal hormones on fitness is expected to vary greatly, depending not only on the group of hormones studied, but also on the fitness proxy used, which can result in contrasting fitness effects. For example, egg androgens (i.e., androstenedione, testosterone, and 5-α dihydrotestosterone) and thyroid hormones (i.e., TH3 and TH4) can lead to increased behavioral competitiveness, boost early growth, and ultimately survival (von Engelhardt & Groothuis 2011; Groothuis *et al.* 2005a; Hsu *et al.* 2017; Ruuskanen *et al.* 2018). However, high concentrations of egg androgens can also impair the offspring's immune system (reviewed by Gil 2008) and oxidative condition (e.g., Treidel *et al.* 2013), potentially decreasing survival. Regarding egg glucocorticoids, they have been reported to alter behaviors such as antipredator response, impair physiological systems beneficial for offspring (e.g., immune system), and decrease offspring body condition, growth rate, and ultimately survival (Haussmann *et al.* 2012; Polich *et al.* 2018; Stier *et al.* 2009). Second, egg hormones might have sex-specific effects on offspring fitness. A meta-analysis on bird studies showed that experimental manipulation of androgen and glucocorticoid concentrations via egg injection tended to lead to larger effect sizes in male compared to female offspring (Podmokła *et al.* 2018b). However, the generality of this finding is debated, as the study only examined the relationship between experimentally injected eggs and offspring fitness proxies in seven populations of wild birds - six of which provided data only on androgens, and one on glucocorticoids. Third, the influence of maternal hormones on fitness might be ontogeny-dependent. Maternal effects are expected and have been observed to weaken throughout offspring ontogeny; however, recent meta-analyses on maternal and parental effects more generally found only partial (Moore *et al.* 2019) or inconclusive evidence for such a weakening effect (Sánchez-Tójar *et al.* 2020). Fourth, theory predicts that the evolution of maternal effects generally, and of hormone allocation specifically, should be shaped by conflicting fitness optima between offspring, mothers, and fathers (Bebbington & Groothuis 2021; Marshall & Uller 2007). For example, high concentrations of egg androgens might increase maternal reproductive success and offspring survival by boosting offspring growth. But at the same time, high concentrations of egg androgens might decrease maternal life expectancy by increasing her workload or even reduce reproductive success if the increased workload cannot be fulfilled through parental care. Overall, due to the differential effects expected and observed on reproductive and survival proxies, the net fitness consequences of egg hormone concentrations are difficult to predict. Finally, besides biological reasons, methodological aspects might also add

uncertainty to our understanding of the evolutionary importance of maternal egg hormones (reviewed by Groothuis *et al.* 2020; Groothuis & Von Engelhardt 2005). A meta-analysis examining the effects of different experimental manipulations of egg androgens and glucocorticoids in wild and captive bird populations provided evidence that factors such as study type (correlational *vs* experimental), manipulation methods (e.g., maternal *vs* egg manipulation), and hormone doses (e.g., within natural range *vs* supra-physiological concentrations) can influence the relationship between egg hormones and fitness proxies (Podmokła *et al.* 2018).

Here, we used an evidence synthesis approach to test the extent to which an increase in prenatal maternal egg hormone concentrations relates to fitness in wild populations of egg-laying vertebrates and to understand the patterns of the effects (i.e., the sources of heterogeneity). We tested six pre-registered hypotheses and predictions, which we divided into biological and methodological categories (pre-registration: Mentesana *et al.* 2021b). By pre-registering our study, we enhanced the distinction between hypothesis testing and hypothesis generation, thereby improving the credibility of our research findings (Ihle *et al.* 2017). Note that we mostly focused on the relationship between egg hormones and offspring fitness rather than parental fitness because the former has received the most empirical attention. Indeed, due to lack of studies, we could not investigate the relationship between egg hormones and paternal fitness and could only conduct an exploratory analysis for egg hormones and maternal fitness.

1. Biological hypotheses:

BH.1. Before testing our pre-registered predictions for each group of hormones, we first calculated the overall relationship between egg hormones and offspring fitness proxies, regardless of hormone group, and estimated heterogeneity across effect sizes. Next, for each hormone group, we predicted a positive association between egg androgens (BH.1.1) and thyroid hormones (BH.1.3) with offspring fitness proxies, and a negative one with egg glucocorticoids (BH.1.2). We did not have directional predictions for progesterone and estradiol (BH.1.4 & BH.1.5) due to insufficient data (reviewed by Gil 2008; Groothuis *et al.* 2019). For the remaining hypotheses and predictions, we assumed the same hormone-specific directional predictions as detailed in BH.1.

BH.2. Few studies to date have addressed whether egg hormones have sex-specific fitness consequences for the offspring in wild populations. Given some evidence that androgens and glucocorticoids may have a stronger impact on males than females (Podmokła *et al.* 2018), we predicted for these two hormones larger effect sizes in males. We also tested for a sex-specific effect in the remaining hormones (thyroids, progesterone,

and estradiol), but because there is little data on their association, we did not have directional sex-specific predictions.

BH.3. Since there is partial evidence that maternal effects weaken throughout ontogeny (Moore *et al.* 2019), we expected the relationships between egg hormones and offspring fitness proxies to lessen from early to later life stages. For both altricial and precocial species, we predicted that the association would weaken from hatching through the nesting period (excluding hatching day) and into adulthood (i.e., before and after offspring independence; BH.3.1). Also, for altricial species, we predicted that the association would become weaker throughout the nesting phase (BH.3.2).

2. Methodological hypotheses:

MH.1. Egg hormones are indicated as key mechanisms underlying maternal effects. We therefore predicted that the association between egg hormones and fitness proxies would be stronger in experimental than correlational studies.

MH.2. Since females can modify the transfer of circulating hormones to their eggs (reviewed by Groothuis & Schwabl 2008), we predicted that the association between egg hormones and fitness proxies would be stronger for direct egg manipulations than for female manipulations.

MH.3. The dose of hormones used may differentially affect offspring fitness. While high doses of hormones within the natural range are expected to result in stronger fitness effects, doses exceeding naturally occurring concentrations may lead to weaker or even opposite effects on fitness. Following Podmokła *et al.* (2018), we predicted that manipulations within the natural range would lead to stronger fitness effects than supra-physiological concentrations for androgens, but not glucocorticoids, whereas we did not have directional predictions for the remaining hormones (thyroids, progesterone, and estradiol).

## 2. Materials and Methods

### 2.1. Protocol

The study protocol was pre-registered on the Open Science Framework before data extraction began (Mentesana *et al.* 2021b). Our pre-registration specified the purpose of the study, our a priori hypotheses and predictions, the data collection procedures, and the entire analytic plan. Unless stated otherwise, we adhered to these pre-registered plans. Throughout, we follow the Preferred Reporting Items for Systematic reviews and Meta-Analyses in ecology and evolutionary

biology (PRISMA-EcoEvo Checklist: Supplementary Figure 1 & Supplementary Information S1; O'Dea *et al.* 2021).

## 2.2. Information sources and search

We conducted a systematic search of the literature from 1993 (i.e., the year when the first paper on maternal hormones was published) until the 22nd of August 2022. We did an advance search in both Web of Science Core Collection and Scopus. The search query was designed to find studies that: (1) tested inter- and/or transgenerational (plasticity) effects; (2) measured or experimentally manipulated egg hormone concentrations or maternal hormone concentrations just before or during egg laying; and (3) measured maternal and/or offspring fitness proxies. The full keyword search string, as well as further details, is provided in Supplementary Information S2.

## 2.3. Study selection and eligibility criteria

Our searches yielded 2404 records which we deduplicated using the R package revtool v.0.3.0 (Westgate 2018). We screened the titles and abstracts of all records using Rayyan (Ouzzani *et al.* 2016). In total, 178 articles passed the title-and-abstract screening and were subjected to full-text screening (for detailed information on the decision tree used see Supplementary Figure 2). We included studies that were conducted in free-living populations of vertebrates, measured or experimentally manipulated maternal and egg hormones, and reported fitness proxies. For both mother and offspring, we considered all fitness components related to their reproduction and/or survival (for our full list of fitness proxies used see Supplementary Information 3). For the offspring, we also considered morphological information because they are commonly accepted as predictors of post-fledging survival (e.g., meta-analysis by Ronget *et al.* 2018).

After full-text screening, we identified 79 articles meeting our criteria (see PRISMA diagram in Supplementary Figure 3). Three observers (LM, NMA and PBD) evenly performed all the screening and double-screened ~20% of the records to confirm the reproducibility of the procedure (12% for title-and-abstract screening and 30% for full-text screening). The three observers collectively discussed and resolved conflicting decisions (~7% conflicted papers).

Although our pre-registered plan aimed to test for the relationship between egg hormones and fitness proxies in all egg-laying vertebrate groups, only 2 out of the 79 potentially suitable studies that we found were conducted on egg-laying vertebrates other than birds; thus, we decided to focus our study solely on birds.

## 2.4. Data collection and extraction

The data extraction was evenly performed by three observers (LM, NMA and PBD). Two observers (LM and AST) double-checked 36% of the extracted data to confirm the reproducibility of this process, LM double-checked the remaining data in search of potential errors, and both observers discussed and fixed any typos found.

We extracted data from text, tables or figures. For the latter, we used the R package metaDigitise v.1.0.1 (Pick *et al.* 2019). In addition, we also obtained raw data directly from supplementary materials and/or from published datasets. Complete data extraction was possible for 22 articles (29% of the dataset). To obtain the missing data from the remaining articles, we contacted 32 authors (55 articles) using a standardized email template (Supplementary Information S4), which allowed us to obtain complete data from an additional 21 articles (38% of the dataset). Eight authors communicated that the data were lost, one author expressed that they were working outside science and did not have time to search for the data, another could not provide the data due to being on sabbatical leave, and five authors did not reply. During author correspondence, one author made us aware of suitable data in the supplementary section of another study (Duckworth *et al.* 2015), which we added to our dataset. In total, we obtained complete data from 47 articles and partial data for 10 articles (N = 57); that is, we had to exclude 20 in-principle suitable articles (26%) due to preventable reporting issues (for more discussion, see Hennessy *et al.* 2022).

### 2.5. Extracted variables

From each article we extracted information regarding the study (e.g., title, publication year, author information, journal), the study subjects (e.g., species, developmental mode, mean clutch size), the study area (e.g., location, latitude, longitude and elevation of study area), the methodology of the study (e.g., egg collection, sampling method, subject of the experiment, method and dose use), the fitness proxy studied (e.g., maternal/offspring, trait measured, offspring life history stage), the results (e.g., statistical test used, statistical value, mean values for control and experimental groups), and information on how data were provided (e.g., data location, data source). For the complete list of continuous and categorical moderators see Supplementary Information S5.

Note that for some fitness proxies, it is difficult to fully disentangle maternal and offspring fitness (e.g., hatching or fledging success; discussed by Wolf & Wade (2001). In such cases, following the recommendations of Wolf & Wade (2001), we took a conservative approach, assigning to the mother those fitness proxies measured in the offspring that are largely or wholly controlled by her.

2.6. Effect size calculation

To quantify the relationship between egg hormones and fitness, we used Pearson's correlation coefficients (hereafter Pearson's *r*). For correlation coefficients other than Pearson's *r* (e.g., Spearman's rho correlation coefficient), percentages, absolute values or inferential statistics (e.g., F-value, t-value, Chi-square value), we calculated Pearson's *r* using the equations from Nakagawa & Cuthill (2007) and Lajeunesse (2013; Supplementary Information S6). For tests that lacked information on directionality, we determined it from the text, figures, or by directly asking the authors via email. We then calculated the variance in Pearson's *r* (*Vr*) as $Vr = (1 - r^2)^2 \div (N - 1)$ (Borenstein *et al.* 2009). For experimental studies in which two groups were compared (control *vs* treatment), we transformed the means, SDs, and sample sizes of the fitness proxies of interest of each group into biserial correlations coefficients and calculated their sampling variance following Jacobs & Viechtbauer (2017) with the function 'escalc' from the R package "metafor" v.2.4-0 (Viechtbauer 2010). Biserial correlations are conceptually equivalent and directly comparable to *r* (Jacobs & Viechtbauer 2017). Meta-analyses involving both Pearson's and biserial correlation coefficients need to be based on the raw coefficients; this is why we did not use Fisher's *r*-to-*z* transformation (Jacobs & Viechtbauer 2017). When more than two groups were compared (e.g., control *vs* treatment 1 *vs* treatment 2), we calculated effect sizes between the control and each experimental group and divided the sample size of the control group by the number of times the control group was used (N/2 in the previous example) to account for shared-control non-independence. Whenever more than one effect size could be extracted or calculated for the same data (e.g., when the same result was provided using more than one statistical model), we chose one source using the following order of preference: (a) Pearson's *r*; (b) other correlation coefficients, (c) mean/percentages/absolute differences between control and experimental groups, and (d) inferential statistics. The sample size assigned to each effect size (i.e., number of clutches or number of offspring) reflected the replication level for the specific trait and experimental design reported by the authors in the papers. If the sample size was not reported at that level or the reported sample size did not correspond to the correct replication level, we used the sample size for the other replication level (if available) or estimated it from the reported degrees of freedom (df + 1) whenever possible (18 effect sizes).

In all cases, effect sizes were coded so that fitness proxies were positively correlated with overall fitness. Specifically, for 'egg mortality,' 'hatching failure,' and 'offspring mortality,' the sign of the effect sizes (k = 5 effect sizes, N = 3 studies) was inverted to ensure comparability across all fitness proxies.

## 2.7. Phylogenies and variance-covariance matrices

We built phylogenetic trees by searching for species in the Open Tree Taxonomy (Rees & Cranston 2017) and extracting their phylogenetic information from the Open Tree of Life (Hinchliff *et al.* 2015) using the R package "rotl" v.3.0.5 (Michonneau *et al.* 2016). We calculated tree branch length following Grafen (1989) with the R package "ape" v.5.4-1 (Paradis & Schliep 2019) and constructed a phylogenetic correlation matrix that was included in all the phylogenetic multilevel models (Figure 1). In addition, we specified sampling variances as a variance-covariance matrix rather than a vector of sampling variances. This variance-covariance matrix had the sampling variance for each effect size on the diagonal and the covariance between these measures, which was calculated assuming a 0.5 correlation between effect size sampling variances from the same study (more in Noble *et al.* 2017), as off-diagonal elements (see Supplementary Information S7 for a sensitivity analysis using a vector of sampling variances rather than the variance-covariance matrix).

## 2.8. General information on analyses

All phylogenetic multilevel meta-analyses and meta-regressions were run using the R package "metafor" v.2.4-0 (Viechtbauer 2010) and included the following random effects: (a) study ID, which encompasses effect sizes extracted from the same study, (b) laboratory ID, which contains effect sizes obtained from the same laboratory, (c) population ID, which considers effect sizes derived from the same population, (d) species name, which encompasses effect sizes derived from the same species, (e) phylogeny, which consists of the phylogenetic correlation matrix and informs about the relatedness between species, and (f) effect ID, which is a unit-level effect and represents residual/within-study variance.

We present our results by first describing the direction and magnitude of the effect, followed by the statistical significance, and finally by reporting the heterogeneity around the mean estimates. This approach places less focus on dichotomizing the results (discussed by e.g., Amrhein *et al.* 2019) and provides crucial insights into the variance across effect sizes, helping assess the generalizability of our findings (discussed by e.g., Gurevitch *et al.* 2018). We provide information on the effect size, both 95% confidence intervals (CI) and 95% prediction intervals (PI), and p-values. Confidence intervals show the range in which the overall average effect is likely to be found, whereas prediction intervals (which incorporate heterogeneity) estimate the likely range in which 95% of effects are expected to occur in studies (past or future) similar to those included in the current study (IntHout *et al.* 2016). For the intercept-only models, we first

calculated $Q$ and then calculated total heterogeneity ($\sigma^2$), its source (or relative heterogeneity; $I^2$) and two different metrics to understand its magnitude – a mean-standardized metric (*CV*) and a variance-mean-standardized metric (*M*; more in Yang *et al.* 2024). For meta-regression models, we estimated the percentage of heterogeneity explained by the moderators using $R^2_{marginal}$ (Nakagawa & Schielzeth 2013). We graphically presented the results as orchard plots using the R package "orchaRd" v.2.0 (Nakagawa *et al.* 2023).

2.9. Main effect model

To estimate the overall effect size for the association between egg hormones and fitness across studies (BH.1), we fitted a phylogenetic multilevel intercept-only meta-analytic model for our entire dataset (k = 443 effect sizes, N = 57 studies). Note that this model was not pre-registered and includes data for all hormones although we predicted a positive association for androgens and thyroids, and a negative association for glucocorticoids. We used this approach to understand the overall heterogeneity in our data before we proceeded to explain it with our meta-regressions (see below).

To assess the robustness of our results to the choice of effect size, we additionally ran three sensitivity models that, although not stated in our pre-registration, we considered important to confirm that our results were robust to the choice of effect size. For that, we ran three intercept-only models using three subsets of the data: i) Pearson's *r* (k = 166, N = 40), ii) biserial correlations (k = 277, N = 39), and iii) log-response ratio coefficients (lnRR, k = 274, N = 38; calculated using the function 'escalc' from the R package "metafor" v.2.4-0, Viechtbauer 2010). These three sensitivity models reached similar conclusions as the main effect model (Supplementary Information S8).

2.10. Biological hypotheses

To investigate the association between maternal egg hormones and offspring fitness proxies (BH.1), we first subsetted the data to only include effect sizes on offspring fitness (k = 355, N = 45). We then split the data into three subsets containing information on androgens (i.e., androstenedione, 5α-dihydrotestosterone and testosterone; BH.1.1), glucocorticoids (BH.1.2) and thyroids (i.e., TH3 and TH4; BH.1.3). Next, we fitted three multilevel intercept-only meta-analytic models. Note that in our pre-registration we also proposed to study egg progesterone and estradiol (BH.1.4 & BH.1.5); however, we could not study these hormones because no study fulfilled our eligibility criteria. To assess our results robustness to our choice of analytical strategy for these three hypotheses (BH.1.1 - BH.1.3), we ran a phylogenetic multilevel meta-regression

with ‘hormone type’ (levels: androgens, glucocorticoids and thyroids) as a moderator. Although the mean estimate for glucocorticoids (k = 56, N = 8) and thyroids (k = 36, N = 3) was slightly different between the two analytical approaches, likely due to the much smaller sample sizes for those hormones compared to androgens (k = 263, N = 34), the estimates remained statistically non-significant (Supplementary Information S9).

To test if maternal egg hormones have sex-specific fitness consequences for the offspring (hypothesis BH.2) we used the small subset of studies for which offspring sex was available (k = 45, N = 8). We fitted a phylogenetic multilevel meta-regression including ‘hormone type’ (levels: androgen and thyroids), ‘sex’ (levels: female, male), and their interaction as moderators. Note that we initially planned on including ‘glucocorticoids’ as another level within the moderator ‘hormone type’; however, none of the glucocorticoid studies provided offspring sex.

To test if maternal egg hormones have age-specific fitness consequences for the offspring (BH.3) we fitted two meta-regression models. In the first model (BH.3.1), we originally planned to test if the hormone-fitness relationship differed depending on ‘hormone type’ (levels: androgens, glucocorticoids, and thyroids), ‘offspring life history stage’ (levels: hatching, before independence, and after independence), and their interaction. However, our final model contained only androgen data and the moderator ‘offspring life history stage’ with three levels (hatching, before independence, and after independence; k = 263, N = 34) because we did not have enough data points for the levels glucocorticoids, thyroids and after independence (i.e., we had less than 5 data points per level; more in our pre-registration Mentesana *et al.* 2021b). In the second model (BH.3.2), we included ‘hormone type’ (levels: androgens, glucocorticoids and thyroids), ‘offspring relative age’ as a continuous moderator, and their interaction (k = 215, N = 19).

### 2.11. Methodological hypotheses

We performed three phylogenetic multilevel meta-regressions to investigate whether the relationship between maternal egg hormones and both maternal and offspring fitness proxies differed between correlational and experimental studies. In the first model (MH.1), we included ‘hormone type’ (levels: androgens, glucocorticoids and thyroids), ‘study type’ (levels: correlational and experimental), and their interaction as moderators (k = 404, N = 54). In the second model (MH.2), we fitted ‘hormone type’, ‘experimental study object’ (levels: egg and mother), and their interaction as moderators (k = 250, N = 36). Note that due to a lack of correlational studies testing for the effects of thyroids on fitness and of experimental studies manipulating maternal thyroid hormones, we could only test the first two models (MH.1 & MH.2) for androgens and glucocorticoids. In the third model (MH.3), we originally planned to include ‘hormone type’,

‘experimental dose’ (levels: within natural range and supraphysiological), and their interaction as moderators; however, we could only do so for androgens (k = 89, N = 19) due to a lack of data for the other two hormones, and thus, our final model only included ‘experimental dose’ as moderator.

2.12. Exploratory hypotheses

We performed three additional pre-registered biological and methodological exploratory meta-regressions for which we did not have directional predictions (Mentesana *et al.* 2021b). To test the effect of maternal egg hormones on maternal fitness (BEH.1) we subsetted the data to only include effect sizes on maternal fitness proxies (k = 88, N = 46). We originally planned to run a phylogenetic multilevel meta-regression with ‘hormone type’ (levels: androgens, glucocorticoids and thyroids) as a moderator. However, we could only test this for androgens and glucocorticoids (k = 85, N = 43). To test if the effect of maternal egg hormones on offspring fitness differs depending on the matrix in which the hormone was measured/manipulated (BEH.2), we ran a phylogenetic multilevel meta-regression using ‘egg site measured’ (levels: albumen and yolk) as moderator. We tested this hypothesis only for glucocorticoids (k = 38, N = 4), which was the hormone that had sufficient data (i.e., at least 5 data points per level) in both the yolk and albumen. Finally, to test the effect of maternal egg hormones on fitness depending on the sampling technique used, we ran a phylogenetic multilevel meta-regression using ‘egg sampling method’ (levels: biopsy and entire egg removed) as moderator (MEH.1), and we were able to test this only for androgens (k = 91, N = 9).

2.13. Publication bias tests

We assessed publication bias, both small-study and decline effects, following Nakagawa *et al.* (2022). All models employed the same random effect structure as reported for the analyses above. We used the complete database containing information on both maternal and offspring fitness. For small-study effects, we conducted a multilevel meta-regression, including the ‘square root of the inverse of the sample size’ of each effect size as a moderator. Additionally, we investigated temporal patterns in the data that could highlight the existence of decline effects (i.e., effect sizes decreasing over time; Sánchez-Tójar *et al.* 2018; Trikalinos & Ioannidis 2005) by conducting a multilevel meta-regression where the ‘year of publication’ was mean-centered and included as a moderator. Note that these two analyses using the entire dataset were not pre-registered, but we added them to provide an informative overall overview of small-study and decline effects. We also tested for the existence of small-study and decline effects for each

hormone separately by rerunning the same two models described above but adding the interaction between the ‘square root of the inverse of the sample size’ or ‘year of publication’ and 'hormone type' (levels: androgens, glucocorticoids, and thyroids), respectively.

In addition, to test whether effect sizes are larger in studies that selectively or incompletely report statistically significant results, we conducted two meta-regression models. To test the effect of partial or selective reporting, we included 'partial or selective reporting' (levels: yes and no), 'hormone type,' and their interaction as moderators. To examine the effect of data reporting completeness, we initially planned to include 'data reporting completeness' (levels: complete and incomplete), 'hormone type,' and their interaction as moderators. However, due to insufficient data for glucocorticoids and thyroids, we could only test this for androgens (k = 337, N = 53), so our final model included 'data reporting completeness' as the only moderator. We also had pre-registered a meta-regression to test if blind data collection leads to smaller effect sizes as shown in previous studies (Freeberg *et al.* 2024; Holman *et al.* 2015; Keaney *et al.* 2024), but we could not perform this test because only 3 studies reported to have used a blind data collection protocol.

Lastly, we ran a multi-moderator multilevel meta-regression to estimate adjusted overall effects, and the percentage of heterogeneity explained by those moderators combined ($R^2_{marginal}$). For this, we mean-centered and included 'study type' as a moderator, as well as the interaction between the 'square root of the inverse of the sample size' and 'hormone type', and the interaction between 'year' and 'hormone type'.

## 3. Results

Overall, we obtained 443 effect sizes from 57 studies across 19 bird species, 4 bird orders, and 34 different populations (Figure 1). All studies but one were conducted in the Northern Hemisphere. Three species accounted for almost 60% of the effect sizes (i.e., *Sturnus unicolor*, k = 105 effect sizes; *Parus major*, k = 92; *Ficedula albicollis*, k = 69). Androgens were the most studied hormones (k = 338), followed by glucocorticoids (k = 66) and thyroid hormones (k = 39), and most effect sizes (77.2%) were obtained from experiments (k = 342). The fitness proxies studied corresponded mainly to offspring (k = 355) rather than mothers (k = 88), and those proxies

were primarily obtained before offspring independence (k = 315) and without quantifying them by offspring sex (k = 310).

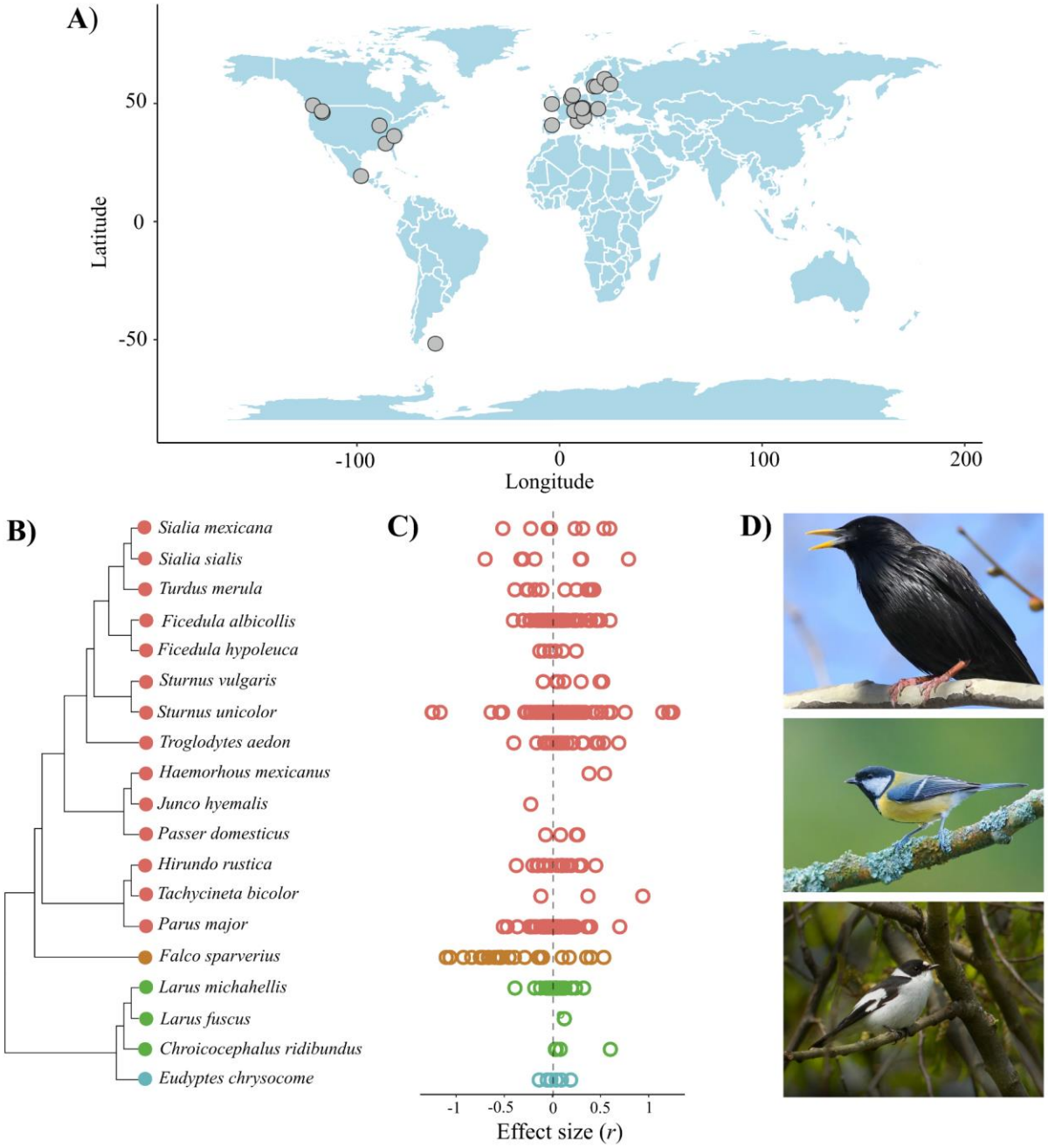


Figure 1: Overview of the meta-analytic data analyzed to test the fitness consequences of maternal egg hormones. A) Global map showing the location of each study included in the meta-analysis. Each point represents one study area for which one or more effect sizes were obtained. (B) Phylogenetic tree of the 19 bird species included in the meta-analysis, along with (C) their respective effect sizes. Each color represents a different bird order (red: Passeriformes, brown: Falconiformes, green: Charadriiformes, and blue: Sphenisciformes). D) Images of the bird species that contributed to most of the data (top to bottom: *Sturnus unicolor*, *Parus major*, *Ficedula albicollis*). All images are copyright-free (CC BY-SA 4.0 DEED. Authors: Peterwchen, Luc Viatour, Frank Vassen, respectively) and were extracted from www.commons.wikipedia.org

### 3.1. Main effect model

The intercept-only model revealed a negative, near-zero, and statistically nonsignificant association between maternal egg hormones and fitness - both offspring and maternal - across wild birds (Figure 2; $r$ = -0.073, 95% CI: [-0.325, 0.179], 95% PI: [-0.844, 0.699], p-value = 0.571, k = 443, n = 57). Total absolute heterogeneity ($\sigma^2$ = 0.138; $Q$ = 48193, p-value < 0.001), relative heterogeneity ($I^2_{total}$ = 95.2%), and the two magnitude metrics ($CV_{total}$ = 5.1, $M_{total}$ = 0.6) showed high values of heterogeneity, which ranged between the 50$^{th}$ and 75$^{th}$ percentile of empirically derived ones in ecological and evolutionary meta-analyses (more in Yang *et al.* 2024; for details see Supplementary Figure 4 & Supplementary Information S10).

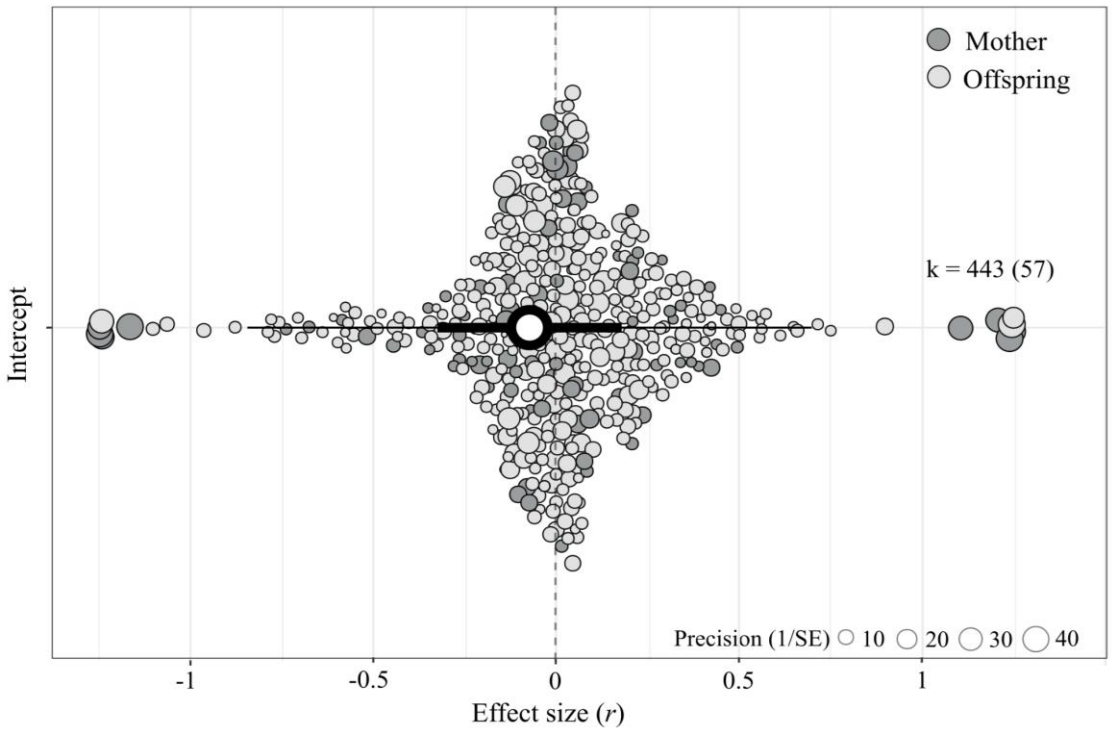



Figure 2: Overall, egg hormone concentrations (androgens, glucocorticoids, and thyroid hormones) show a near-zero association with both maternal (dark gray circles; ca. 20% of effect sizes) and offspring (light gray circles; ca. 80% of effect sizes) fitness in wild birds. Orchard plot of the phylogenetic multilevel (intercept-only) model showing the meta-analytic mean (black circle), 95% confidence intervals (thick whisker), 95% prediction intervals (thin whisker), and individual effect sizes scaled by their precision (gray circles). k is the number of individual effect sizes and the number of studies is shown in brackets.

## 3.2. Biological hypotheses

### 3.2.1. Offspring fitness effect sizes (BH.1)

A non-pre-registered intercept-only model confirmed the negative, near-zero, and statistically nonsignificant association between maternal egg hormones and offspring fitness across wild birds ($r$ = -0.073, [95% CI = -0.427, 0.282], [95% PI = -0.924, 0.779], p-value = 0.687, k = 355; N = 43) as well as its high levels of heterogeneity ($\sigma^2$ = 0.155; Q = 13192, p-value < 0.001; $I^2_{total}$ = 94.9%; $CV_{total}$ = 5.4, $M_{total}$ = 0.6).

Moreover, effect sizes for the association between all three types of egg hormones and offspring fitness remained small and statistically nonsignificant (Figure 3; Table 1) and the moderator hormone type explained negligible heterogeneity ($R^2_{marginal}$ = 0.4%). Moreover, none of the three estimates differed statistically from each other (p-value androgens *vs* glucocorticoids = 0.314; androgens *vs* thyroids = 0.411; glucocorticoids *vs* thyroids = 0.969).

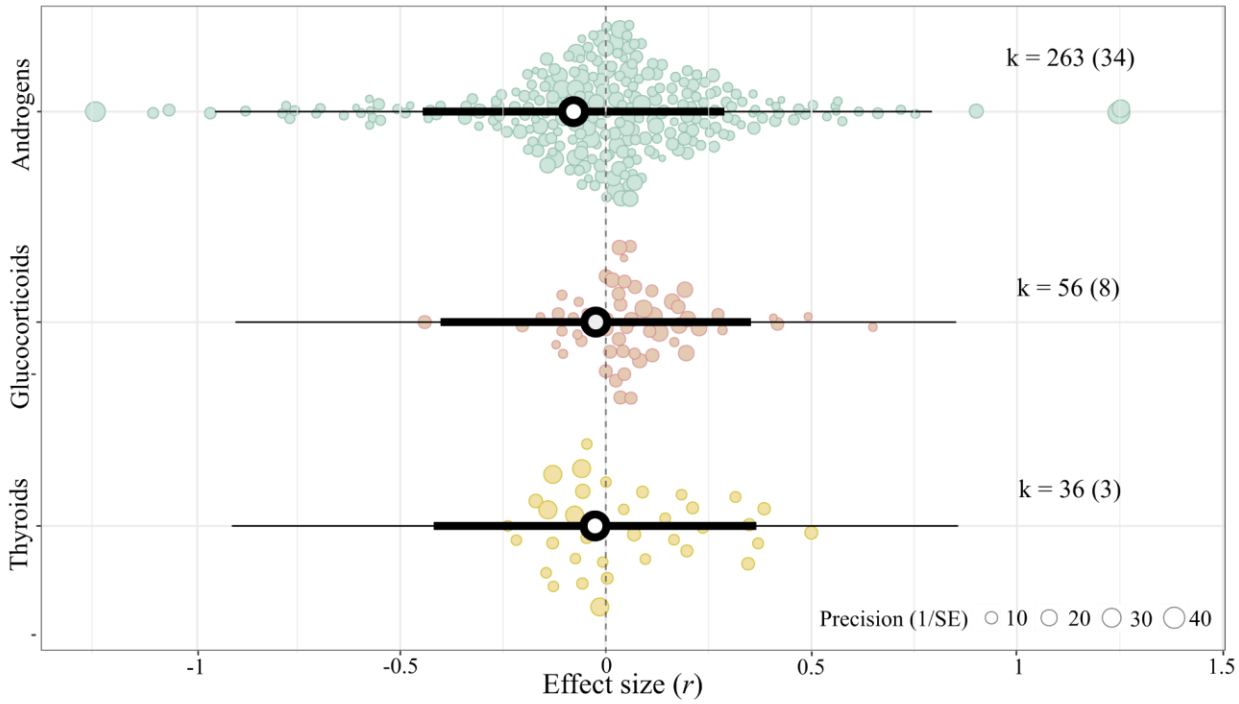



Figure 3: Regardless of hormone type, egg hormone concentrations show a near-zero association with offspring fitness in wild birds. Orchard plot of the phylogenetic multilevel meta-regression showing the mean estimates (black circles), 95% confidence intervals (thick whisker), 95% prediction intervals (thin whisker), and individual effect sizes scaled by their precision (coloured circles). k is the number of individual effect sizes, and the number of studies is shown in between brackets.

| | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | p-value | k | N |
|---|---|---|---|---|---|---|---|---|
| Androgens | -0.086 | -0.446 | 0.274 | -0.946 | 0.774 | 0.638 | 263 | 34 |
| Glucocorticoids | -0.028 | -0.398 | 0.343 | -0.892 | 0.837 | 0.882 | 56 | 8 |
| Thyroids | -0.024 | -0.410 | 0.360 | -0.896 | 0.847 | 0.901 | 36 | 3 |

Table 1: Results of a phylogenetic multilevel meta-regression (without intercept) model testing the association of egg hormone concentrations on offspring fitness depending on hormone type (androgens, glucocorticoids, and thyroids). CI: confidence interval; PI: prediction interval; L: lower; U: upper; k = number of effect sizes; N = number of studies.

3.2.2. Sex-specific offspring fitness effect sizes (BH.2)

The association between both egg androgens and thyroids with offspring fitness tended to be negative regardless of offspring sex, but none of the estimates were statistically significant (Figure 4; Table 2). Both moderators combined explained a moderate amount of heterogeneity for this subset of the data ($R^2_{marginal}$ = 3.8%). The post-hoc tests showed no statistical difference between males and females for both androgens (p-value = 0.262) and thyroids (p-value = 0.065).

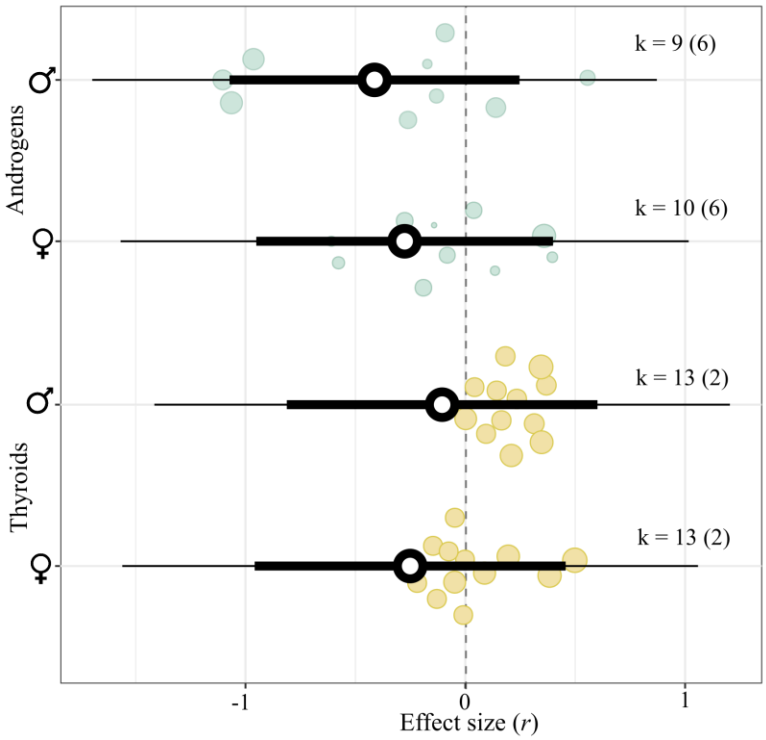



Figure 4: Regardless of offspring sex, androgen and thyroid egg hormone concentrations show a near-zero association with offspring fitness in wild birds. Orchard plots of the phylogenetic multilevel meta-regression showing male and female mean estimates (black circles), 95% confidence intervals (thick whisker), 95% prediction intervals (thin whisker), and individual effect sizes scaled by their precision (coloured circles). k indicates the number of individual effect sizes and, in between brackets, the number of studies.

| | | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | p-value | k | N |
|---|---|---|---|---|---|---|---|---|---|
| Androgens | female | -0.276 | -0.951 | 0.399 | -1.568 | 1.016 | 0.414 | 10 | 6 |
| | male | -0.413 | -1.072 | 0.247 | -1.697 | 0.871 | 0.213 | 9 | 6 |
| Thyroids | female | -0.249 | -0.957 | 0.458 | -1.559 | 1.060 | 0.481 | 13 | 2 |
| | male | -0.104 | -0.811 | 0.603 | -1.413 | 1.205 | 0.768 | 13 | 2 |

Table 2: Results of the phylogenetic multilevel meta-regression models testing sex-specific egg hormone concentrations associations on offspring fitness. CI: confidence interval; PI: prediction interval; L: lower; U: upper; k = number of effect sizes; N = number of studies.

### 3.2.3. Age-specific offspring fitness effect sizes (BH.3)

For both altricial and precocial species, androgens (the only hormone for which we had enough data at different categorical age stages to perform this test; BH.3.1) tended to be positively associated with offspring fitness at hatching and negatively before and after independence. However, effect sizes were small and statistically nonsignificant (Figure 5A; hatching: $r$ = 0.015, [95% CI = -0.376, 0.405], [95% PI = -0.906, 0.936], p-value = 0.940, k = 21, N = 7; before independence: $r$ = -0.067, [95% CI = -0.434, 0.299], [95% PI = -0.979, 0.844], p-value = 0.718 , k = 226, N = 31; after independence: $r$ = -0.164, [95% CI = -0.594, 0.267], [95% PI = -1.102, 0.775], p-value = 0.454, k = 16, N = 4), and the moderator explained negligible heterogeneity ($R^2_{marginal}$ = 0.6%). None of the three age stages differed statistically from each other (p-value hatching *vs* before independence = 0.313; hatching *vs* after independence = 0.176; before *vs* after independence = 0.345).

In altricial species, the association between egg androgen or thyroid concentrations and offspring fitness proxies tended to weaken throughout the nesting phase, while the association with glucocorticoids showed the opposite trend (BH.3.2); however, none of those slopes were statistically significant (Figure 5B-D; Table 3) and the moderator only explained little heterogeneity ($R^2_{marginal}$ = 2.0%).

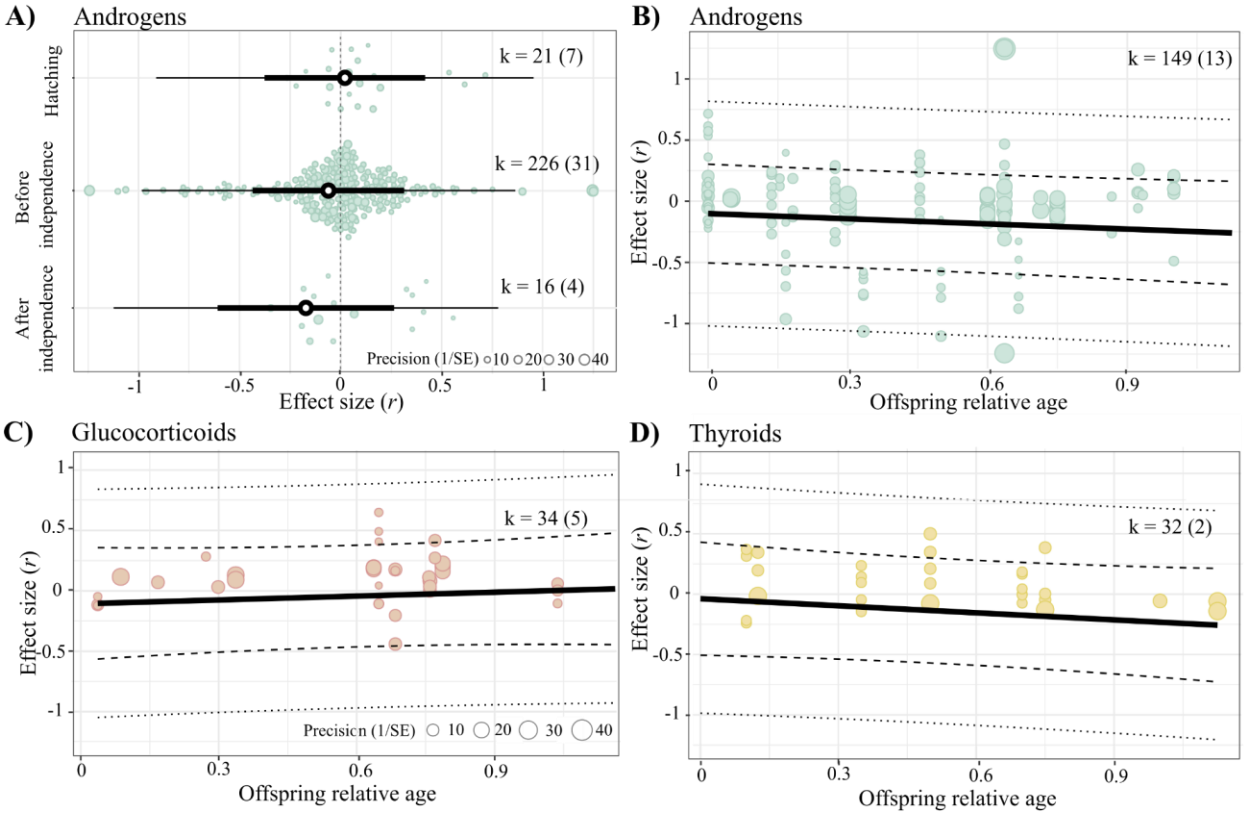



Figure 5: Regardless of offspring age and hormone type, egg hormone concentrations show a near-zero association with offspring fitness in wild birds. A) Egg androgen concentrations and offspring fitness in both altricial and precocial species. Orchard plot of the phylogenetic multilevel meta-regression displaying mean estimates (black circles), 95% confidence intervals (thick whiskers), 95% prediction intervals (thin whiskers), and individual effect sizes scaled by precision (colored circles). Panels B-D refer to B) androgen, C) glucocorticoid and D) thyroid egg hormone concentrations with offspring fitness in altricial species. We plotted a separate bubble plot for each of the hormone types analyzed in a single phylogenetic multilevel meta-regression, showing model estimates (solid lines), 95% CIs (thick dashed lines), 95% prediction intervals (thin dashed lines), and individual effect sizes by precision (colored circles: green for androgens, red for glucocorticoids, yellow for thyroids). For all panels, k denotes the number of effect sizes and, in brackets, the number of studies.

| | Estimate | SE | t-value | df | p-value | L 95% CI | U 95% CI |
|---|---|---|---|---|---|---|---|
| Intercept (Androgens) | -0.102 | 0.205 | -0.499 | 209 | 0.619 | -0.506 | 0.301 |
| Glucocorticoids | -0.002 | 0.132 | -0.018 | 209 | 0.986 | -0.257 | 0.262 |
| Thyroids | 0.063 | 0.138 | 0.458 | 209 | 0.647 | -0.209 | 0.336 |
| Offspring age | -0.141 | 0.087 | -1.615 | 209 | 0.108 | -0.313 | 0.031 |
| Glucocorticoids * Offspring age | 0.237 | 0.191 | 1.241 | 209 | 0.216 | -0.140 | 0.614 |
| Thyroids * Offspring age | -0.054 | 0.180 | -0.299 | 209 | 0.766 | -0.409 | 0.302 |

Table 3: Results of phylogenetic multilevel meta-regression models testing whether the association between egg hormone concentrations and offspring fitness vary throughout the nesting phase in altricial bird species. df: degrees of freedom; CI: confidence interval. Offspring age was standardized as the age of the offspring divided by the average nesting time and was calculated for each study based on information reported by the authors (more in Table S5.1).

## 3.3. Methodological hypotheses

### 3.3.1. Fitness effect sizes in correlational *vs* experimental studies (MH.1)

For the two hormone types for which sufficient data existed, androgen and glucocorticoids, the associations between concentrations and both maternal and offspring fitness tended to be negative, near-zero, and statistically nonsignificant (Figure 6A; Table 4A). The moderator explained little heterogeneity ($R^2_{marginal}$ = 1.5%), and the differences between correlational and experimental were statistically nonsignificant regardless of hormone type (p-value androgens = 0.124; glucocorticoids = 0.341).

### 3.3.2. Fitness effect sizes in experimental studies performed on the mother *vs* the egg (MH.2)

The negative but statistically nonsignificant association between egg androgen or glucocorticoid concentrations and offspring fitness remained independent of whether hormonal manipulations were performed on the mother or the egg (p-value mother *vs* eggs = androgens 0.412; glucocorticoids 0.911; Figure 6B; Table 4B). The moderator explained negligible heterogeneity ($R^2_{marginal}$ = 0.5%).

3.3.3. Fitness effect sizes in experimental studies conducted within *vs* above the natural physiological range (MH.3)

Although the association between egg androgens and offspring fitness tended to be positive at doses above the physiological range and negative at doses within the natural range, none of the estimates were statistically significant (Figure 6C; Table 4C) or differed from each other (p-value = 0.175). Moreover, the moderator explained little heterogeneity ($R^2_{marginal}$ = 2.9%).

Table 4: Estimates of three phylogenetic multilevel meta-regressions used to investigate the relationship between maternal egg hormones and fitness in: A) correlational *vs* experimental studies, B) experiments that manipulated the mother's hormone concentrations *vs* the egg's hormone concentrations, and C) experiments that manipulated maternal egg hormones within *vs* above the natural physiological range. CI: confidence interval; PI: prediction interval; L: lower; U: upper; k = number of effect sizes; N = number of studies.

| Fitness effects | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | p-value | k | N |
| **A)** *Correlational vs experimental studies* | | | | | | | | | |
| Androgens | Correlational | -0.001 | -0.268 | 0.267 | -0.787 | 0.786 | 0.996 | 88 | 8 |
| | Experimental | -0.096 | -0.346 | 0.154 | -0.877 | 0.685 | 0.450 | 250 | 39 |
| Glucocorticoids | Correlational | 0.068 | -0.247 | 0.384 | -0.735 | 0.872 | 0.670 | 13 | 2 |
| | Experimental | -0.054 | -0.336 | 0.228 | -0.846 | 0.737 | 0.706 | 53 | 8 |
| **B)** *Experimental subject* | | | | | | | | | |
| Androgens | Mother | -0.172 | -0.572 | 0.228 | -1.057 | 0.713 | 0.398 | 21 | 4 |
| | Egg | -0.073 | -0.419 | 0.273 | -0.935 | 0.790 | 0.679 | 185 | 26 |
| Glucocorticoids | Mother | -0.052 | -0.479 | 0.376 | -0.950 | 0.847 | 0.812 | 17 | 3 |
| | Egg | -0.069 | -0.451 | 0.314 | -0.946 | 0.809 | 0.724 | 27 | 3 |
| **C)** *Experimental dose* | | | | | | | | | |
| Androgens | Natural range | -0.113 | -0.450 | 0.223 | -1.112 | 0.885 | 0.505 | 82 | 17 |
| | Supraphysiol. range | 0.188 | -0.343 | 0.718 | -0.891 | 1.267 | 0.484 | 7 | 2 |

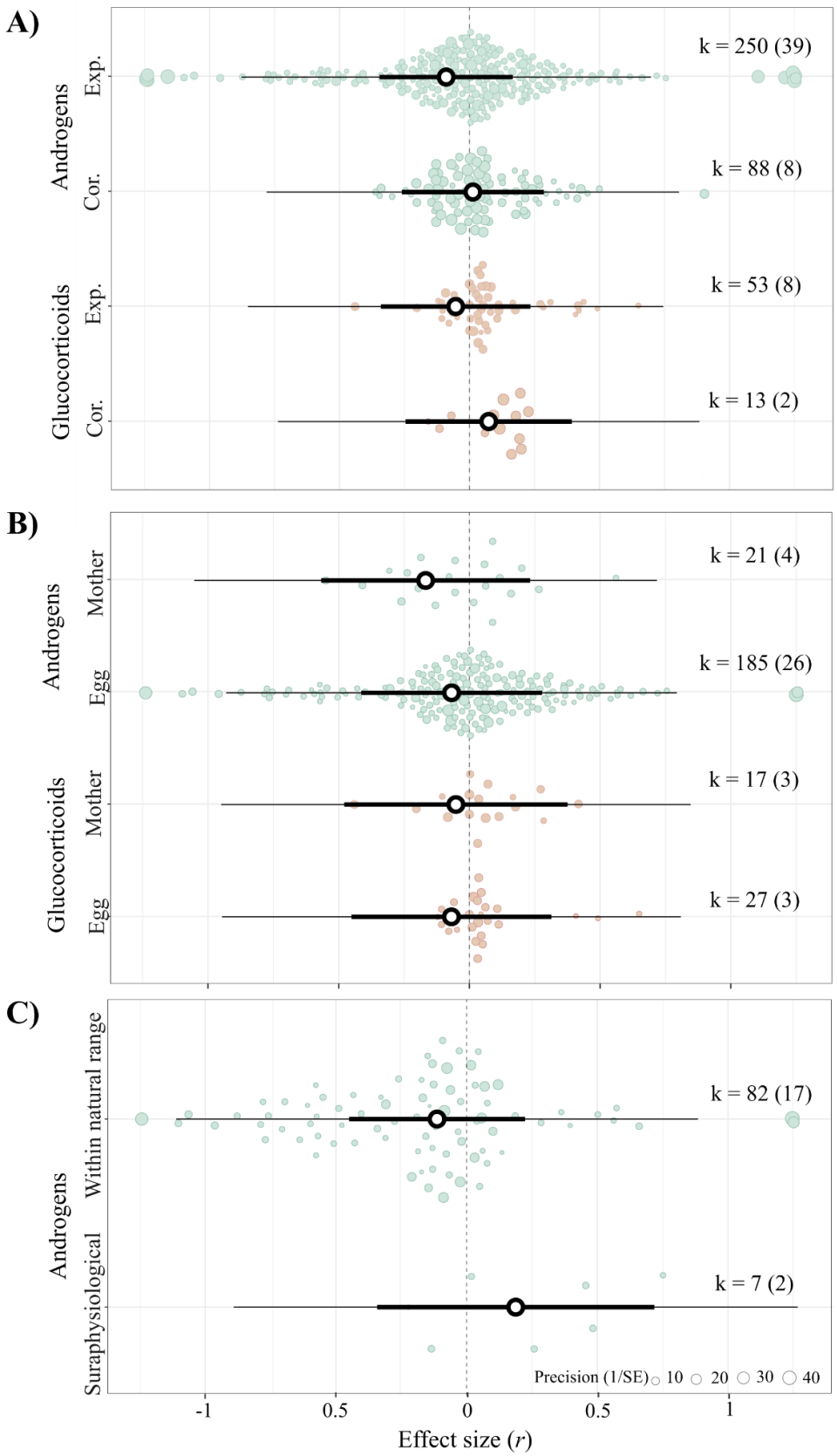

A)
Androgens
Exp.
Cor.
Glucocorticoids
Exp.
Cor.
k = 250 (39)
k = 88 (8)
k = 53 (8)
k = 13 (2)
B)
Androgens
Mother
Egg
Glucocorticoids
Mother
Egg
k = 21 (4)
k = 185 (26)
k = 17 (3)
k = 27 (3)
C)
Androgens
Within natural range
Suraphysiological
k = 82 (17)
k = 7 (2)
Precision (1/SE) 10 20 30 40
-1
0.5
0
0.5
1
Effect size (r)

Figure 6: The relationship between androgen and glucocorticoid egg hormone concentrations and fitness is near-zero, irrespective of A) the type of study performed (i.e., correlational or experimental), B) the study object on which the experiment was performed (i.e., the mother or the egg), and, solely for androgens, C) the experimental dose used (i.e., within or above the natural range). Orchard plots of three phylogenetic multilevel meta-regression models showing mean estimates (black circles), 95% confidence intervals (thick whisker), 95% prediction intervals (thin whisker), and individual effect sizes scaled by their precision (coloured circles). k indicates the number of individual effect sizes and, in between brackets, the number of studies.

### 3.4. Exploratory biological and methodological hypotheses

The association between egg hormones and maternal fitness was statistically nonsignificant both for androgens and glucocorticoids, and we lacked enough data for thyroid hormones (BEH.1; androgens: $r$ = -0.033, [95% CI = -0.148, 0.083], [95% PI = -0.935, 0.869], p-value = 0.574, k = 75, N = 36; glucocorticoids: $r$ = 0.092, [95% CI = -0.209, 0.392], [95% PI = -0.852, 1.035], p-value = 0.546, k = 10, N = 8). The moderator explained negligible heterogeneity ($R^2_{marginal}$ = 0.8%), and the two hormone groups did not differ statistically from each other (p-value = 0.444).

For glucocorticoids, the only hormone with sufficient, but still limited, data for both yolk and albumen (k = 19, N = 2), we found a statistically nonsignificant positive association between hormone concentrations and offspring fitness regardless of the egg site where the hormone was measured (BEH.2; albumen: $r$ = 0.028, [95% CI = -0.164, 0.219], [95% PI = -0.270, 0.325], p-value = 0.772; yolk: $r$ = 0.094, [95% CI = -0.100, 0.287], [95% PI = -0.205, 0.393], p-value = 0.332; $R^2_{marginal}$ = 8.1%; p-value albumen *vs* yolk = 0.625).

Finally, we found that the association between maternal egg androgen concentrations on fitness did not differ statistically between the two sampling techniques used (p-value = 0.196; MEH.1). When androgens were measured using an egg biopsy, they tended to negatively associated with fitness, whereas when the entire egg was removed, they tended to positively associated with fitness (egg biopsy: $r$ = -0.205, [95% CI = -0.708, 0.298], [95% PI = -1.070, 0.659], p-value = 0.208, k = 19, N = 2; entire egg removed: $r$ = 0.166, [95% CI = -0.094, 0.427], [95% PI = -0.584, 0.916], p-value = 0.208, k = 72, N = 7; $R^2_{marginal}$ = 15.5%).

Due to small sample sizes, these results should be interpreted cautiously as the models likely are heavily underpowered.

### 3.5. Publication bias

The evidence for the existence of overall small-study effects across the full dataset was statistically nonsignificant and the sign was contrary to our expectations (Figure 7A; intercept = -

0.148, [95% CI = -0.449, 0.153], p-value = 0.333; slope = 0.500, [95% CI = -0.180,1.180], p-value = 0.149; k = 443, N = 57; $R^2_{marginal}$ = 0.9%), suggesting that positive effect sizes are associated with effect sizes of lower precision across the full dataset combining all three hormones. In other words, some negative effect sizes of lower precision likely remain unpublished in the scientific literature, and although adjusting for their absence would not seemingly change the statistical significance of the mean effect, it could result in a ca. 100% increase in its magnitude from -0.073 to -0.148. Our model testing for hormone-specific small-study effects did not show statistically significant evidence of small-study effects for any of the three hormones (Figures 7B-D; $R^2_{marginal}$ = 1.2%), but the tendency described above suggesting that some negative effect sizes of lower precision may remain unpublished was present for all three hormones (adjusted effects: androgens: *r* = -0.092, [95% CI = -0.377, 0.192], [95% PI = -0.910, 0.725], k = 338, N = 46; glucocorticoids: *r* = -0.035, [95% CI = -0.352, 0.282], [95% PI = -0.864, 0.795], k = 66, N = 10; thyroids: *r* = 0.029, [95% CI = -0.354, 0.411], [95% PI = -0.828, 0.885], k = 39, N = 3; see also Supplementary Information S11).

We also found no clear evidence of decline effects either when combining all hormones together (intercept = -0.069, [95% CI = -0.322, 0.183]; slope = 0.001, [95% CI = -0.008, 0.010]; k = 443, N = 57; $R^2_{marginal}$ = 0.0%), or when testing for hormone-specific patterns via meta-regression (Supplementary Information S12; $R^2_{marginal}$ = 0.4%). The adjusted hormone-specific effects from the latter model remained statistically nonsignificant (androgens: *r* = -0.084, [95% CI = -0.343, 0.175], [95% PI = -0.864, 0.696], k = 338, N = 46; glucocorticoids: *r* = -0.039, [95% CI = -0.317, 0.239], [95% PI = -0.825, 0.747], k = 66, N = 10 ; thyroids: *r* = -0.027, [95% CI = -0.613, 0.559], [95% PI = -0.967, 0.913], k = 39, N = 3).

We were able to categorize whether researchers did or did not selectively report data for only 38.6% of effect sizes (k = 171, N = 25). Authors partially or selectively reported data in 88 cases, while in 83 cases they did not. For thyroid hormones, but not for androgens or glucocorticoids, we found evidence that selectively and/or partially reported results influenced the effect size (Supplementary Information S13). However, these results should be interpreted with extreme caution, as only one study (k = 10) on thyroid hormones selectively/partially reported outcomes, making it impossible to determine whether the observed effect size is due to reporting biases or the study's nature. The moderator explained a relatively high amount of heterogeneity ($R^2_{marginal}$ = 15.9%) and, within each hormone group, the two levels (i.e., no *vs* yes partial/selective reporting) did not differ statistically from each other (androgens: p-value = 0.796; glucocorticoids: p-value = 0.145; thyroids: p-value = 0.378).

We were able to examine the effect of data reporting completeness only for androgens. We found no evidence of an effect of data reporting completeness (complete: $r$ = -0.081, [95% CI = -0.337, 0.174], [95% PI = -0.905, 0.742], k = 304, N = 45; incomplete: $r$ = -0.040, [95% CI = -0.324, 0.244], [95% PI = -0.873, 0.793], k = 34, N = 8; p-value complete *vs* incomplete = 0.553) and this moderator explained little variance ($R^2_{marginal}$ = 0.1%). However, our results should be interpreted cautiously because they do not account for the fact that 229 potentially suitable effect sizes (34.1% of 672) could not be included in our final dataset due to incomplete reporting information.

Last, the multi-moderator model that we originally planned to estimate the total amount of heterogeneity explained by complete-case moderators, only contained a test for both small-study and decline effects simultaneously due to the lack of complete-case moderators in our dataset. This multi-moderator model resulted in similar results as those of the uni-moderator model testing for publication bias (Supplementary Information S14, $R^2_{marginal}$ = 1.3%).

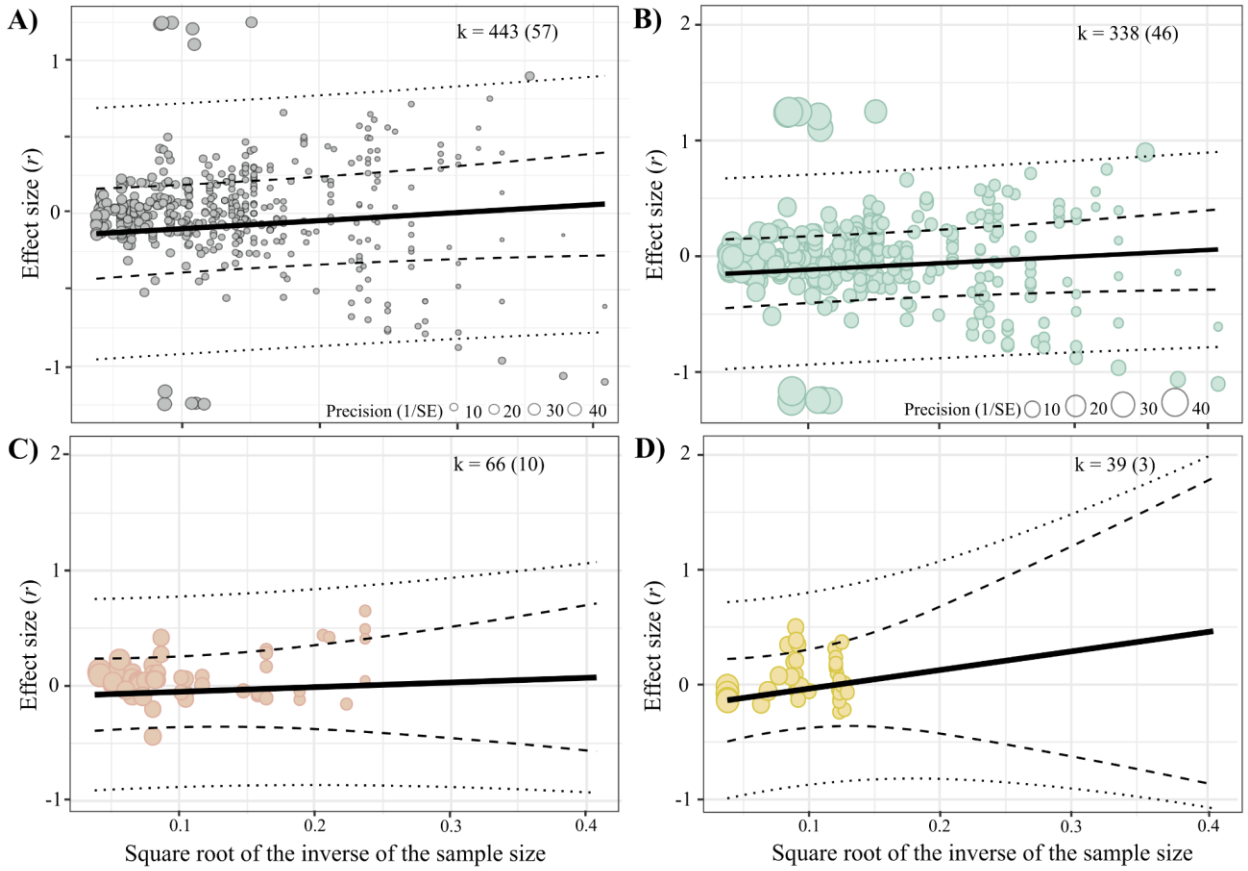


Figure 7: Orchard plots showing small-study effects biases in our dataset. A) For all studies, larger positive effect sizes were associated with effect sizes of lower prediction. For B) androgens, C) glucocorticoids, and D) thyroids no clear association was present. k indicates the number of individual effect sizes and, in between brackets, the number of studies.

## 4. Discussion

Our pre-registered systematic review and meta-analysis found that the fitness effects of egg hormones across 19 wild bird species are neither as strong nor as clear as previously suggested. The overall mean effect of androgens, glucocorticoids, and thyroid hormones on both maternal and offspring fitness was statistically indistinguishable from zero. However, those effects were also highly heterogeneous, suggesting that, once differences in precision are accounted for, the magnitude and direction of the relationship between hormone concentrations and fitness is context specific. We tested several biological and methodological moderators that we predicted could explain the differences in fitness effects for offspring, for which most of the data exists, but they only accounted for a small percentage of the heterogeneity. We also did not find any compelling evidence of publication bias either when testing for small study, decline or reporting effects. Below, we discuss our findings in detail, examining our hypotheses in specific contexts, and highlighting knowledge gaps, biases in the current literature, and areas for future research.

The influence of egg hormone concentrations on fitness is expected to vary by hormone group: an increase in egg androgens and thyroid hormones is predicted to increase fitness, while an increase in glucocorticoids is expected to decrease it (e.g., von Engelhardt & Groothuis 2011; Groothuis *et al.* 2005a; Haussmann *et al.* 2012; Hsu *et al.* 2017; Polich *et al.* 2018; Ruuskanen *et al.* 2018; Stier *et al.* 2009). However, our results showed that the relationship between hormone concentration and offspring fitness proxies was similarly small and statistically indistinguishable from zero for all three hormone groups, with 'hormone type' explaining a negligible 0.4% of the heterogeneity among effect sizes. Our exploratory analysis testing the relationship with maternal fitness for the two hormones for which sufficient data were available (i.e., androgens and glucocorticoids) also showed a small and statistically nonsignificant relationship. In all, our results robustly show no support for a consistent overall relationship between egg hormones and fitness in wild bird populations.

We found no overall evidence for a sex-specific effect of egg androgens and thyroid hormones on offspring fitness. Our result contrasts with that by Podmokła *et al.* (2018), who showed that eggs injected with androgens have a stronger fitness effect on male than female

birds. The discrepancy between studies may be due to three non-mutually exclusive factors. First, the effect size reported by Podmokła *et al.* (2018) includes offspring traits besides reproduction and survival, such as behavior and physiology. Second, their study focused primarily on captive birds, with only six wild bird populations included in their analysis. Third, evidence suggests that sex-specific sensitivity to egg androgens occurs in bird species with sexual size dimorphism (Tschirren 2015). Since our dataset mainly consists of species without such dimorphism, this may have contributed to the differences with Podmokła *et al.* (2018). However, our finding aligns with a meta-analysis by Moore et al. (2019), which showed that maternal effects in animal populations do not have a sex-specific influence on offspring. Nonetheless, it is worth noting that sex-specific fitness consequences of egg hormones in wild birds are still vastly understudied. In fact, we could only test this hypothesis on a small subset of the data (12.7% of all eligible effect sizes) and for only two hormones. Therefore, although conclusions from our test should be treated with caution, given that sex as a moderator explained a non-negligible percentage of the heterogeneity, we encourage researchers to investigate and report offspring fitness proxies separately for male and female offspring.

The relationship between egg hormones and offspring fitness proxies showed no overall change with age. Theory predicts that maternal effects should weaken as offspring encounter more environments throughout ontogeny (e.g., Mousseau & Dingle 1991), and a meta-analysis of vertebrate species supported this prediction (Moore *et al.* 2019). The absence of an ontogenetic decline in our study might be because of egg hormones being weak already at hatching, when they are expected to be strong compared to other stages. Moreover, Moore *et al.* (2019) found that maternal effects declined more for traits with higher maternal influence such as morphological traits compared to other traits such as behavior and physiology. Most studies providing offspring life-stage or age focused on androgens and, thus, we urge researchers to gather more data on glucocorticoids and thyroid hormones to better understand the relationship between hormones across life stages.

Discrepancies among studies in the magnitude, sign, and statistical significance of the effect of maternal egg hormones on fitness proxies have long been attributed to methodological differences across studies (e.g., Groothuis *et al.* 2019; Podmokła *et al.* 2018), which we explored in detail but found no evidence for. Specifically, we found that egg androgens and glucocorticoids had no overall effect on fitness proxies regardless of study type (correlational or experimental) or experimental focus (mother or egg). The experimental dose being within the natural range or above physiological concentrations, the site (or matrix) in which hormones were measured or manipulated (albumen or yolk), and the sampling technique used (egg biopsy or entire egg

collection) explained little to negligible amounts of the observed heterogeneity. Hence, in wild bird populations, the overall null effect between hormones and fitness proxies appears consistent, even when accounting for several methodological differences across studies. Yet, due to lack of reporting in the original studies, many of these tests could only be performed for small subsets of data, so their results should be interpreted cautiously.

By using a pluralistic approach to estimate total heterogeneity and stratify it across the different hierarchical levels explored, we found that the overall association between egg hormones and both maternal and offspring fitness was highly heterogeneous and stemming from within-study variation and among-species differences (Supplementary Information S9.2 & S10). The high levels of heterogeneity associated with within- compared to among-study variation, which was zero in all models, suggest that the small and statistically nonsignificant overall effect found would be generalizable among studies as long as (i) we define replication as the testing of the null hypothesis at the among-study level and, (ii) importantly, if within-study level variation - both biological and methodological, are accounted for (Yang *et al.* 2024). However, our moderator analyses testing for both biological and methodological differences did not explain a substantial amount of heterogeneity, limiting our potential for generalization. We hypothesized that such differences may stem from researchers reporting multiple fitness proxies within studies (mean ± SD = 3.0 ± 1.9), and therefore conducted an additional (non-pre-registered) meta-regression using the entire dataset. Although none of the fitness proxy-specific mean estimates were statistically significantly different from zero (see Supplementary Figure 5; Supplementary Information S15), the fitness proxy used accounts for a non-negligible percentage of the heterogeneity ($R^2_{marginal}$ = 3.2%; i.e., the percentage of heterogeneity explained corresponded to our second highest estimate). Among-species differences, and particularly, phylogenetic relationships among the 19 bird species were associated with high levels of heterogeneity, which indicates that the size and magnitude of the overall association between egg hormones and fitness proxies differ among taxonomic groups (Figure 1C). We explored whether species-specific characteristics (e.g., life-history traits, breeding traits, behaviors) that are strongly correlated with phylogeny may explain part of that heterogeneity. For this, we ran a (non-pre-registered) model to compare precocial vs altricial species and found that developmental mode explains little heterogeneity ($R^2_{marginal}$ = 0.3%; see Supplementary Information S16). To better understand the factors driving these differences, future research could focus on understanding if species-specific characteristics (e.g., life-history, ecology) might be driving the phylogenetic signal observed in our study as well as investigate whether egg hormones and fitness are associated in other bird species not yet covered.

Three additional factors might explain the heterogeneity that we found across effect sizes. First, theory predicts that anticipatory maternal effects evolve when mothers and offspring encounter similar environments, or when environmental fluctuations are low (Hoyle & Ezard 2012; Kuijper & Hoyle 2015). One context not accounted for in this meta-analysis, due to its complexity, is the shared environment of mothers and offspring. Egg hormones may significantly influence fitness proxies only when maternal and offspring environments match, and/or when mothers can predict the environment their offspring will encounter. Second, maternal effects are multivariate. Mothers influence their offspring's prenatal environment by depositing various compounds into the egg besides hormones (e.g., immunoglobulins, antioxidants, fatty acids, proteins; (Lovern & Wade 2001; McCormick 1999; Mentesana *et al.* 2019; Valcu *et al.* 2019; Weiss *et al.* 2011), which may interact (Groothuis *et al.* 2019; Williams & Groothuis 2015). If these egg components do not act independently on fitness traits, and their correlations are weak or different combinations have similar outcomes, testing the effects of single components independently, as we did, is likely to result in high heterogeneity. Third, there is growing evidence that embryos are not passive responders to maternal signals (Bebbington & Groothuis 2021; Groothuis *et al.* 2019). Embryos seem adapted to receiving maternal hormone signals, as they have hormone receptors before producing their own hormones (Godsave *et al.* 2002; Kumar *et al.* 2019). However, recent evidence suggests that embryos can rapidly metabolize maternal hormones in the egg into potentially inactive steroid forms (Kumar *et al.* 2018; Vassallo *et al.* 2014). If the hormone concentrations measured or manipulated in the egg do not reflect the amount actually incorporated by the offspring, this could explain the variation we found across effect sizes.

## 5. Conclusions

Thirty years after the discovery of egg hormones, our meta-analysis showed an increase in their concentrations to have a near zero and highly variable overall mean effect on fitness proxies in wild bird populations. We could not identify any predictor explaining much of the high levels of heterogeneity found within studies and among species, even after accounting for various biological (e.g., hormone type, offspring sex and age) and methodological (e.g., study type, experimental subject, dose) factors.

Our meta-analysis raises doubts about the premise that egg hormone-mediated maternal effects are an adaptive mechanism through which mothers adjust hormone concentrations to shape offspring phenotypes and optimize maternal and/or offspring fitness. Moreover, by summarizing all existing literature to date, our study reveals that the theory supporting the

evolutionary relevance of maternal egg hormones is based on a biased dataset. Among all egg-laying vertebrates, only birds have been studied in the wild to test for a relationship between egg hormones and fitness. Furthermore, 60% of the data comes from just three bird species (*Sturnus unicolor*, *Parus major*, and *Ficedula albicollis*), primarily from Northern Hemisphere populations. Most studies (76%) focus on androgen hormones, and the dataset covers a narrow range of fitness traits. We hope our work will encourage researchers to fill these gaps and address some of the many questions that remain unanswered in this fascinating area of research. Only by doing so can we truly understand whether maternal egg hormone concentrations are adaptive or not.

## Acknowledgements

We thank all the authors who replied to our email, addressed our questions, and kindly shared their data: Alexandra Bentz, Amber Rice, Barbara Tschirren, Bibiana Montoya, Bin-Yan Hsu, E. Diego Rubolini, Keith Bowers, Gergely Hegyi, Ismael Galvan, Jaime Muriel, János Török, Jose Carlos Noguera, Keith W. Sockman, Kristen J. Navara, Marco Parolini, Scott Sakaluk, and Suvi Ruuskanen. We also want to thank Yefeng Yang for their feedback on interpreting heterogeneity results. This research was funded by Max Planck Gesellschaft to MH.

## Author's contributions

Conceptualization: LM, MH and AST. Data curation: LM, PBD, NMA and AST. Formal Analysis: LM and AST. Funding acquisition: MH. Investigation: LM, PBD, and NMA. Methodology: LM and AST. Project administration: LM. Software: LM and AST. Supervision: LM and AST. Validation: LM, PBD, NMA, and AST. Visualization: LM and NMA. Writing – original draft: LM. Writing – review & editing: all authors.

## References

Amrhein, V., Greenland, S. & Blake McShane. (2019). Scientists rise up against statistical significance." Nature 567.7748 (2019): 305-307. *Nature*, 567, 305–307.

Bebbington, K. & Groothuis, T.G.G. (2021). Who listens to mother? A whole-family perspective on the evolution of maternal hormone allocation. *Biol. Rev*., 4.

Bentz, A.B., Navara, K.J. & Siefferman, L. (2013). Phenotypic plasticity in response to breeding density in tree swallows: An adaptive maternal effect? *Horm. Behav*., 64, 729–736.

Bentz, A.B., Sirman, A.E., Wada, H., Navara, K.J. & Hood, W.R. (2016). Relationship between maternal environment and DNA methylation patterns of estrogen receptor alpha in wild Eastern Bluebird (*Sialia sialis*) nestlings: a pilot study. *Ecol. Ev.*

Borenstein, M., Hedges, L. V. & Higgins, J. P. T., Rothstein, H.R. (2009). Introduction to meta-analysis.

Bowers, E.K., Bowden, R.M., Sakaluk, S.K. & Thompson, C.F. (2015). Immune activation generates corticosterone-mediated terminal reproductive investment in a wild bird. *Am. Nat.*, 185, 769–783.

Bowers, E.K., Bowden, R.M., Thompson, C.F. & Sakaluk, S.K. (2016). Elevated corticosterone during egg production elicits increased maternal investment and promotes nestling growth in a wild songbird. *Horm. Behav*., 83, 6–13.

Clairardin, S.G., Barnett, C.A., Sakaluk, S.K. & Thompson, C.F. (2011). Experimentally increased in ovotestosterone leads to increased plasma bactericidal activity and decreased cutaneous immune response in nestling house wrens. *J. Exp. Biol*., 214, 2778–2782.

Clotfelter, E.D., O'Neal, D.M., Gaudioso, J.M., Casto, J.M., Parker-Renga, I.M., Snajdr, E.A., et al. (2004). Consequences of elevating plasma testosterone in females of a socially monogamous songbird: Evidence of constraints on male evolution? *Horm. Behav*., 46, 171–178.

Duckworth, R.A., Belloni, V. & Anderson, S.R. (2015). Cycles of species replacement emerge from locally induced maternal effects on offspring behavior in a passerine bird. *Science* (80)., 347, 875–877.

Eising, C.M. & Groothuis, T.G.G. (2003). Yolk androgens and begging behaviour in black-headed gull chicks: an experimental field study. *Anim. Behav*., 66, 1027–1034.

von Engelhardt, N. & Groothuis, T.G.G. (2011). Maternal Hormones in Avian Eggs. *Horm. Reprod. Vertebr.* - Vol. 4. Elsevier Inc.

Freeberg, T.M., Benson, S.A. & Burghardt, G.M. (2024). Minimizing observer bias in animal behavior studies revisited: Improvement, but a long way to go. *Ethology*, 130, e13446.

Galván, I. & Alonso-Alvarez, C. (2010). Yolk testosterone shapes the expression of a melanin-based signal in great tits: An antioxidant-mediated mechanism*? J. Exp. Biol*., 213, 3127–3130.

Gil, D. (2008). Chapter 7 Hormones in Avian Eggs: Physiology, Ecology and Behavior. *Advances in the Study of Behavior*.

Gil, D., Ninni, P., Lacroix, A., De Lope, F., Tirard, C., Marzal, A., et al. (2006). Yolk androgens in the barn swallow (Hirundo rustica): A test of some adaptive hypotheses. *J. Evol. Biol.*, 19, 123–131.

Godsave, S.F., Lohmann, R., Vloet, R.P.M. & Gahr, M. (2002). Androgen receptors in the embryonic zebra finch hindbrain suggest a function for maternal androgens in perihatching survival. *J. Comp. Neurol.,* 453, 57–70.

Grafen, A. (1989). The phylogenetic regression. *Philos. Trans. R. Soc. London. B, Biol. Sci.*, 326, 119–157.

Groothuis, T.G.G. & Schwabl, H. (2008). Hormone-mediated maternal effects in birds: mechanisms matter but what do we know of them? *Philos. Trans. R. Soc. London B Biol. Sci*., 363.

Groothuis, T.G., Kumar, N. & Hsu, B.-Y. (2020). Explaining discrepancies in the study of maternal effects: the role of context and embryo. *Curr. Opin. Behav. Sci.*, 36, 185–192.

Groothuis, T.G.G., Müller, W., von Engelhardt, N., Carere, C. & Eising, C. (2005a). Maternal hormones as a tool to adjust offspring phenotype in avian species. *Neurosci. Biobehav. Rev.*, 29, 329–352.

Groothuis, T.G.G., Eising, C.M., Dijkstra, C. & Müller, W. (2005b). Balancing between costs and benefits of maternal hormone deposition in avian eggs. *Biol. Lett.*, 1, 78–81.

Groothuis, T.G.G. & Von Engelhardt, N. (2005). Investigating maternal hormones in avian eggs: Measurement, manipulation, and interpretation. *Ann. N. Y. Acad. Sci.*, 1046, 168–180.

Groothuis, T.G.G., Hsu, B.-Y., Kumar, N. & Tschirren, B. (2019). Revisiting mechanisms and functions of prenatal hormone-mediated maternal effects using avian species as a model. *Philos. Trans. R. Soc. B Biol. Sci.*, 374, 20180115.

Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555, 175–182.

Haussmann, M.F., Longenecker, A.S., Marchetto, N.M., Juliano, S.A. & Bowden, R.M. (2012). Embryonic exposure to corticosterone modifies the juvenile stress response, oxidative stress and telomere length. *Proc. R. Soc. London B Biol. Sci.*, 279.

Hegyi, G., Herényi, M., Szöllosi, E., Rosivall, B., Török, J. & Groothuis, T.G.G. (2011). Yolk androstenedione, but not testosterone, predicts offspring fate and reflects parental quality. *Behav. Ecol.*, 22, 29–38.

Hennessy, E.A., Acabchuk, R.L., Arnold, P.A., Dunn, A.G., Foo, Y.Z., Johnson, B.T., et al. (2022). Ensuring Prevention Science Research is Synthesis-Ready for Immediate and Lasting Scientific Impact. *Prev. Sci.*, 23, 809–820.

Hernández, A., Hernández-Martínez, A., Nicolás, L., Torres, R. & Montoya, B. (2022). Competitive events during nest-building influence offspring growth and behaviour, but not testosterone deposition in the egg yolk in Western Bluebirds (*Sialia mexicana*). *Ibis*, 164, 1049–1062.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., et al. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci.*, 112, 12764–12769.

Holman, L., Head, M.L., Lanfear, R. & Jennions, M.D. (2015). Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording. *PLOS Biol.*, 13, 1–12.

Hoyle, R.B. & Ezard, T.H.G. (2012). The benefits of maternal effects in novel and in stable environments. *J. R. Soc. Interface*, 9, 2403–2413.

Hsu, B.-Y., Dijkstra, C. & Groothuis, T.G.G. (2017). Organizing effects of adverse early-life condition on body mass, compensatory growth and reproduction: experimental studies in rock pigeons. *J. Avian Biol.*, 48, 1166–1176.

Hsu, B.-Y., Doligez, B., Gustafsson, L. & Ruuskanen, S. (2019). Transient growth-enhancing effects of elevated maternal thyroid hormones at no apparent oxidative cost during early postnatal period. *J. Avian Biol.*, 50.

Hsu, B.-Y., Sarraude, T., Cossin-Sevrin, N., Crombecque, M., Stier, A. & Ruuskanen, S. (2020). Testing for context-dependent effects of prenatal thyroid hormones on offspring survival and physiology: an experimental temperature manipulation. *Sci. Rep.*, 10.

Ihle, M., Winney, I.S., Krystalli, A. & Croucher, M. (2017). Striving for transparent and credible research: practical guidelines for behavioral ecologists. *Behav. Ecol.*, 28, 348–354.

IntHout, J., Ioannidis, J.P.A., Rovers, M.M. & Goeman, J.J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6, e010247.

Jacobs, P. & Viechtbauer, W. (2017). Estimation of the biserial correlation and its sampling variance for use in meta-analysis. *Res. Synth. Methods*, 8, 161–180.

Keaney, T.A., Jones, T.M. & Mulder, R.A. (2024). An undergraduate classroom experiment illustrates an effect of observer bias on data collection in animal behaviour. *Anim. Behav.*, 212, 175–182.

Kuijper, B. & Hoyle, R.B. (2015). When to rely on maternal effects and when on phenotypic plasticity? *Evolution*, 69, 950–968.

Kumar, N., van Faassen, M., Kema, I., Gahr, M. & Groothuis, T.G.G. (2018). Early embryonic modification of maternal hormones differs systematically among embryos of different laying order: A study in birds. *Gen. Comp. Endocrinol.*, 269, 53–59.

Kumar, N., Lohrentz, A., Gahr, M. & Groothuis, T.G.G. (2019). Steroid receptors and their regulation in avian extraembryonic membranes provide a novel substrate for hormone mediated maternal effects. *Sci. Rep.*, 9, 11501.

Lajeunesse, M.J. (2013). 13. Recovering Missing or Partial Data from Studies: A Survey of Conversions and Imputations for Meta-analysis. In: Handbook of Meta-analysis in Ecology and Evolution (eds. Koricheva, J., Gurevitch, J. & Mengersen, K.). Princeton University Press, Princeton, pp. 195–206.

López-Rull, I. & Gil, D. (2009). Elevated testosterone levels affect female breeding success and yolk androgen deposition in a passerine bird. *Behav. Processes*, 82, 312–318.

Love, O.P., Chin, E.H., Wynne-Edwards, K.E. & Williams, T.D. (2005). Stress hormones: a link between maternal condition and sex-biased reproductive investment. *Am. Nat.*, 166, 751–66.

Lovern, M.B. & Wade, J. (2001). Maternal plasma and egg yolk testosterone concentrations during embryonic development in green anoles (*Anolis carolinensis*). *Gen. Comp. Endocrinol.*, 124, 226–235.

Marshall, D.J. & Uller, T. (2007). When is a maternal effect adaptive? *Oikos*, 116, 1957–1963.

McCormick, M.I. (1999). Experimental test of the effect of maternal hormones on larval quality of a coral reef fish. *Oecologia*, 118, 412–422.

Mentesana, L., Andersson, M.N., Casagrande, S., Goymann, W., Isaksson, C. & Hau, M. (2021a). Natural variation in yolk fatty acids, but not androgens, predicts offspring fitness in a wild bird. *Front. Zool.*, 18, 1–12.

Mentesana, L., Hau, M., Adreani, M.N., D'amelio, P. & Sánchez-Tójar, A. (2021b). Fitness consequences of egg hormones in the light of evidence synthesis. Retrieved from osf.io/st38j. OSF.

Mentesana, L., Isaksson, C., Goymann, W., Andersson, M.N., Trappschuh, M. & Hau, M. (2019). Female variation in allocation of steroid hormones, antioxidants and fatty acids: a multilevel analysis in a wild passerine bird. *J. Avian Biol*., 50, 1–16.

Michonneau, F., Brown, J.W. & Winter, D.J. (2016). rotl: an R package to interact with the Open Tree of Life data. *Methods Ecol. Evol*., 7, 1476–1481.

Moore, M.P., Whiteman, H.H. & Martin, R.A. (2019). A mother's legacy: the strength of maternal effects in animal populations. *Ecol. Lett*., 22, 1620–1628.

Mousseau, T.A. & Dingle, H. (1991). Maternal effects in insect life histories. *Annu. Rev. Entomol*. Vol. 36, 511–534.

Mousseau, T.A. & Fox, C.W. (1998). The adaptive significance of maternal effects. *Trends Ecol. Evol.*

Müller, W., Deptuch, K., López-Rull, I. & Gil, D. (2007). Elevated yolk androgen levels benefit offspring development in a between-clutch context*. Behav. Ecol*., 18, 929–936.

Müller, W., Dijkstra, C. & Groothuis, T.G.G. (2009). Maternal yolk androgens stimulate territorial behaviour in black-headed gull chicks. *Biol. Lett.*, 5, 586–588.

Müller, W., Groothuis, T.G.G., Kasprzik, A., Dijkstra, C., Alatalo, R. V & Siitari, H. (2005). Prenatal androgen exposure modulates cellular and humoral immune function of black-headed gull chicks. *Proc. R. Soc. B Biol. Sci*., 272, 1971–1977.

Muriel, J., Pérez-Rodríguez, L., Puerta, M. & Gil, D. (2013). Differential effects of yolk testosterone and androstenedione in embryo development and nestling growth in the spotless starling (*Sturnus unicolor*). *Gen. Comp. Endocrinol*., 194, 175–182.

Muriel, J., Pérez-Rodríguez, L., Puerta, M. & Gil, D. (2015a). Diverse dose-response effects of yolk androgens on embryo development and nestling growth in a wild passerine. *J. Exp. Biol.*, 218, 2241–9.

Muriel, J., Salmón, P., Nunez-Buiza, A., de Salas, F., Pérez-Rodríguez, L., Puerta, M., et al. (2015b). Context-dependent effects of yolk androgens on nestling growth and immune function in a multibrooded passerine. *J. Evol. Biol*., 28, 1476–1488.

Nagy, G., Blázi, G., Hegyi, G. & Török, J. (2016). Side-specific effect of yolk testosterone elevation on second-to-fourth digit ratio in a wild passerine. *Sci. Nat*., 103.

Nakagawa, S. & Cuthill, I.C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists*. Biol. Rev*., 82, 591–605.

Nakagawa, S., Lagisz, M., Jennions, M.D., Koricheva, J., Noble, D.W.A., Parker, T.H., et al. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods Ecol. Evol.*, 13, 4–21.

Nakagawa, S., Lagisz, M., O'Dea, R.E., Pottier, P., Rutkowska, J., Senior, A.M., et al. (2023). orchaRd 2.0: An R package for visualising meta-analyses with orchard plots. Methods *Ecol. Evol.*, 14, 2003–2010.

Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining R 2 from generalized linear mixed-effects models. *Methods Ecol. Evol.*, 4, 133–142.

Navara, K.J., Hill, G.E. & Mendonça, M.T. (2005). Variable effects of yolk androgens on growth, survival, and immunity in eastern bluebird nestlings. *Physiol. Biochem. Zool.*, 78, 570–578.

Navara, K.J., Hill, G.E. & Mendonça, M.T. (2006). Yolk testosterone stimulates growth and immunity in house finch chicks. *Physiol. Biochem. Zool*., 79, 550–555.

Noble, D.W.A., Lagisz, M., O'dea, R.E. & Nakagawa, S. (2017). Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. *Mol. Ecol.*, 26, 2410–2425.

Noguera, J.C., Kim, S.-Y. & Velando, A. (2013). Maternal testosterone influences a begging component that makes fathers work harder in chick provisioning. *Horm. Behav.*, 64, 19–25.

Noguera, J.C., da Silva, A. & Velando, A. (2020). Egg corticosterone can stimulate telomerase activity and promote longer telomeres during embryo development. *Mol. Ecol.*

O'Dea, R.E., Lagisz, M., Jennions, M.D., Koricheva, J., Noble, D.W.A., Parker, T.H., et al. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biol. Rev*., 96, 1695–1722.

Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Syst. Rev*., 5, 1–10.

Paradis, E. & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.

Parolini, M., Possenti, C.D., Romano, A., Caprioli, M., Rubolini, D. & Saino, N. (2018). Physiological increase of yolk testosterone level does not affect oxidative status and telomere length in gull hatchlings. *PLoS One*, 13.

Partecke, J., Hegyi, G., Fitze, P.S., Gasparini, J. & Schwabl, H. (2020). Maternal effects and urbanization: Variation of yolk androgens and immunoglobulin in city and forest blackbirds. *Ecol. Evol.*

Pick, J.L., Nakagawa, S. & Noble, D.W.A. (2019). Reproducible, flexible and high-throughput data extraction from primary literature: The metaDigitise r package. *Methods Ecol. Evol*., 10, 426–431.

Pilz, K.M., Quiroga, M., Schwabl, H. & Adkins-Regan, E. (2004). European starling chicks benefit from high yolk testosterone levels during a drought year. *Horm. Behav*., 46, 179–192.

Pitala, N., Ruuskanen, S., Laaksonen, T., Doligez, B., Tschirren, B. & Gustafsson, L. (2009). The effects of experimentally manipulated yolk androgens on growth and immune function of male and female nestling collared flycatchers *Ficedula albicollis*. *J. Avian Biol*., 40, 225–230.

Podlas, K., Helfenstein, F. & Richner, H. (2013). Brood Reduction via Intra-clutch Variation in Testosterone - An Experimental Test in the Great Tit. *PLoS One*, 8.

Podmokła, E., Drobniak, S.M. & Rutkowska, J. (2018a). Chicken or egg? Outcomes of experimental manipulations of maternally transmitted hormones depend on administration method - a meta-analysis. *Biol. Rev.*, 93, 1499–1517.

Podmokła, E., Drobniak, S.M. & Rutkowska, J. (2018b). Chicken or egg? Outcomes of experimental manipulations of maternally transmitted hormones depend on administration method – a meta-analysis. *Biol. Rev*., 93, 1499–1517.

Poisbleau, M., Müller, W., Carslake, D., Demongin, L., Groothuis, T.G.G., van Camp, J., et al. (2012). Penguin chicks benefit from elevated yolk androgen levels under sibling competition. *PLoS One*, 7.

Polich, R.L., Bodensteiner, B.L., Adams, C.I.M. & Janzen, F.J. (2018). Effects of augmented corticosterone in painted turtle eggs on offspring development and behavior. *Physiol. Behav.*, 183, 1–9.

Possenti, C.D., Romano, A., Caprioli, M., Rubolini, D., Spiezio, C., Saino, N., et al. (2016). Yolk testosterone affects growth and promotes individual-level consistency in behavioral lateralization of yellow-legged gull chicks. *Horm. Behav*., 80, 58–67.

Rees, J.A. & Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodivers. data* J., e12581.

Reinhold, K. (2002). Maternal Effects and the Evolution of Behavioral and Morphological Characters: A Literature Review Indicates the Importance of Extended Maternal Care. *J. Hered*., 93, 400–405.

Rice, A.M., Vallin, N., Kulma, K., Arntsen, H., Husby, A., Tobler, M., et al. (2013). Optimizing the trade-off between offspring number and quality in unpredictable environments: Testing the role of differential androgen transfer to collared flycatcher eggs. *Horm. Behav*., 63, 813–822.

Ronget, V., Gaillard, J.-M., Coulson, T., Garratt, M., Gueyffier, F., Lega, J.-C., et al. (2018). Causes and consequences of variation in offspring body mass: meta-analyses in birds and mammals. *Biol. Rev*., 93, 1–27.

Rubolini, D., Romano, M., Boncoraglio, G., Ferrari, R.P., Martinelli, R., Galeotti, P., et al. (2005). Effects of elevated egg corticosterone levels on behavior, growth, and immunity of yellow-legged gull (*Larus michahellis*) chicks. *Horm. Behav*., 47, 592–605.

Rubolini, D., Romano, M., Martinelli, R. & Saino, N. (2006). Effects of elevated yolk testosterone levels on survival, growth and immunity of male and female yellow-legged gull chicks. *Behav. Ecol. Sociobiol*., 59, 344–352.

Ruuskanen, S., Darras, V.M., Visser, M.E. & Groothuis, T.G.G. (2016). Effects of experimentally manipulated yolk thyroid hormone levels on offspring development in a wild bird species. *Horm. Behav.*, 81, 38–44.

Ruuskanen, S., Doligez, B., Gustafsson, L. & Laaksonen, T. (2012a). Long-term effects of yolk androgens on phenotype and parental feeding behavior in a wild passerine. *Behav. Ecol. Sociobiol*., 66, 1201–1211.

Ruuskanen, S., Doligez, B., Pitala, N., Gustafsson, L. & Laaksonen, T. (2012b). Long-term fitness consequences of high yolk androgen levels: Sons pay the costs. *Funct. Ecol*., 26, 884–894.

Ruuskanen, S., Doligez, B., Tschirren, B., Pitala, N., Gustafsson, L., Groothuis, T.G.G., et al. (2009). Yolk androgens do not appear to mediate sexual conflict over parental investment in the collared flycatcher *Ficedula albicollis*. *Horm. Behav*., 55, 514–519.

Ruuskanen, S., Groothuis, T.G.G., Baugh, A.T., Schaper, S. V, de Vries, B. & van Oers, K. (2018). Maternal egg hormones in the mating context: The effect of pair personality. *Funct. Ecol.*, 32, 439–449.

Ruuskanen, S. & Laaksonen, T. (2010). Yolk hormones have sex-specific long-term effects on behavior in the pied flycatcher (*Ficedula hypoleuca*). *Horm. Behav*., 57, 119–127.

Ruuskanen, S. & Laaksonen, T. (2013). Sex-specific effects of yolk androgens on begging behavior and digestion in pied flycatchers. *J. Avian Biol*., 44, 331–338.

Sánchez-Tójar, A., Lagisz, M., Moran, N.P., Nakagawa, S., Noble, D.W.A. & Reinhold, K. (2020). The jury is still out regarding the generality of adaptive 'transgenerational' effects. *Ecol. Lett.*, 23, 1715–1718.

Sánchez-Tójar, A., Nakagawa, S., Sánchez-Fortún, M., Martin, D.A., Ramani, S., Girndt, A., et al. (2018). Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias. *Elife*, 7, e37385.

Schwabl, H. (1993). Yolk is a source of maternal testosterone for developing birds. *Proc. Natl. Acad. Sci.,* 90, 11446–11450.

Schwabl, H. & Partecke, J. (2020). Developmental programming of the adrenocortical stress response by yolk testosterone depends on sex and life history stage. *J. Exp. Biol*., 223.

Sockman, K.W. & Schwabl, H. (2000). Yolk androgens reduce offspring survival. *Proc. R. Soc. B Biol. Sci*., 267, 1451–1456.

Sockman, K.W., Weiss, J., Webster, M.S., Talbott, V. & Schwabl, H. (2008). Sex-specific effects of yolk-androgens on growth of nestling American kestrels. *Behav. Ecol. Sociobiol*., 62, 617–625.

Stier, K.S., Almasi, B., Gasparini, J., Piault, R., Roulin, A. & Jenni, L. (2009). Effects of corticosterone on innate and humoral immune functions and oxidative stress in barn owl nestlings. *J. Exp. Biol.*, 212, 2085–91.

Strange, M.S., Bowden, R.M., Thompson, C.F. & Sakaluk, S.K. (2016). Pre- and Postnatal Effects of Corticosterone on Fitness-Related Traits and the Timing of Endogenous Corticosterone Production in a Songbird. *J. Exp. Zool. Part A Ecol. Genet. Physiol*., 325, 347–359.

Strasser, R. & Schwabl, H. (2004). Yolk testosterone organizes behavior and male plumage coloration in house sparrows (*Passer domesticus). Behav. Ecol. Sociobiol*., 56, 491–497.

Tilgar, V., Mägi, M., Lind, M., Lodjak, J., Moks, K. & Mänd, R. (2016). Acute embryonic exposure to corticosterone alters physiology, behaviour and growth in nestlings of a wild passerine. *Horm. Behav*., 84, 111–120.

Treidel, L.A., Whitley, B.N., Benowitz-Fredericks, Z.M. & Haussmann, M.F. (2013). Prenatal exposure to testosterone impairs oxidative damage repair efficiency in the domestic chicken (*Gallus gallus*). *Biol. Lett*., 9.

Trikalinos, T.A. & Ioannidis, J.P.A. (2005). Assessing the Evolution of Effect Sizes over Time. In: Publication Bias in Meta-Analysis. John Wiley & Sons, Ltd, pp. 241–259.

Tschirren, B. (2015). Differential effects of maternal yolk androgens on male and female offspring: A role for sex-specific selection? *PLoS One*, 10.

Tschirren, B., Fitze, P.S. & Richner, H. (2007). Maternal modulation of natal dispersal in a passerine bird: An adaptive strategy to cope with parasitism? *Am. Nat*., 169, 87–93.

Tschirren, B., Postma, E., Gustafsson, L., Groothuis, T.G.G. & Doligez, B. (2014a). Natural selection acts in opposite ways on correlated hormonal mediators of prenatal maternal effects in a wild bird population. *Ecol. Lett*., 17, 1310–5.

Tschirren, B., Postma, E., Gustafsson, L., Groothuis, T.G.G. & Doligez, B. (2014b). Natural selection acts in opposite ways on correlated hormonal mediators of prenatal maternal effects in a wild bird population. *Ecol. Lett.,* 17, 1310–1315.

Tschirren, B., Saladin, V., Fitze, P.S., Schwabl, H. & Richner, H. (2005). Maternal yolk testosterone does not modulate parasite susceptibility or immune function in great tit nestlings. *J. Anim. Ecol*., 74, 675–682.

Uller, T. (2008). Developmental plasticity and the evolution of parental effects. *Trends Ecol. Evol*., 23, 432–438.

Uller, T. (2012). The Evolution of Parental Care. In: The Evolution of Parental Care (ed. Royle NJ, Smiseth PT, K.M.). New York: Oxford University Press, pp. 247–267.

Valcu, C.M., Scheltema, R.A., Schweiggert, R.M., Valcu, M., Teltscher, K., Walther, D.M., et al. (2019). Life history shapes variation in egg composition in the blue tit *Cyanistes caeruleus*. *Commun. Biol.*, 2, 1–14.

Vassallo, B.G., Paitz, R.T., Fasanello, V.J. & Haussmann, M.F. (2014). Glucocorticoid metabolism in the in ovo environment modulates exposure to maternal corticosterone in Japanese quail embryos (*Coturnix japonica*). *Biol. Lett*., 10.

Verboven, N., Monaghan, P., Evans, D.M., Schwabl, H., Evans, N., Whitelaw, C., et al. (2003). Maternal condition, yolk androgens and offspring performance: A supplemental feeding

experiment in the lesser black-backed gull (*Larus fuscus*). *Proc. R. Soc. B Biol. Sci.,* 270, 2223–2232.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor. *J. Stat. Softw*., 36, 1–48.

Weber, B.M., Bowers, E.K., Terrell, K.A., Falcone, J.F., Thompson, C.F. & Sakaluk, S.K. (2018). Pre- and postnatal effects of experimentally manipulated maternal corticosterone on growth, stress reactivity and survival of nestling house wrens. *Funct. Ecol*., 32, 1995–2007.

Weiss, S.L., Kennedy, E.A., Safran, R.J. & Mcgraw, K.J. (2011). Pterin-based ornamental coloration predicts yolk antioxidant levels in female striped plateau lizards (*Sceloporus virgatus*). *J. Anim. Ecol.*, 80, 519–527.

Westgate, M.J. (2018). revtools: Tools to support evidence synthesis (0.3.0) [Computer software]., https://CRAN.R-project.org/package=revtools.

Williams, T.D. & Groothuis, T. (2015). Egg quality, embryonic development, and post-hatching phenotype: an integrated perspective. Nests, eggs, and incubation: new ideas about avian reproduction, 113-126.

Wolf, J.B., Leamy, L.J., Roseman, C.C. & Cheverud, J.M. (2011). Disentangling prenatal and postnatal maternal genetic effects reveals persistent prenatal effects on offspring growth in mice. *Genetics*, 189, 1069–1082.

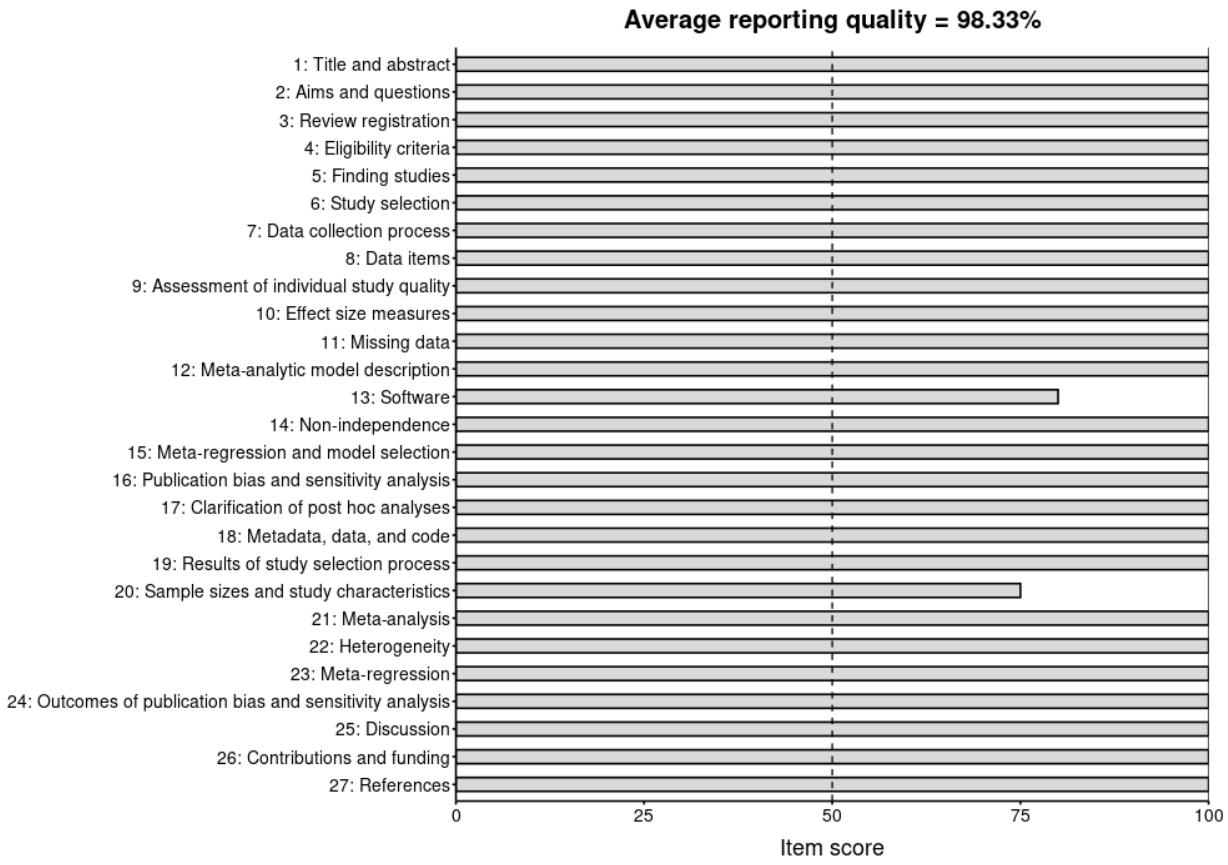
Wolf, J.B. & Wade, M.J. (2001). On the assignment of fitness to parents and offspring: whose fitness is it and when does it matter? *J. Evol. Biol.*, 14, 347–356.

Wolf, J.B. & Wade, M.J. (2016). Evolutionary genetics of maternal effects. *Evolution*, 70, 827–839.

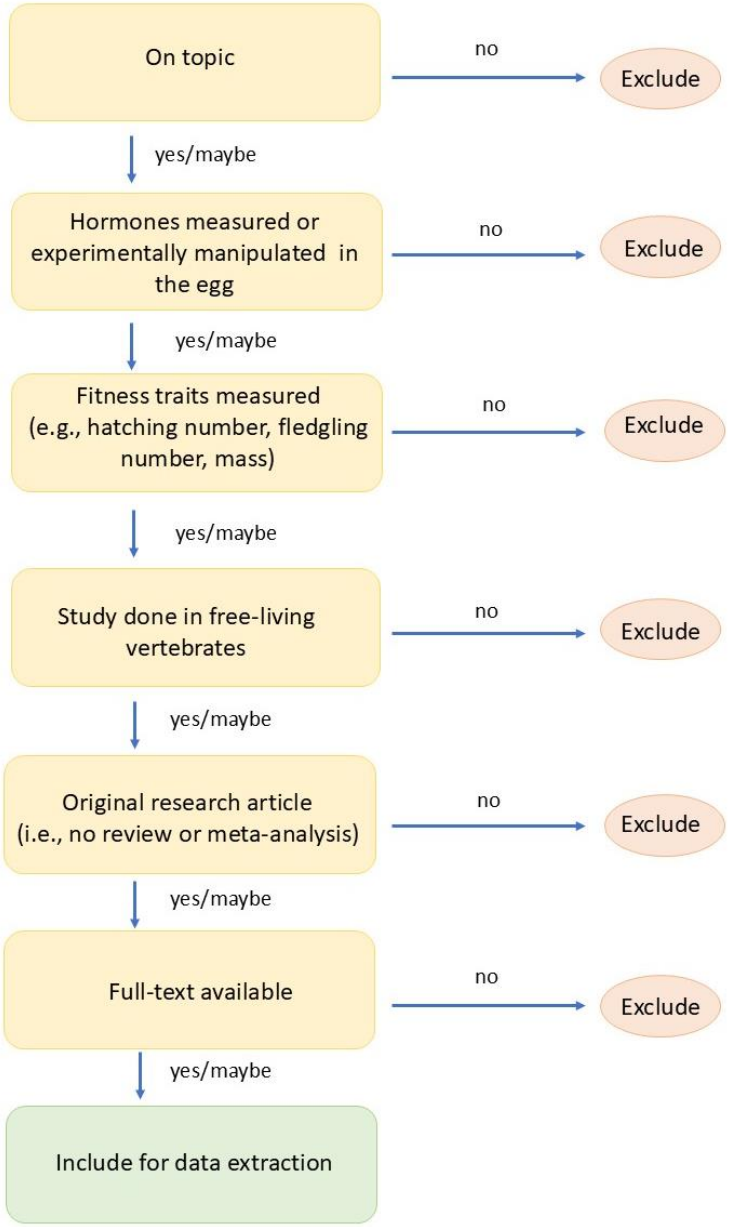
Yang, Y., Noble, D.W.A., Senior, A.M., Lagisz, M. & Nakagawa, S. (2024). Interpreting prediction intervals and distributions for decoding biological generality in meta-analyses. *bioRxiv*.

Yin, J., Zhou, M., Lin, Z., Li, Q.Q. & Zhang, Y. (2019). Transgenerational effects benefit offspring across diverse environments: a meta-analysis in plants and animals. *Ecol. Lett.*, 22, 1976–1986.

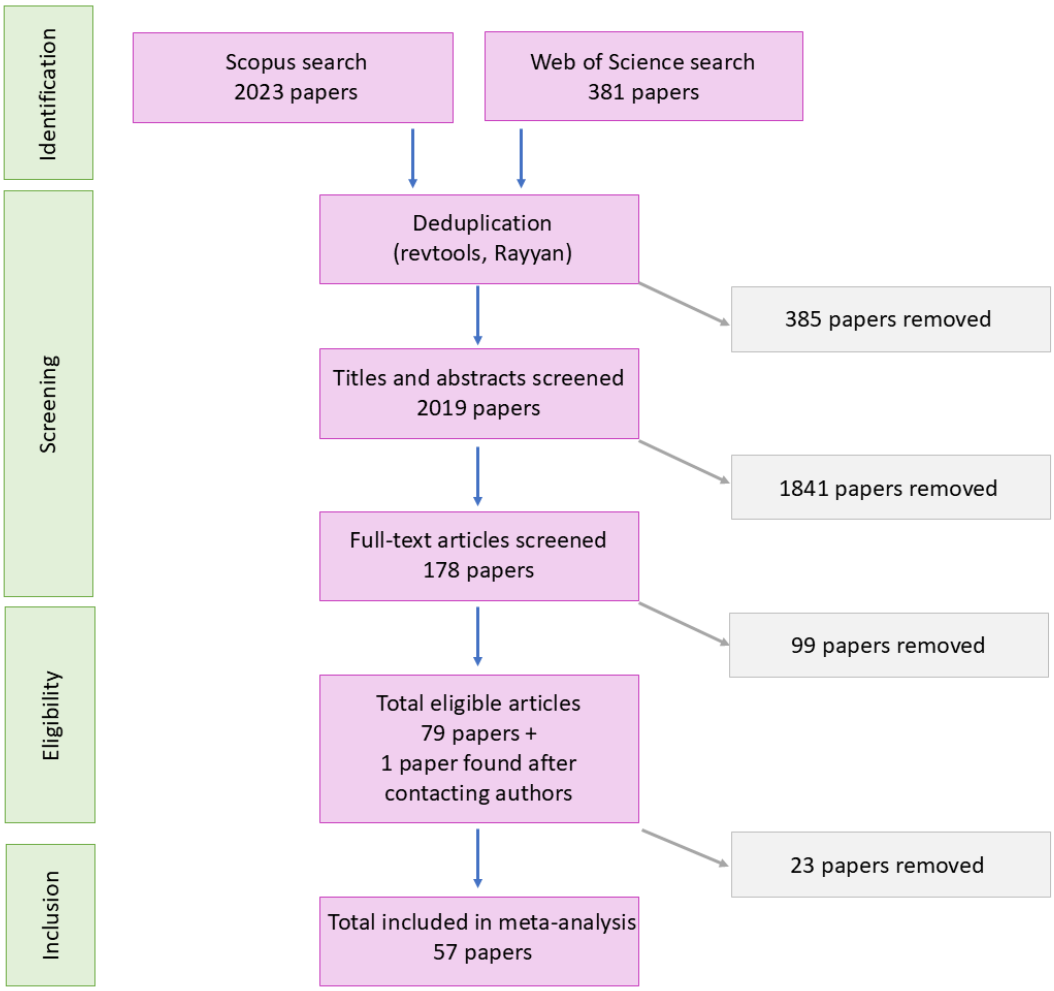
**Supplementary Figure 1.** Quality assessment of our systematic review and meta-analysis testing the extent to which prenatal maternal hormone deposition into eggs relates to fitness in wild birds. For it, we filled out the 'Interactive PRISMA-EcoEvo Checklist' (https://prisma-ecoevo.shinyapps.io/checklist; O'Dea *et al.* 2021).



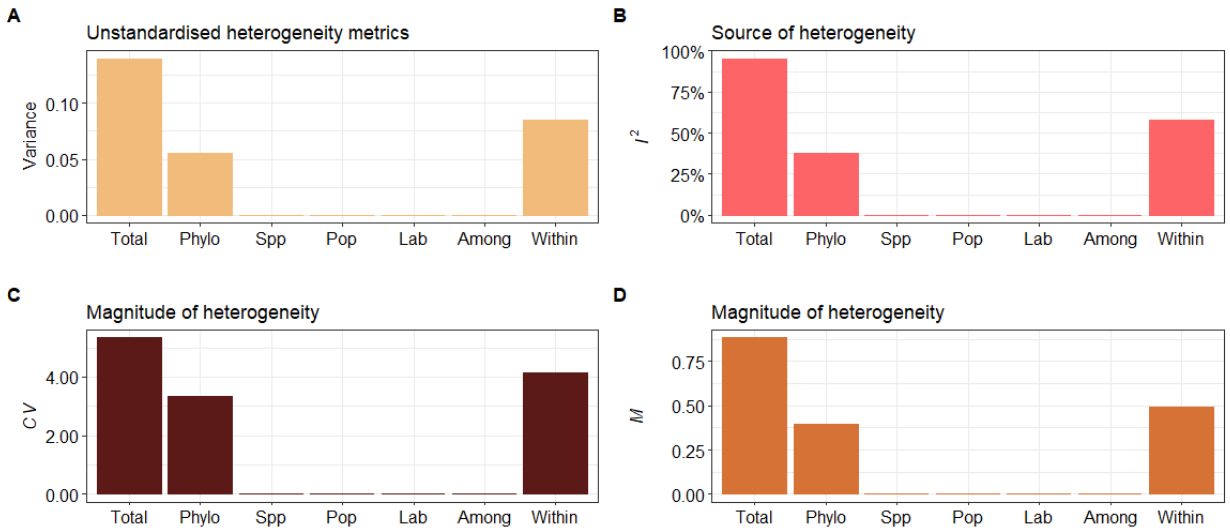
**Supplementary Figure 2.** Complete decision tree used for both title-and-abstract and full-text screening. Note that in our pre-registration (Mentesana *et al.* 2021b), we had separate trees for title-and-abstract and full-text screening, but here we present a unified and complete version for clarity.



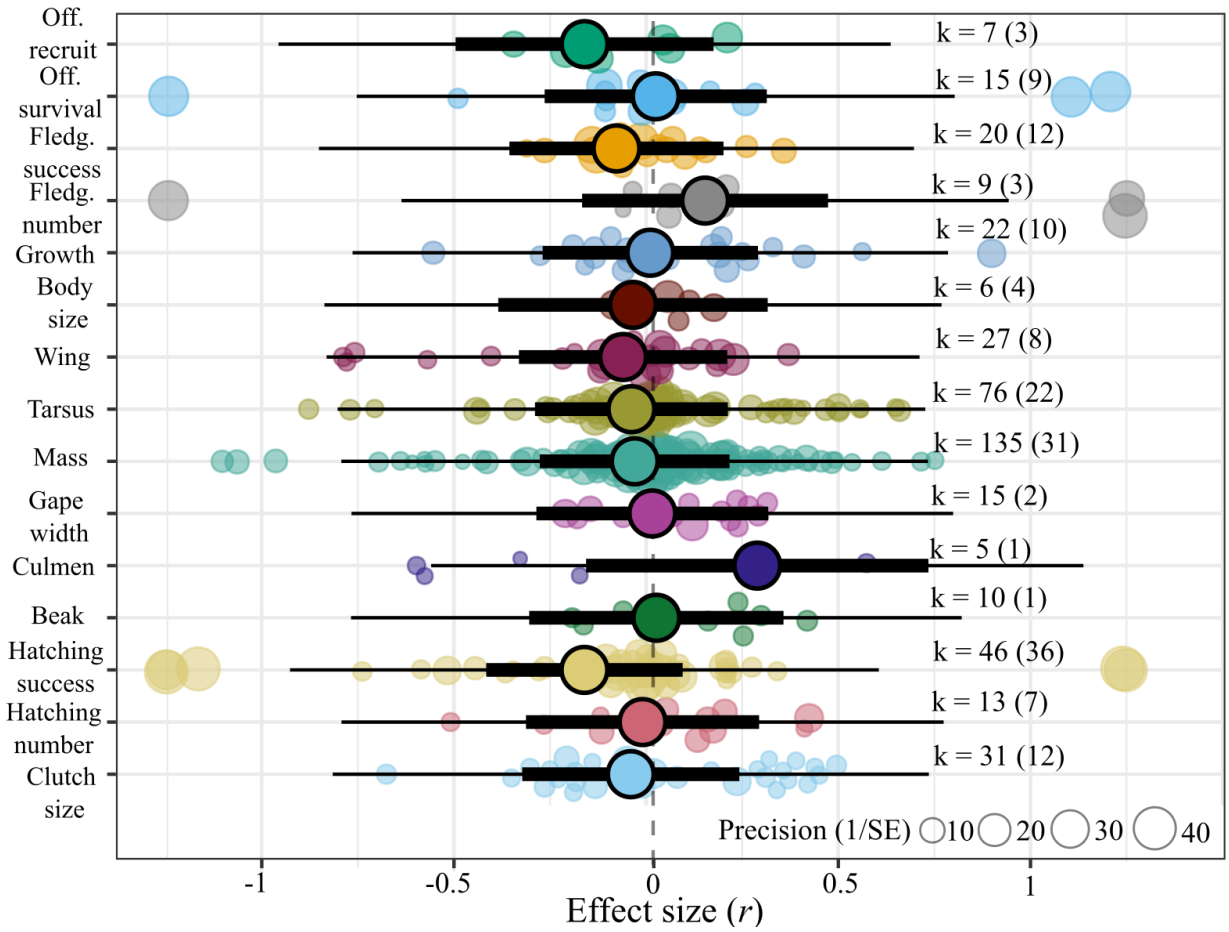
**Supplementary Figure 3.** PRISMA flow diagram summarizing our literature search. Of 80 eligible articles, we could not include 20 studies in the meta-analysis because articles reported incomplete information (e.g., missing sample sizes) and the authors did not provide such information despite our author contacting efforts (see "Materials and Methods" section), 1 article did not explore the effect of a hormone *per se* but that of a hormone blocker, and 2 studies were conducted in turtles.



**Supplementary Figure 4.** Heterogeneity metrics and stratification for a meta-analysis testing the association between maternal egg hormones and fitness across 19 bird species. The heterogeneity is quantified using A) raw variance ($\sigma^2$), B) source measure $I^2$, C) magnitude measure $CV$, and D) magnitude measure $M$, and stratified at phylogenetic (Phylo), non-phylogenetic among-species (Spp), among-population (Pop), among-laboratory (Lab), among- (Among) and within-study (Within) levels.



**Supplementary Figure 5.** Effect sizes obtained from the full dataset (considering both maternal and offspring fitness) using a phylogenetic multilevel meta-regression model with 'fitness proxies' as the sole moderator. Orchard plot showing mean estimates (circles with black outlines), 95% confidence intervals (thick whisker), 95% prediction intervals (thin whisker), and individual effect sizes scaled by their precision (coloured circles). k indicates the number of individual effect sizes and, in between brackets, the number of studies. Note that this model was not pre-registered and was conducted after obtaining the results to better understand them.



**Supplementary Information S1.** Quality assessment of our systematic review and meta-analysis testing the extent to which prenatal maternal hormone deposition into eggs relates to fitness in wild birds. For it, we filled out the 'Interactive PRISMA-EcoEvo Checklist' (https://prisma-ecoevo.shinyapps.io/checklist; O'Dea *et al.* 2021).

| Checklist item | Item score | Sub-item number | Sub-item | Reported by authors? |
|---|---|---|---|---|
| Title and abstract | 100% | 1.1 | Identify the review as a systematic review, meta-analysis, or both | Yes |
| | | 1.2 | Summarise the aims and scope of the review | Yes |
| | | 1.3 | Describe the data set | Yes |
| | | 1.4 | State the results of the primary outcome | Yes |
| | | 1.5 | State conclusions | Yes |
| | | 1.6 | State limitations | Yes |
| Aims and questions | 100% | 2.1 | Provide a rationale for the review | Yes |
| | | 2.2 | Reference any previous reviews or meta-analyses on the topic | Yes |
| | | 2.3 | State the aims and scope of the review (including its generality) | Yes |
| | | 2.4 | State the primary questions the review addresses (e.g. which moderators were tested) | Yes |
| | | 2.5 | Describe whether effect sizes were derived from experimental and/or observational comparisons | Yes |
| Review registration | 100% | 3.1 | Register review aims, hypotheses (if applicable), and methods in a time-stamped and publicly accessible archive and provide a link to the registration in the methods section of the manuscript. Ideally registration occurs before the search, but it can be done at any stage before data analysis. | Yes |

| | | | | |
|---|---|---|---|---|
| | | 3.2 | Describe deviations from the registered aims and methods | Yes |
| | | 3.3 | Justify deviations from the registered aims and methods | Yes |
| Eligibility criteria | 100% | 4.1 | Report the specific criteria used for including or excluding studies when screening titles and/or abstracts, and full texts, according to the aims of the systematic review (e.g. study design, taxa, data availability) | Yes |
| | | 4.2 | Justify criteria, if necessary (i.e. not obvious from aims and scope) | Yes |
| Finding studies | 100% | 5.1 | Define the type of search (e.g. comprehensive search, representative sample) | Yes |
| | | 5.2 | State what sources of information were sought (e.g. published and unpublished studies, personal communications) | Yes |
| | | 5.3 | Include, for each database searched, the exact search strings used, with keyword combinations and Boolean operators | Yes |
| | | 5.4 | Provide enough information to repeat the equivalent search (if possible), including the timespan covered (start and end dates) | Yes |
| Study selection | 100% | 6.1 | Describe how studies were selected for inclusion at each stage of the screening process (e.g. use of decision trees, screening software) | Yes |
| | | 6.2 | Report the number of people involved and how they contributed (e.g. independent parallel screening) | Yes |

| | | | | |
|---|---|---|---|---|
| Data collection process | 100% | 7.1 | Describe where in the reports data were collected from (e.g. text or figures) | Yes |
| | | 7.2 | Describe how data were collected (e.g. software used to digitize figures, external data sources) | Yes |
| | | 7.3 | Describe moderator variables that were constructed from collected data (e.g. number of generations calculated from years and average generation time) | Yes |
| | | 7.4 | Report how missing or ambiguous information was dealt with during data collection (e.g. authors of original studies were contacted for missing descriptive statistics, and/or effect sizes were calculated from test statistics) | Yes |
| | | 7.5 | Report who collected data | Yes |
| | | 7.6 | State the number of extractions that were checked for accuracy by co-authors | Yes |
| Data items | 100% | 8.1 | Describe the key data sought from each study | Yes |
| | | 8.2 | Describe items that do not appear in the main results, or which could not be extracted due to insufficient information | Yes |
| | | 8.3 | Describe main assumptions or simplifications that were made (e.g. categorising both 'length' and 'mass' as 'morphology') | NA |
| | | 8.4 | Describe the type of replication unit (e.g. individuals, broods, study sites) | Yes |

| | | | | |
|---|---|---|---|---|
| Assessment of individual study quality | 100% | 9.1 | Describe whether the quality of studies included in the systematic review or meta-analysis was assessed (e.g. blinded data collection, reporting quality, experimental versus observational) | Yes |
| | | 9.2 | Describe how information about study quality was incorporated into analyses (e.g. meta-regression and/or sensitivity analysis) | Yes |
| Effect size measures | 100% | 10.1 | Describe effect size(s) used | Yes |
| | | 10.2 | Provide a reference to the equation of each calculated effect size (e.g. standardised mean difference, log response ratio) and (if applicable) its sampling variance | Yes |
| | | 10.3 | If no reference exists, derive the equations for each effect size and state the assumed sampling distribution(s) | NA |
| Missing data | 100% | 11.1 | Describe any steps taken to deal with missing data during analysis (e.g. imputation, complete case, subset analysis) | Yes |
| | | 11.2 | Justify the decisions made to deal with missing data | Yes |
| Meta-analytic model description | 100% | 12.1 | Describe the models used for synthesis of effect sizes | Yes |
| | | 12.2 | The most common approach in ecology and evolution will be a random-effects model, often with a hierarchical/multilevel structure. If other types of models are chosen (e.g. common/fixed effects model, | NA |

| | | | | |
|---|---|---|---|---|
| | | | unweighted model), provide justification for this choice | |
| Software | 80% | 13.1 | Describe the statistical platform used for inference (e.g. R) | Yes |
| | | 13.2 | Describe the packages used to run models | Yes |
| | | 13.3 | Describe the functions used to run models | Yes |
| | | 13.4 | Describe any arguments that differed from the default settings | No |
| | | 13.5 | Describe the version numbers of all software used | Yes |
| Non-independence | 100% | 14.1 | Describe the types of non-independence encountered (e.g. phylogenetic, spatial, multiple measurements over time) | Yes |
| | | 14.2 | Describe how non-independence has been handled | Yes |
| | | 14.3 | Justify decisions made | Yes |
| Meta-regression and model selection | 100% | 15.1 | Provide a rationale for the inclusion of moderators (covariates) that were evaluated in meta-regression models | Yes |
| | | 15.2 | Justify the number of parameters estimated in models, in relation to the number of effect sizes and studies (e.g. interaction terms were not included due to insufficient sample sizes) | Yes |

| | | | | |
|---|---|---|---|---|
| | | 15.3 | Describe any process of model selection | NA |
| Publication bias and sensitivity analysis | 100% | 16.1 | Describe assessments of the risk of bias due to missing results (e.g. publication, time-lag, and taxonomic biases) | Yes |
| | | 16.2 | Describe any steps taken to investigate the effects of such biases (if present) | Yes |
| | | 16.3 | Describe any other analyses of robustness of the results, e.g. due to effect size choice, weighting or analytical model assumptions, inclusion or exclusion of subsets of the data, or the inclusion of alternative moderator variables in meta-regressions | Yes |
| Clarification of post hoc analyses | 100% | 17.1 | When hypotheses were formulated after data analysis, this should be acknowledged. | Yes |
| Metadata, data, and code | | 18.1 | Share metadata (i.e. data descriptions) | Yes |
| | | 18.2 | Share data required to reproduce the results presented in the manuscript | Yes |
| | | 18.3 | Share additional data, including information that was not presented in the manuscript (e.g. raw data used to calculate effect sizes, descriptions of where data were located in papers) | Yes |
| | | 18.4 | Share analysis scripts (or, if a software package with graphical user interface (GUI) was used, then describe full model specification and fully specify choices) | Yes |

| | | | | |
|---|---|---|---|---|
| Results of study selection process | 100% | 19.1 | Report the number of studies screened | Yes |
| | | 19.2 | Report the number of studies excluded at each stage of screening | Yes |
| | | 19.3 | Report brief reasons for exclusion from the full text stage | Yes |
| | | 19.4 | Present a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-like flowchart (www.prisma-statement.org). | Yes |
| Sample sizes and study characteristics | 75% | 20.1 | Report the number of studies and effect sizes for data included in meta-analyses | Yes |
| | | 20.2 | Report the number of studies and effect sizes for subsets of data included in meta-regressions | Yes |
| | | 20.3 | Provide a summary of key characteristics for reported outcomes (either in text or figures; e.g. one quarter of effect sizes reported for vertebrates and the rest invertebrates) | Yes |
| | | 20.4 | Provide a summary of limitations of included moderators (e.g. collinearity and overlap between moderators) | NA |
| | | 20.5 | Provide a summary of characteristics related to individual study quality (risk of bias) | No |
| Meta-analysis | 100% | 21.1 | Provide a quantitative synthesis of results across studies, including estimates for the mean effect size, with confidence/credible intervals | Yes |

| | | | | |
|---|---|---|---|---|
| Heterogeneity | | 22.1 | Report indicators of heterogeneity in the estimated effect (e.g. I2, tau2 and other variance components) | Yes |
| Meta-regression | | 23.1 | Provide estimates of meta-regression slopes (i.e. regression coefficients) and confidence/credible intervals | Yes |
| | | 23.2 | Include estimates and confidence/credible intervals for all moderator variables that were assessed (i.e. complete reporting) | Yes |
| | | 23.3 | Report interactions, if they were included | Yes |
| | | 23.4 | Describe outcomes from model selection, if done (e.g. R2 and AIC) | NA |
| Outcomes of publication bias and sensitivity analysis | 100% | 24.1 | Provide results for the assessments of the risks of bias (e.g. Egger's regression, funnel plots) | Yes |
| | | 24.2 | Provide results for the robustness of the review's results (e.g. subgroup analyses, meta-regression of study quality, results from alternative methods of analysis, and temporal trends) | Yes |
| Discussion | 100% | 25.1 | Summarise the main findings in terms of the magnitude of effect | Yes |
| | | 25.2 | Summarise the main findings in terms of the precision of effects (e.g. size of confidence intervals, statistical significance) | Yes |
| | | 25.3 | Summarise the main findings in terms of their heterogeneity | Yes |
| | | 25.4 | Summarise the main findings in terms of their biological/practical relevance | Yes |

| | | | | |
|---|---|---|---|---|
| | | 25.5 | Compare results with previous reviews on the topic, if available | Yes |
| | | 25.6 | Consider limitations and their influence on the generality of conclusions, such as gaps in the available evidence (e.g. taxonomic and geographical research biases) | Yes |
| Contributions and funding | 100% | 26.1 | Provide names, affiliations, and funding sources of all co-authors | Yes |
| | | 26.2 | List the contributions of each co-author | Yes |
| | | 26.3 | Provide contact details for the corresponding author | Yes |
| | | 26.4 | Disclose any conflicts of interest | Yes |
| References | 100% | 27.1 | Provide a reference list of all studies included in the systematic review or meta-analysis | Yes |
| | | 27.2 | List included studies as referenced sources (e.g. rather than listing them in a table or supplement) | Yes |

**Supplementary Information S2:** The full keyword search string used, with a minor syntax edit at the beginning of the string for the electronic search engine Web of Science[†] ("TS =…") and Scopus ("ALL…"), was:

(("maternal effect*" OR "parental effect*" OR "maternal invest*" OR "parental invest*" OR "maternal program*" OR "parental program*" OR "maternal hormon*" OR "epigenetic effect*" OR "epigenetic inheritance" OR "trans-generational phenotypic plasticity" OR "trans-generational plasticity" OR "transfer* from mother* to offspring" OR "transfer* from parent* to offspring" OR "transgeneration* induc*" OR "transgenerational adaptive plasticity" OR "transgenerational induction" OR "transgenerational phenotypic plasticity" OR "transgenerational plasticity" OR "transmit* between generations" OR "transmitted from mother* to offspring" OR "transmitted from parent* to offspring" OR "*generation* effect*" OR "anticipatory effect*" OR "bet hedging" OR "carry-over" OR "carryover" OR "developmental conditioning" OR "developmental programming" OR "fetal programming" OR "maternal conditioning" OR "maternal match*" OR "paternal conditioning" OR "silver spoon" OR "transgenerational process*" OR "*cross generation* effect" OR "imprinting" OR "inter-generational process*" OR "adaptive cross-generational hypothesis" OR "adaptive epigenetic memory" OR "between generation* effect*" OR "embryo* conditioning" OR "environmental matching hypothes*" OR "intergenerational phenotypic plasticity" OR "intergenerational plasticity" OR "maternal hormon* effect*" OR "maternal environment enhanc* offspring" OR "maternal environment* effect*" OR "maternal mediation hypothes*" OR "maternal mismatch hypothes*" OR "maternal rearing condition*" OR "mother-to-offspring transfer*" OR "mothers forewarn offspring" OR "parent-to-offspring transfer*" OR "parental environment* effect*" OR "parental rearing condition*" OR "paternal environment* effect*" OR "phenotypeenvironment *matching" OR "phenotype environment *match" OR "plasticity *cross generation*" OR "prim* the* offspring" OR "prime* offspring" OR "telescoping generation*" OR "priming" OR "generational memory") AND ("yolk steroid* hormon*" OR " yolk androgen*" OR "yolk androstenedion*" OR "yolk testosteron*" OR "yolk 5α-dihydrotestosteron*" OR "yolk DHT" OR "yolk glucocorticoid*" OR "yolk corticosteron*" OR "yolk cortisol" OR "yolk progesteron*" OR "yolk estrogen*" OR "yolk T4" OR "yolk T3" OR "yolk thyroid" OR "egg steroid* hormon*" OR "egg androgen*" OR "egg androstenedion*" OR "egg testosteron*" OR "egg 5α-dihydrotestosteron*" OR "egg DHT" OR "egg glucocorticoid*" OR "egg corticosteron*" OR "egg cortisol" OR "egg progesteron*" OR "egg estrogen*" OR "egg T4" OR "egg T3" OR "egg thyroid" OR "egg cort*" OR "yolk cort*" OR "egg estradiol" OR "yolk estradiol" OR "egg GC*" OR "yolk GC*" OR "egg hormon*" OR "yolk hormon*" OR "album* steroid* hormon*" OR "album* androgen*" OR "album* androstenedion*" OR " album* testosteron*" OR " album* 5α-dihydrotestosteron*" OR "album* DHT" OR "album* glucocorticoid*" OR "album* corticosteron*" OR "album* cortisol" OR "album* progesteron*" OR "album* estrogen*" OR "album* T4" OR "album* T3" OR "album* thyroid" OR "album* cort*" OR "album* estradiol" OR "album* GC*" OR "album* hormon*") AND ("fitness" OR "breed* success*" OR "clutch*" OR "hatch*" OR "fledg*" OR "embryo* development*" OR "reproductive success" OR "surviv*" OR "condition" OR "size" OR "growth" OR "mass" OR "tarsi" OR "tarsus" OR "wing*" OR "quality" OR "weight" OR "bone*" OR "standard length" OR "total length"))

[†] Databases covered with Web of Science Core Collection: Science Citation Index Expanded (SCIE); Social Sciences Citation Index (SSCI); Arts & Humanities Citation Index (AHCI); Emerging Sources Citation Index (ESCI)).

**Supplementary Information S3:** Full list of fitness proxies used in the current meta-analysis. Note that for some fitness proxies, it is difficult to fully disentangle maternal and offspring fitness (e.g., hatching or fledging success; discussed by Wolf & Wade 2001). In such cases, following the recommendations of Wolf & Wade (2001), we took a conservative approach, assigning to the mother those fitness proxies measured in the offspring that are largely or wholly controlled by her. k: number of effect sizes; N: number of studies. Fitness proxies shown in grey were initially extracted but could not be included in the final analysis due to incomplete data.

* Fitness proxies for offspring when recaptured as breeding adults.

| | Maternal fitness proxies | k | N | Offspring fitness proxies | k | N |
|---|---|---|---|---|---|---|
| 1 | Clutch size | 29 | 11 | Offspring survival | 13 | 8 |
| 2 | Egg mortality | 2 | 1 | Beak flank width | 10 | 1 |
| 3 | Hatching probability | 3 | 3 | Culmen | 5 | 1 |
| 4 | Hatching success | 43 | 34 | Fledgling number | 9 | 3 |
| 5 | Hatching failure | 1 | 1 | Fledging probability | 1 | 1 |
| 6 | Maternal lifetime reproductive success | 2 | 1 | Fledging success | 14 | 10 |
| 7 | Maternal longevity | 2 | 1 | Fledging success (fledglings/hatchings) | 1 | 1 |
| 8 | Offspring recruitment | 4 | 1 | Flipper length | 1 | 1 |
| 9 | Maternal survival | - | - | Gape width | 15 | 2 |
| 10 | Maternal number of reproductive events over different years | - | - | Growth (mass) | 3 | 1 |
| 11 | | | | Growth (tarsus) | 3 | 1 |
| 12 | | | | Growth (PC1) | 2 | 1 |
| 13 | | | | Growth (PC2) | 2 | 1 |
| 14 | | | | Growth rate | 1 | 1 |
| 15 | | | | Growth rate (flipper length) | 1 | 1 |
| 16 | | | | Growth rate (mass) | 6 | 4 |
| 17 | | | | Growth rate (tarsus) | 1 | 1 |
| 18 | | | | Growth rate (ulna) | 1 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 19 | | | | Clutch size* | 2 | 1 |
| 20 | | | | Hatching number* | 11 | 6 |
| 21 | | | | Offspring survival years | 2 | 1 |
| 22 | | | | Head-bill length | 1 | 1 |
| 23 | | | | Offspring mortality | 2 | 1 |
| 24 | | | | Pre-fledging survival probability | 4 | 1 |
| 25 | | | | Head length | 1 | 1 |
| 26 | | | | Mass | 135 | 31 |
| 27 | | | | Mass gain | 2 | 1 |
| 28 | | | | Recruitment | 1 | 1 |
| 29 | | | | Structural body size (mass and bill length) | 1 | 1 |
| 30 | | | | Structural body size (mass and tarsus) | 5 | 4 |
| 31 | | | | Tarsus length | 76 | 22 |
| 32 | | | | Wing length | 26 | 7 |
| 33 | | | | Fledgling success_egg1 | - | - |
| 34 | | | | Fledgling success_egg2 | - | - |
| 35 | | | | Fledgling success_egg3 | - | - |
| 36 | | | | Fledgling success_egg4 | - | - |
| 37 | | | | Fledgling success_egg5 | - | - |
| 38 | | | | Fledgling success_egg6 | - | - |
| 39 | | | | Growth rate (head length) | - | - |
| 40 | | | | Growth rate (wing) | - | - |
| 41 | | | | Survival | - | - |
| 42 | | | | Tail length | - | - |
| 43 | | | | Ulna length | | |

**Supplementary Information S4:** Standardized email template sent to those authors whose articles had incomplete information to extract effect sizes and moderator information.

Dear XXX,

My name is Lucia Mentesana. I am a postdoctoral researcher at the Max Planck for Biological Intelligence (Germany). Together with Prof. Dr. Michaela Hau, Dr. Nicolas Adreani, Dr. Pietro B. D'Amelio & Dr. Alfredo Sánchez-Tójar, we are currently conducting a meta-analysis to test to what extent prenatal maternal egg hormone deposition relates to fitness in free-living birds. For this, we conducted a systematic literature search that identified 77 articles with relevant data for our meta-analysis and from where we have extracted all available information. Among these valuable studies, we found your work!

The reason why I am contacting you is because after carefully reading your work, we realized that we would need some extra information about your work to be able to include it in our meta-analysis (see details in the Word files attached). We were wondering if you could help us obtaining the missing information so that we can include your important work in our meta-analysis and make the results of it more robust, complete and generalizable. We would then of course make sure to cite your work in our manuscript.

We are aiming to finalize our analyses in about a month, so please, if you think you can provide us the missing information, but may need some extra time, do let us know. Please, do not hesitate to contact us if you have any questions.

Looking forward to hearing from you!
Thank you very much in advance!
Warm regards,
Lucia Mentesana et al.

**Supplementary Information S5:** Information on the continuous and categorical moderators used in the meta-regression models.

**Table S5.1:** Description of the continuous moderator used for meta-regression models.

| Continuous moderator | Description |
|---|---|
| Offspring relative age | Age of the offspring/Average nesting time across nests (both values obtained from information reported in the paper), with 0 = hatching day. median = 0.500; range = 0 - 1.125; mean = 0.445, SD = 0.304; k = 215; N = 19 |
| Square root of the inverse of the sample size | This is simply the square root of the inverse of the sample size used to calculate the effect size and its sampling variance. This moderator was used to test for small-study effects following (Nakagawa *et al.* 2022) |
| Year of publication (mean-centered) | The year when the study was published - the mean year of publication across our dataset. Mean-centering was performed to aid interpretation |

**Table S5.2:** Description of the categorical moderators used for meta-regression models.

| Categorical moderator | Moderator levels | Description |
|---|---|---|
| Study type | 1. Correlational<br>2. Experimental | Type of study |
| Egg hormone | 1. Androgens (androstenedione, 5α-dihydrotestosterone & testosterone)<br>2. Glucocorticoids<br>3. Thyroids (TH3 & TH4) | Type of hormone measured/manipulated |
| Experimental subject | 1. Mother<br>2. Egg | Subject that was experimentally manipulated |
| Hormone matrix | 1. Albumen<br>2. Yolk | Egg site where the hormone was measured |
| Experimental dose | 1. Supraphysiological<br>2. Within natural range | Range of hormone dose used for the experiment as stated by authors |
| Egg sampling method | 1. Entire egg removed<br>2. Biopsy | Method of matrix collection |
| Offspring life history stage | 1. Hatching<br>2. Before independence<br>3. After independence | Offspring age categories |
| Offspring sex | 1. Female<br>2. Male | Sex of the offspring determined in the paper |
| Data reporting | 1. Complete<br>2. Incomplete | Incomplete whenever authors did not report sample sizes, test for directionality, insufficient information on statistical models, insufficient information on statistical inference (e.g., missing p-values, CrI or CI) |
| Partial or selective reporting | 1. Yes<br>2. No | No: when authors did not report or partially reported results, for example, because the results were not statistically significant |

**Supplementary Information S6:** Effect size calculation.

**Table S6:** Equations used for converting the reported results into Pearson correlation coefficients (*r*; from Lajeunesse 2013; Nakagawa & Cuthill 2007). The sampling variance of *r* was calculated as $Vr = (1 - r^2)^2 \div (N - 1)$ (Borenstein *et al.* 2009).

| Statistical test reported in the original paper | Equation |
|---|---|
| Absolute value | $r = [(Nexp_survived * Ncontrol_not_survived) - (Ncontrol_survived * Nexp_not_survived)] \div \text{sqrt}[(Nexp_survived + Ncontrol_survived) * (Nexp_not_survived + Ncontrol_not_survived) * (Nexp_survived + Nexp_not_survived) * (Ncontrol_survived + Ncontrol_not_survived)]$ |
| Chi-square value *($X^2$)* | $r = \sqrt{X^2 \div N}$ |
| *F*-value | $r = \sqrt{F \div (F + N - 2)}$ |
| Spearman's rho correlation coefficient ($\rho$) | $r = 2 * \sin(\pi\rho \div 6)$ |
| *t*-value | $r = t \div \sqrt{t^2 + (N - 2)}$ |
| *z*-score | $r = z \div \sqrt{N}$ |

**Supplementary Information S7:** Sensitivity analyses conducted to confirm the robustness of our results using a vector of sampling variances instead of the variance-covariance matrix. These analyses were performed for the intercept-only model of i) the entire database and ii) only offspring fitness traits, and iii) the hormone-type meta-regression for the offspring database. [*non-pre-registered analyses*]

| Models | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | Q | p-value | k | N |
|---|---|---|---|---|---|---|---|---|---|
| Intercept only models | | | | | | | | | |
| i) For the entire database | -0.052 | -0.257 | 0.153 | -0.764 | 0.659 | 24490 | 0.617 | 443 | 57 |
| ii) For the offspring database | -0.043 | -0.304 | 0.218 | -0.733 | 0.646 | 6953 | 0.745 | 355 | 43 |
| iii) Meta-regression model for the offspring database | | | | | | | | | |
| Androgens | -0.055 | -0.319 | 0.209 | -0.750 | 0.640 | 6917 | 0.681 | 263 | 34 |
| Glucocorticoids | 0.007 | -0.268 | 0.282 | -0.692 | 0.706 | | 0.959 | 56 | 8 |
| Thyroids | -0.011 | -0.290 | 0.269 | -0.712 | 0.690 | | 0.940 | 36 | 3 |

CI: confidence interval; PI: prediction interval; L: lower; U: upper; *Q:* heterogeneity; k: number of effect sizes; N: number of studies.

**Supplementary Information S8:** Sensitivity analyses testing the robustness of our results to different choices of effect size types. Note that these analyses used data on both maternal and offspring fitness.

**Table S8:** Model result table comparing the estimates from the original model (i.e., intercept-only model analyzing both Pearson and Biserial correlations) to those of the three intercept-only models only including: (a) Pearson's *r* correlations, (b) Biserial correlations, or (c) log-response ratio coefficients (lnRR). [*non-pre-registered analyses*]

| | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | $I^2$ total | p-value | k | N |
|---|---|---|---|---|---|---|---|---|---|
| Original (full) model | | | | | | | | | |
| | -0.070 | -0.328 | 0.188 | -0.84 | 0.709 | 95.3 | 0.593 | 443 | 57 |
| Sensitivity analysis: three intercept-only models | | | | | | | | | |
| a) Pearson | 0.039 | -0.029 | 0.107 | -0.333 | 0.411 | 87.6 | 0.264 | 166 | 40 |
| b) Biserial | -0.079 | -0.336 | 0.177 | -0.946 | 0.787 | 94.9 | 0.543 | 277 | 39 |
| c) lnRR | -0.004 | -0.011 | 0.004 | -0.079 | 0.071 | 98.5 | 0.313 | 274 | 38 |

CI: confidence interval; PI: prediction interval; L: lower; U: upper; $I^2$: total relative heterogeneity; k: number of effect sizes; N: number of studies.

Model a) "Pearson" used the subset of the data containing Pearson's *r* correlations that were either extracted directly from the original articles or transformed either from other types of correlations or from inferential statistics. Model b) "biserial" and c) "lnRR" used the subset of the data that were extracted as means, SDs and sample sizes, and then used for calculating either biserial correlations or lnRR (more in the section "2.6. Effect size calculation" from the main manuscript). lnRR cannot be calculated for non-ratio scale data, explaining the reduction in sample size in (c) compared to (b). None of the three sensitivity analyses deviated from the conclusion of the main model, that is, the meta-analytic mean, though slightly negative, is non-statistically significantly from zero. However, note that these models do not differentiate between hormone groups, and that a) the full model contains both experimental and correlational data, whereas the subset of data used for b) and c) corresponds to results from experimental data only.

**Supplementary Information S9:** Sensitivity analyses testing the robustness of our results for predictions BH.1.1 - BH.1.3. Note that these analyses used data on offspring fitness only.

**Table S9.1:** Model result table showing the estimates from three intercept-only models including the data of each egg hormone separately compared to the estimates shown in the main text, which come from a meta-regression including hormone type as a moderator with three levels.

| Models | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | *Q* | p-value | k | N |
|---|---|---|---|---|---|---|---|---|---|
| BH.1.1 (androgens) | -0.057 | -0.425 | 0.300 | -0.972 | 0.858 | 12413 | 0.761 | 263 | 34 |
| BH.1.2 (glucocorticoids) | 0.077 | -0.032 | 0.186 | -0.225 | 0.378 | 176 | 0.164 | 56 | 8 |
| BH.1.3 (thyroids) | 0.022 | -0.083 | 0.127 | -0.331 | 0.375 | 200 | 0.672 | 36 | 3 |

CI: confidence interval; PI: prediction interval; L: lower; U: upper; *Q:* heterogeneity; k: number of effect sizes; N: number of studies.

**Table S9.2:** Heterogeneity metrics and stratification from three intercept-only models including the data of each egg hormone separately.

| | Androgens | | | | Glucocorticoids | | | | Thyroids | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $I^2$ (%) | *CV* | *M* | $\sigma^2$ | $I^2$ (%) | *CV* | *M* | $\sigma^2$ | $I^2$ (%) | *CV* | *M* | $\sigma^2$ |
| Total | 95.10 | 7.46 | 0.57 | 0.18 | 74.52 | 1.83 | 0.51 | 0.02 | 80.84 | 7.56 | 0.88 | 0.03 |
| Among-study (Study ID) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Among-laboratory (Laboratory ID) | 0.00 | 0.00 | 0.00 | 0.00 | 42.17 | 1.38 | 0.38 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| Among-population (Population ID) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Among-species (Species name) | 6.17 | 1.90 | 0.15 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Phylogenetic relationships (Species phylogeny) | 55.76 | 5.71 | 0.44 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Within-Study (Effect ID) | 33.17 | 4.40 | 0.34 | 0.06 | 32.35 | 1.20 | 0.34 | 0.01 | 80.84 | 7.56 | 0.88 | 0.03 |

$I^2$: relative heterogeneity, *CV*: mean-standardized heterogeneity, *M*: variance-mean-standardized heterogeneity, $\sigma^2$: total heterogeneity.

**Table S9.3**. Small-study effects analyses for each of the three intercept-only models including the data of each egg hormone separately.

| Hormone | Intercept | L 95% CI | U 95% CI | p-value | Sqrt (1/N) | L 95% CI | U 95% CI | p-value | $R^2_{marginal}$ | k | N |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Androgen | -0.152 | -0.465 | 0.161 | 0.340 | 0.475 | -0.332 | 1.282 | 0.248 | 0.8% | 338 | 46 |
| Glucocorticoids | 0.057 | -0.085 | 0.198 | 0.426 | 0.242 | -0.855 | 1.338 | 0.661 | 0.9% | 66 | 10 |
| Thyroids | -0.227 | -0.467 | 0.013 | 0.063 | 3.016 | 0.422 | 5.611 | 0.024 | 23.8% | 39 | 3 |

CI: confidence interval; PI: prediction interval; L: lower; U: upper; k: number of effect sizes; N: number of studies.

**Supplementary Information S10:** Heterogeneity metrics and stratification for a meta-analysis testing the association between maternal egg hormones and fitness - both offspring and maternal - across 19 bird species.

| | $I^2$ (%) | $CV$ | $M$ | $\sigma^2$ |
|---|---|---|---|---|
| Total | 95.2 | 5.10 | 0.63 | 0.14 |
| Among-study (Study ID) | 0.00 | 0.00 | 0.00 | 0.00 |
| Among-laboratory (Laboratory ID) | 0.00 | 0.00 | 0.00 | 0.00 |
| Among-population (Population ID) | 0.00 | 0.00 | 0.00 | 0.00 |
| Among-species (Species name) | 0.00 | 0.00 | 0.00 | 0.00 |
| Phylogenetic relationships (Species phylogeny) | 36.17 | 3.15 | 0.39 | 0.05 |
| Within-Study (Effect ID) | 59.03 | 4.02 | 0.49 | 0.09 |

$I^2$: relative heterogeneity, $CV$: mean-standardized heterogeneity, $M$: variance-mean-standardized heterogeneity, $\sigma^2$: total heterogeneity.

$I^2_{total}$ indicates that heterogeneity in our dataset is, on average, ca. 20 times larger than statistical noise, and $CV_{total}$ indicates that heterogeneity is, on average, ca. 5 times larger than the meta-analytic mean. $M_{total}$ was also high, indicating that absolute heterogeneity is large relative to the magnitude of the meta-analytic mean. These analyses also suggest that once the phylogenetic relationships between the 19 bird species included in our meta-analysis have been accounted for, the only important source of heterogeneity is the within-study level, indicating that it is within-study level predictors that are most likely driving the observed context-dependency for the hypothesis that maternal egg hormones associate with fitness. Said differently, despite the high levels of heterogeneity observed, the meta-analytic mean observed could still be generalized at the between-study level. if we define replication as the testing of the null hypothesis at the between-study level and when within-study variations (methodological and biological) can be accounted for.

**Supplementary Information S11:** Hormone-specific small-study effects analyses across the full dataset (i.e., considering both maternal and offspring fitness).

| | Mean | SE | t-value | df | p-value | L 95% CI | U 95% CI |
|---|---|---|---|---|---|---|---|
| Intercept (Androgens) | -0.166 | 0.155 | -1.071 | 437 | 0.285 | -0.471 | 0.139 |
| Glucocorticoids | 0.076 | 0.104 | 0.726 | 437 | 0.468 | -0.129 | 0.281 |
| Thyroids | -0.044 | 0.200 | -0.220 | 437 | 0.826 | -0.438 | 0.350 |
| sqrt(1/N) | 0.530 | 0.380 | 1.393 | 437 | 0.164 | -0.218 | 1.277 |
| Glucocorticoids * sqrt(1/N) | -0.130 | 0.954 | -0.136 | 437 | 0.892 | -2.005 | 1.746 |
| Thyroids * sqrt(1/N) | 1.183 | 2.097 | 0.564 | 437 | 0.573 | -2.938 | 5.304 |

df: degrees of freedom; L: lower; U: upper.

**Supplementary Information S12:** Decline effects analyses across the full dataset (i.e., considering both maternal and offspring fitness).

| | Mean | SE | t-value | df | p-value | L 95% CI | U 95% CI |
|---|---|---|---|---|---|---|---|
| Intercept (Androgens) | -0.084 | 0.132 | -0.635 | 437 | 0.526 | -0.343 | 0.175 |
| Glucocorticoids | 0.044 | 0.065 | 0.680 | 437 | 0.497 | -0.084 | 0.173 |
| Thyroids | 0.064 | 0.268 | 0.240 | 437 | 0.811 | -0.463 | 0.591 |
| Year | -0.001 | 0.005 | -0.103 | 437 | 0.918 | -0.011 | 0.010 |
| Glucocorticoids * Year | 0.005 | 0.009 | 0.543 | 437 | 0.587 | -0.013 | 0.023 |
| Thyroids * Year | -0.002 | 0.050 | -0.036 | 437 | 0.971 | -0.100 | 0.096 |

df: degrees of freedom; L: lower; U: upper.

**Supplementary Information S13:** Results of the phylogenetic multilevel meta-regression models testing whether the associations between hormone concentrations and both maternal and offspring fitness differ between effect sizes from studies categorized as partially or selectively reporting results.

| | Partial or selective reporting | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | p-value | k | N |
|---|---|---|---|---|---|---|---|---|---|
| Androgens | No | -0.020 | -0.087 | 0.047 | -0.324 | 0.283 | 0.547 | 54 | 6 |
| | Yes | -0.008 | -0.081 | 0.066 | -0.313 | 0.298 | 0.840 | 59 | 13 |
| Glucocorticoids | No | 0.090 | -0.009 | 0.190 | -0.222 | 0.402 | 0.075 | 14 | 3 |
| | Yes | -0.021 | -0.137 | 0.095 | -0.339 | 0.297 | 0.718 | 19 | 5 |
| Thyroids | No | 0.093 | -0.137 | 0.322 | -0.282 | 0.467 | 0.428 | 15 | 1 |
| | Yes | 0.223 | 0.042 | 0.404 | -0.124 | 0.570 | 0.042 | 10 | 1 |

CI: confidence interval; PI: prediction interval; L: lower; U: upper; k = number of effect sizes; N = number of studies.

**Supplementary Information S14:** Pseudo-all-in model to estimate the total amount of heterogeneity explained by complete-case moderators (i.e., small-study and decline effects). Note that these analyses used data on both maternal and offspring fitness.

| | Mean | SE | t-value | df | p-value | L 95% CI | U 95% CI |
|---|---|---|---|---|---|---|---|
| Intercept (Androgens) | -0.165 | 0.158 | -1.040 | 434 | 0.299 | -0.476 | 0.146 |
| Glucocorticoids | 0.063 | 0.109 | 0.582 | 434 | 0.561 | -0.150 | 0.277 |
| Thyroids | -0.213 | 0.429 | -0.497 | 434 | 0.620 | -1.055 | 0.629 |
| sqrt(1/N) | 0.546 | 0.391 | 1.396 | 434 | 0.163 | -0.223 | 1.316 |
| Year | 0.001 | 0.005 | 0.141 | 434 | 0.888 | -0.010 | 0.011 |
| Glucocorticoids * sqrt(1/N) | -0.107 | 0.965 | -0.110 | 434 | 0.912 | -2.003 | 1.790 |
| Thyroids * sqrt(1/N) | 1.675 | 2.341 | 0.716 | 434 | 0.475 | -2.927 | 6.277 |
| Glucocorticoids * Year | 0.004 | 0.009 | 0.431 | 434 | 0.667 | -0.015 | 0.0232 |
| Thyroids * Year | 0.024 | 0.056 | 0.425 | 434 | 0.671 | -0.086 | 0.134 |

df: degrees of freedom; L: lower; U: upper.

**Supplementary Information S15:** Results of the phylogenetic multilevel meta-regression model for our entire dataset, with 'fitness proxies' as the only moderator. Following our pre-registration, from all possible ‘fitness proxies’ (see Supplementary Information S2), we only included those levels that had at least 5 data points (14 levels, 19 bird species). Also, we included the same random effects as in the other meta-regression models: study ID, laboratory ID, population ID, species name, species phylogeny, and effect ID.

| Category | Fitness proxy reported in paper | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | p-value | k | N |
|---|---|---|---|---|---|---|---|---|---|
| Clutch size | | -0.040 | -0.322 | 0.242 | -0.814 | 0.736 | 0.871 | 31 | 12 |
| Hatching number | Hatching number | -0.009 | -0.313 | 0.294 | -0.793 | 0.774 | 0.952 | 13 | 7 |
| | Egg mortality | | | | | | | | |
| Hatching success | Hatching success | -0.161 | -0.416 | 0.095 | -0.927 | 0.606 | 0.217 | 46 | 36 |
| | Hatching probability | | | | | | | | |
| | Hatching failure | | | | | | | | |
| Beak flank width | | 0.027 | -0.304 | 0.357 | -0.768 | 0.821 | 0.874 | 10 | 1 |
| Culmen | | 0.289 | -0.156 | 0.735 | -0.559 | 1.138 | 0.203 | 5 | 1 |
| Gape width | | 0.016 | -0.286 | 0.318 | -0.767 | 0.799 | 0.917 | 15 | 2 |
| Mass | | -0.030 | -0.277 | 0.217 | -0.793 | 0.733 | 0.813 | 135 | 31 |
| Tarsus | | -0.038 | -0.289 | 0.213 | -0.803 | 0.726 | 0.764 | 76 | 22 |
| Wing | Wing | -0.060 | -0.331 | 0.211 | -0.831 | 0.711 | 0.664 | 27 | 8 |
| | Flipper length | | | | | | | | |
| Structural body size | Structural body size (mass and tarsus) | -0.034 | -0.385 | 0.316 | -0.837 | 0.769 | 0.848 | 6 | 4 |
| | Structural body size (mass and | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | bill length) | | | | | | | | |
| Growth | Growth (PC1) | 0.011 | -0.269 | 0.291 | -0.764 | 0.785 | 0.940 | 22 | 10 |
| | Growth (PC2) | | | | | | | | |
| | Growth (mass) | | | | | | | | |
| | Growth (tarsus) | | | | | | | | |
| | Growth rate | | | | | | | | |
| | Growth rate (mass) | | | | | | | | |
| | Growth rate (tarsus) | | | | | | | | |
| | Growth rate (ulna) | | | | | | | | |
| | Growth rate (flipper length) | | | | | | | | |
| | Mass gain | | | | | | | | |
| Fledgling number | | 0.153 | -0.167 | 0.473 | -0.637 | 0.943 | 0.347 | 9 | 3 |
| Fledgling success | Fledgling success | -0.077 | -0.356 | 0.201 | -0.851 | 0.697 | 0.586 | 20 | 12 |
| | Fledgling probability | | | | | | | | |
| | Pre-fledgling survival probability | | | | | | | | |
| | Fledgling success (fledglings/hatchings) | | | | | | | | |
| Offspring survival | Offspring survival | 0.025 | -0.264 | 0.314 | -0.753 | 0.803 | 0.866 | 15 | 9 |
| | Offspring mortality | | | | | | | | |
| Offspring recruit | Offspring recruitment | -0.160 | -0.496 | 0.176 | -0.956 | 0.636 | 0.349 | 7 | 3 |
| | Recruitment | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Offspring survival years | | | | | | | | |

CI: confidence interval; PI: prediction interval; L: lower; U: upper; k = number of effect sizes; N = number of studies.

**Supplementary Information S16:** Results of the phylogenetic multilevel meta-regression model for our entire dataset, with 'developmental model' (levels: altricial and precocial) as the only moderator. We included the same random effects as in the other meta-regression models: study ID, laboratory ID, population ID, species name, species phylogeny, and effect ID.

| | Mean | L 95% CI | U 95% CI | L 95% PI | U 95% PI | p-value | k | N |
|---|---|---|---|---|---|---|---|---|
| Altricial | -0.090 | -0.364 | 0.183 | -0.880 | 0.699 | 0.516 | 406 | 46 |
| Precocial | -0.018 | -0.391 | 0.355 | 0.847 | 0.811 | 0.924 | 37 | 11 |

CI: confidence interval; PI: prediction interval; L: lower; U: upper; k = number of effect sizes; N = number of studies.