The effects of sex hormones on immune function: a meta-analysis

Yong Zhi Foo¹,²,*, Shinichi Nakagawa³,⁴, Gillian Rhodes¹ and Leigh W. Simmons¹,²

¹ARC Centre of Excellence in Cognition and its Disorders, School of Psychology, University of Western Australia, 35 Stirling Hwy, Crawley, 6009, WA, Australia
²Centre for Evolutionary Biology & School of Animal Biology, University of Western Australia, 35 Stirling Hwy, Crawley, 6009, WA, Australia
³Department of Zoology, University of Otago, 340 Great King Street, Dunedin 9016, New Zealand
⁴Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, UNSW Sydney, NSW 2052, Australia

Running title: Sex hormones and immune function

*Author for correspondence (E-mail: 21193971@student.uwa.edu.au, Tel.: +61 (8) 6488 3573).

ABSTRACT
The effects of sex hormones on immune function have received much attention, especially following the proposal of the immunocompetence handicap hypothesis. Many studies, both experimental and correlational, have been conducted to test the relationship between immune function and the sex hormones testosterone in males and oestrogen in females. However, the results are mixed. We conducted four cross-species meta-analyses to investigate the relationship between sex hormones and immune function: (1) the effect of testosterone manipulation on immune function in males, (2) the correlation between circulating testosterone level and immune function in males, (3) the effect of oestrogen manipulation on immune function in females, and (4) the correlation between circulating oestrogen level and immune function in females. The results from the experimental studies showed that testosterone had a medium-sized immunosuppressive effect on immune function. The effect of oestrogen, on the other hand, depended on the immune measure used. Oestrogen suppressed cell-mediated immune function while reducing parasite loads. The overall correlation (meta-analytic relationship) between circulating sex hormone level and immune function was not statistically significant for either testosterone or oestrogen despite the power of meta-analysis. These results suggest that correlational studies have limited value for testing the effects of sex hormones on immune function. We found little evidence of publication bias in the four data sets using indirect tests. There was a weak and positive relationship between year of publication and effect size for experimental studies of testosterone that became non-significant after we controlled for castration and immune measure, suggesting that the temporal trend was due to changes in these moderators over time. Graphical analyses suggest that the temporal trend was due to an increased use of cytokine measures across time. We found substantial heterogeneity in effect sizes, except in correlational studies of testosterone,
even after we accounted for the relevant random and fixed factors. In conclusion, our results provide good evidence that testosterone suppresses immune function and that the effect of oestrogen varies depending on the immune measure used.

Key words: honest signals, immune function, immunocompetence handicap

hypothesis, meta-analysis, oestrogen, secondary sexual traits, sex hormones, sexual selection, testosterone, trade-offs.

CONTENTS

I. INTRODUCTION ........................................................................................................... 4

(1) The effect of testosterone on immune function in males ...................................... 6
(2) The effect of oestrogen on immune function in females ...................................... 7
(4) This study ............................................................................................................. 8

II. METHODS .................................................................................................................. 10

(1) Literature search .................................................................................................. 10
(2) Inclusion/exclusion criteria .................................................................................... 11
(3) Effect size extraction/calulation ............................................................................. 12
(4) Coding of papers .................................................................................................. 13
(5) Building phylogenies ............................................................................................ 15
(6) Meta-analyses ....................................................................................................... 15
(7) Publication bias ..................................................................................................... 18

III. RESULTS ................................................................................................................. 19

(1) Testosterone .......................................................................................................... 19
(a) Experimental studies ............................................................................................ 19
(b) Correlational studies ............................................................................................. 21
(2) Oestrogen .............................................................................................................. 22
(a) Experimental studies ............................................................................................ 22
(b) Correlational studies ............................................................................................. 24

IV. DISCUSSION ............................................................................................................. 26

(1) The relationship between testosterone and immune function in males ............ 26
(2) The relationship between oestrogen and immune function in females .......... 27
(3) Publication bias ..................................................................................................... 30
(4) Sex differences in immune function ..................................................................... 32
(5) Heterogeneity in the effects of sex hormones on immune function .............. 32
(6) Limitations and future directions .......................................................................... 34

V. CONCLUSIONS ......................................................................................................... 35
I. INTRODUCTION

There has been a long-standing interest in the effects of sex hormones on immune function (Ansar Ahmed, Penhale & Talal, 1985; Grossman, 1985; Schuurs & Verheul, 1990; Klein, 2004; Bouman, Heineman & Faas, 2005). Across species, females typically show stronger immune responses to parasitic challenges compared to males (Klein, 2004; Zuk & McKean, 1996). The prevalence and intensity of parasite infections also tend to be lower in females (Klein, 2004; Schuurs & Verheul, 1990). These sex differences were found even when experiments control carefully for parasite exposure (Daniels & Belosevic, 1994; Klein, Gamble & Nelson, 1999). These results suggest that sex differences in immune function are at least partly caused by physiological differences between the sexes. Among the physiological factors that differ between the sexes, sex hormones appear to be prime candidates as factors affecting immune function. The presence of testosterone and oestrogen receptors on various immune organs and immune cells suggests that sex hormones can influence the immune system directly (Alexander & Stimson, 1988; Cutolo et al., 1996; Danel et al., 1983; Roberts, Walker & Alexander, 2001; Wunderlich et al., 2002).

Furthermore, the removal of gonads, one of the main sources of sex hormones, can alter immune functioning (e.g. Kamis, Ahmad & Badrul-Munir, 1992; Rivero et al., 2002; also see review by Klein, 2004).
Much attention has been placed on the effects of sex hormones, especially following the proposal of the influential immunocompetence handicap hypothesis (ICHH; Folstad & Karter, 1992). According to indirect-benefit models of female mate choice, females can obtain genetic benefits for their offspring by favouring the most ornamented males (Fisher, 1958; Pomiankowski, 1988). However, if females are consistent in their preference, genetic variance in fitness-related traits should be lost, leading to the so-called ‘lek paradox’ (Borgia, 1979; Taylor & Williams, 1982; Kirkpatrick & Ryan, 1991). Hamilton & Zuk (1982) proposed that male sexual signals might reflect genes that code for superior parasite resistance. Based on this hypothesis, male genetic variation is maintained through a co-evolutionary arms race where genes that code for good health spread among host individuals while parasites evolve increased virulence in response to the improved immunity of hosts. In elaboration of the Hamilton–Zuk hypothesis, the ICHH suggested that sex hormones, in particular testosterone in males, provide the mechanistic link between sexual signals and genes that code for good health through their effects on both signal development and immune function.

Since its inception, there has been much debate concerning the ICHH, particularly the assumption that sex hormones affect immune function. Roberts, Buchanan & Evans (2004) published a meta-analysis examining the effect of testosterone on immune function using studies from evolutionary biology. They found little support for the hypothesis that testosterone suppresses immune function in males. In the decade since then, many more studies on the effects of testosterone have been published, and meta-analytic techniques have advanced considerably. Therefore, we provide an update on Roberts et al. (2004). In addition, we examine the
relationship between oestrogen and immune function in females, which has not been subjected to meta-analysis previously.

(1) The effect of testosterone on immune function in males

Males, in general, face intense mating competition (Trivers, 1972). As a result, males from many species often develop elaborate testosterone-dependent secondary sexual ornaments for the purpose of fighting for and attracting females (Andersson, 1994). The development of such ornaments is not without costs. Resources (e.g. energy) are limited. Therefore, natural selection is expected to favour an optimum allocation of energetic resources depending on the environmental situation, leading to trade-offs between different fitness components (Stearns, 1977, 1992). According to the ICHH, male sexual ornaments provide honest signals of the males’ immune function due to the trade-off between ornament development and immune health via the effects of testosterone. Ornament development triggers a down-regulation of immune function, which is effected through the immunosuppressive effect of testosterone. This immunosuppression makes it impossible for males of low genetic quality to develop exaggerated ornaments without compromising their health and potential future reproductive success. Thus, only high-genetic-quality males can afford to sustain the high levels of testosterone required for the development of elaborate ornaments.

Many empirical studies have therefore examined the relationship between sex hormones and immune function in vivo, across a wide variety of species and using both experimental and correlational designs. For testosterone, both positive (e.g. Bilbo & Nelson, 2001; Evans, Goldsmith & Norris, 2000; Morales-Montor et al., 2002) and negative (e.g. Alonso-Alvarez et al., 2007a; Belliure, Smith & Sorci, 2004;
Duckworth, Mendonça & Hill, 2001; Duffy et al., 2000) effects have been reported in experimental studies. Similarly, both positive and negative correlations between testosterone and immune function have been reported (e.g. Duffy & Ball, 2002; Peters, 2000; Rantala et al., 2012). In an immunocompetence handicap, individuals of different genetic quality are expected to have different optimal hormone levels due to differences in the marginal fitness benefit for every unit increase in a sexual signal (Getty, 2006). Hence, both positive and negative correlations between testosterone and immune function can be taken as supportive of an immunosuppressive effect of testosterone (Getty, 2006). A positive correlation might suggest that individuals with high genetic quality are able to cope with the reduced immune function that results from their high levels of testosterone. By contrast, a negative correlation might suggest that healthy individuals are trading their survival advantage for increased mating success. 

(2) The effect of oestrogen on immune function in females

In females, trade-offs occur between the allocation of resources to current reproduction and conserving resources for future reproduction (Stearns, 1992; Thornhill & Gangestad, 2008). Oestrogen is critically involved in a number of female reproductive functions, such as fertility and pregnancy (Ellison, 2001). Oestrogen is also implicated in the production of sexual signals in species such as humans (Homo sapiens) (Jasieńska et al., 2004; Moore et al., 2011; Smith et al., 2006) and red-sided garter snakes (Thamnophis sirtalis parietalis) (Parker & Mason, 2012). Therefore, the effect of oestrogen on immune function might be linked to trade-offs involving these reproductive functions. However, the predicted direction of the effect of oestrogen on immune function is unclear. Although an immune-enhancing effect seems intuitive
given that females tend to have better immune function (i.e. lower parasitism and
stronger immune responses) than males, we should note that females can have better
immune function than males even if oestrogen is immunosuppressive. As long as
testosterone suppresses immune function in males and the immunosuppressive effect
is stronger than that of oestrogen in females, we will see better immune function in
females.

Similar to the results of studies looking at the effect of testosterone on immune
function in males, the results of in vivo studies looking at the relationship between
oestrogen and immune function in females have been mixed. Both positive (e.g. Ádori
et al., 2010; Ding & Zhu, 2008; Zhou et al., 2011) and negative (e.g. Douin-Echinard
et al., 2011; Salem et al., 2000) effects have been reported for experimental studies.
Some researchers have observed that the effect of oestrogen appears to depend on the
immune component measured and the oestrogen dosage used (Cutolo et al., 1996;
Klein, 2004). Both positive and negative correlations between oestrogen level and
immune function have also been reported (e.g. Klein & Nelson, 1997; Vainikka et al.,
2004).

(4) This study

The mixed results in the literature make it difficult to assess the general effects
of sex hormones on immune function. In this study, we use meta-analysis to analyse
the results quantitatively from the literature. We provide an update on Roberts et al.
(2004) by including studies conducted since then. We also analyse correlations
between testosterone and immune function, which has not been done before. In
addition, we examine the effect of oestrogen on immune function and the correlation
between circulating oestrogen level and immune function in females by quantitatively
analysing the results from experimental and correlational studies, respectively, neither of which has been done previously. Notably, for all these analyses, we include studies from fields other than evolutionary biology, such as the biomedical sciences.

We conduct four phylogenetic meta-analyses (Verdú & Traveset, 2005; Hadfield & Nakagawa, 2010; Nakagawa & Santos, 2012) to address the following questions: (1) does experimental manipulation of testosterone affect immune function in males? (2) Are the levels of circulating testosterone correlated with immune function in males? (3) Does experimental manipulation of oestrogen affect immune function in females? (4) Are the levels of circulating oestrogen correlated with immune function in females?

We also conduct moderator analyses to investigate the factors that account for variation in effect sizes. First, we look at sample-related variables such as the mating system of the species and whether the individuals were sampled from natural or laboratory populations. Polygamous species face more intense mating competition compared to monogamous species (Andersson, 1994; Darwin, 1871). They tend to have higher sex hormone levels and rely more on sex-hormone-mediated traits for mating success (Andersson, 1994). Therefore, we expect the effect of sex hormones on immune function to be stronger in polygamous species.

We also look at natural versus laboratory populations. Some laboratory populations might have undergone artificial selection for traits that make the populations ideal for laboratory experiments, including traits that might be related to sex hormones, such as aggression. Therefore the results from laboratory populations might be different to those from natural populations.

Second, we look at immune-measure-related variables such as the immune measure used and whether the study measured baseline immunity or immune
reactivity to a pathogenic challenge. Different immune measures can be independent from each other (Adamo, 2004). Furthermore, the effect of oestrogen appears to depend on the immune component measured (Cutolo et al., 1996; Klein, 2004).

Therefore, we include immune measure as a moderator variable.

It has been suggested that measures of immune reactivity to pathogenic challenges are more valid measures of immune function than baseline measures because higher baseline immunity may indicate current infection status rather than actual immunocompetence (Demas et al., 2011; Norris & Evans, 2000). We therefore investigate whether effect sizes are different for baseline immunity versus immune reactivity.

Third, for experimental studies, we look at the dosage of the hormones used (physiological versus supraphysiological dosages) to ascertain whether the effects of sex hormones are biologically relevant or simply due to overdosing. We also look at whether steps were taken to control for endogenous production of sex hormones (i.e. castration for males and ovariectomy for females).

Finally, to validate the robustness of our results, we present results from analyses that provide indirect estimates of publication bias.

II. METHODS

(1) Literature search

We conducted a literature search between January 2013 and June 2013. We searched the online database, Web of Science, using the terms ‘immunocompetence handicap hypothesis’, ‘testosterone AND immun*’, ‘testosterone AND parasit*’, ‘estrogen AND immun*’, and ‘estrogen AND parasit*’. We also searched the Internet
via Google Scholar using similar terms. Since it was impossible to use truncations and wildcards in Google Scholar, we tried to use as many variants of a word as possible to maximise our search potential. For example, we used the terms ‘parasite’ and ‘parasitic’ when ‘parasit*’ was not possible. We included studies reported in Roberts et al. (2004). We also included relevant studies citing Roberts et al. (2004). Fig. 1 summarises the process and outcome of the literature search.

(2) Inclusion/exclusion criteria

Studies were included if they fulfilled the following inclusion/exclusion criteria: (1) Studies must have been *in vivo; in vitro* studies and studies with simulated data were not included. (2) Experimental studies must have manipulated hormone levels and measured immune function post-manipulation. (3) Correlational studies must have measured both circulating hormone levels and immune function. (4) The immune function measures used must have been physiological measures of immune parameters or measures of parasite loads. (5) The individuals tested in the studies must have been adults. This criterion was used to prevent any confounds due to age. (6) A paper must have contained extractable data (i.e. effect size and sample size values or statistics that allowed us to infer these values). For relevant papers that did not contain extractable data, we contacted the authors for their original data sets. These papers were excluded if we could not contact the authors (see Fig. 1 for further information). (7) Laboratory animals that had anomalous immune function due to artificial selection (e.g. certain strains such as MRL+/+ mice) were excluded from the meta-analyses. References for the included and excluded studies can be found in the main reference list. Reasons for exclusion of studies are given as online supporting
information in Table S1. The final data set, including citations to the included studies, is uploaded as part of the electronic supplementary material.

(3) Effect size extraction/calculation

We chose Pearson’s $r$ as our measure of effect size. However, $r$ is not suitable for parametric analyses because it is bounded at –1 and 1, not conforming to a normal distribution (Hedges & Olkin, 1985). Therefore, for the purpose of the statistical analyses, all extracted statistics were converted to Fisher’s $Z$ (hereafter termed $Zr$), which is normally distributed. All results were back-transformed to $r$ to facilitate interpretations.

For experimental studies, most statistics extracted were means and variance (or uncertainty) estimates (i.e. S.D. and S.E.). For studies that reported the means and variance estimates in the form of graphs, we extracted values using the software Graphclick (Arizona-Software 2008). For studies that took multiple measures of immune function across time, we took the means and variance estimates at the time point where both conditions showed peak immune function. If the peak immune function could not be determined or if the treatment and control groups peaked at different time points, we took the means and variance estimates at the time point where the difference in immune function between the treatment and control groups was the largest. We reasoned that such differences are most likely to be biologically significant. For multifactorial studies that contained a sex-hormone-only group and a control group, we focused on the difference between the two groups. For multifactorial studies that did not have a sex-hormone-only group and a control group, we collapsed the other factors and took the results from the main effect of sex
hormone. For studies that did not report the means and variance estimates, we extracted statistics such as proportion infected, *F* values, *t* values, and *P* values.

For correlational studies, most statistics extracted were correlation coefficients (i.e. Pearson’s *r*). For studies that did not report correlation coefficients, we extracted statistics such as *F* values, *P* values, Spearman’s *rho*, Kendall’s *tau*, *β* values, *R*² values, and χ².

All extracted statistics were converted to *Zr* based on the equations in Lipsey & Wilson (2001) using the online calculator on the Campbell Collaboration website. For experimental studies that had more than one dosage group, we combined the treatment groups into one effect size using the function to combine subgroups on the Campbell Collaboration website. We expect the effect estimates for parasite loads to be in the opposite direction to those for measures of immunity. Therefore, for the purpose of the meta-analysis, we reversed the sign of parasite load effect sizes to standardize the direction of all effect sizes. A positive effect size indicates one of the following: a stronger immune response, an increase in the baseline immune level, or a decrease in parasite load.

**(4) Coding of papers**

For each effect size, we recorded the species, sample size, and study identity. We also recorded the following moderator variables.

(1) Mating system: the species were classified into ‘monogamous’ *versus* ‘polygamous’ species based on published information (see Table S2). Monogamous and socially monogamous species, where individuals maintain a long-term pair bond while engaging in extra-pair copulations occasionally, were classified as ‘monogamous’. Polygynous, polyandrous, polygynandrous, or lekking species were
classified as ‘polygamous’. We could not find information about the mating systems of four species. For these species, we consulted the authors of the studies and sought their expert opinions. One species was classified this way. The mating systems of the other three species were treated as missing values because we were unable to contact the authors. Socially monogamous species such as *Passer domesticus, Sturnus vulgaris, and Homo sapiens* are most likely intermediate in terms of the level of sexual selection they experience. Therefore, to check the robustness of our results, we reclassified these three species as polygamous and re-ran those analyses that contained these three species, which include the experimental and correlational studies of testosterone. Our conclusions remained the same after this reclassification.

(2) Natural versus laboratory populations: species sampled from the wild were classified as ‘natural’ while laboratory strains of rats, mice, and guinea pigs were classified as ‘laboratory’.

(3) Immune measure: the immune measures were classified into ‘cell-mediated’, ‘cytokines’, ‘humoral-mediated’, ‘white blood cells’, and ‘parasite load’ measures. ‘Parasite load’ contained both ectoparasites and endoparasites. Our initial analyses showed that the results for both types of parasite loads were very similar in all four meta-analyses. The two types of parasites were therefore combined into one category.

(4) Immune challenge: the immune measures were classified into ‘baseline’ and ‘immune reactivity’ measures.

(5) Gonadectomy, i.e. castration for males and ovariectomy for females (experimental studies only): experimental studies were classified based on whether or not gonadectomy was performed.

(6) Dosage (experimental studies only): hormone dosages were classified into ‘physiological’ and ‘supraphysiological’ levels based on the interpretations of the
authors. Five testosterone and five oestrogen studies were treated as missing values because we could not find any information on the dosages.

(5) Building phylogenies

Using the statistical program R 3.0.3 (R Core Team, 2014), we created the phylogeny for each meta-analysis following Lim, Senior & Nakagawa (2014). We first created one main tree containing the species from all four meta-analyses using the Interactive Tree of Life online tree generator (iTOL) (http://itol.embl.de/), which generates trees based on data from the National Center for Biotechnology Information taxonomy database. Polytomies on the main tree were resolved using published phylogenies (Fabre et al., 2012; Leache, 2009; Mayer & Pavlicev, 2007; Pyron, Burbrink & Wiens, 2013). We then created the sub-trees for each meta-analysis by trimming the main tree, leaving only the species from that particular meta-analysis (see Figs S1–S4). Given the wide variety of species in our data sets, it was difficult to estimate the branch lengths accurately. Therefore, we only used the topology of the sub-trees (i.e. the evolutionary relationship among the species without branch-length estimates) for our meta-analyses. Using the compute.brlen command with the default setting of rho = 1 from the R package ape, we converted the sub-trees to an ultrametric form so that the phylogenetic correlation could be incorporated into our meta-analyses.

(6) Meta-analyses

All analyses were conducted using the statistical program R 3.0.3 (R Core Team, 2014). Using the metafor package (Viechtbauer, 2010), we ran multilevel meta-analyses using linear mixed models (Nakagawa & Santos, 2012; Viechtbauer,
Linear mixed models allow us to control for non-independence in the data arising due to multiple effect sizes originating from the same study, multiple effect sizes originating from the same species, and shared ancestry among species (i.e. phylogenetic relationship – species that are more closely related may show more similar effects compared to species that are less closely related, thus resulting in non-independence in the data structure), by including study identity, species identity, and phylogeny into our models as random factors.

For each meta-analysis, we first checked the statistical significance of the random variables study identity and species identity using likelihood ratio tests. Each random variable was first entered into an intercept-only model (i.e. meta-analysis) and tested for statistical significance. Both variables were then entered simultaneously into the intercept-only model to check whether each variable had a significant effect after accounting for the other. A random variable was included in subsequent models if it had a significant effect and had a significant effect over and above that of the other random variable. We then tested the overall effect size by running an intercept-only model using restricted maximum likelihood (REML) estimation with the selected random variable/s.

We also tested whether it was necessary to control for similarity between species due to common phylogenetic descent by including phylogeny into the intercept model as a random effect. If controlling for phylogeny influenced the magnitude and/or significance of our overall effect size, phylogeny was included as a random effect for all subsequent analyses. If not, subsequent analyses were run without phylogeny.

We computed the heterogeneity statistic $I^2$ by running an intercept-only model without any random effects using the rma function in metafor. The statistic $I^2$ tells us
the percentage of variability in the effect sizes that is not due to sampling error
(Higgins et al., 2003; Higgins & Thompson, 2002). If $I^2$ was moderate to large (i.e. > 50% according to Higgins et al., 2003; Higgins & Thompson, 2002), we proceeded to run moderator analyses.

Moderator analyses were conducted by running meta-regression models using the `rma.mv` function in `metafor`. We first ran single-factor models without the intercept using REML estimation by entering each moderator as a fixed factor together with the random factors to obtain the parameter estimates of each level in each factor after controlling for the random factors. We then ran an automated model selection (Burnham & Anderson, 2002; Grueber et al., 2011) using the package `MuMIn` (Barton, 2014) to identify the moderators that remained in the final model. The model selection was based on the Akaike Information Criterion with sample size correction (AICc; Anderson, 2008; Burnham & Anderson, 2002) obtained from maximum likelihood (ML) estimation. We ran the model selection using only the effect sizes that had no missing data to ensure that the AICc values of the different models were comparable (Nakagawa & Freckleton, 2011). We first generated all the possible models from the moderator variables in the data set. We then averaged the model coefficients (without shrinkage) of all models within two AICc units from the best model, indicated by having the lowest AICc unit. We tested the significance of the moderators that were retained in the final averaged models using the $Q$ test. The $Q$ test provides an omnibus test for each moderator. It is also more conservative compared to testing the $Z$ values derived from the model estimates. We interpreted the variables in the final averaged model using the parameter estimates obtained from the single-factor models.
We analysed the relationship between year of publication and effect size for potential time-lag bias. Time-lag bias is the tendency for some studies to be published faster than others depending on the direction and magnitude of their results, usually with studies having large effects being published first (Jennions & Møller, 2002). We first ran a single-factor meta-regression model with year of publication entered as a moderator together with the random factors using the rma.mv function in metafor. We then entered year of publication together with the other moderators into the AICc model selection to see whether it was retained in the final averaged model and, if it was, whether it significantly predicted effect size after controlling for the rest of the fixed and random factors.

We looked for potential missing effect sizes by running two funnel plot asymmetry analyses using the meta-analytic residuals (sensu Nakagawa & Santos, 2012) extracted from the final averaged model using the MCMCglmm function in the MCMCglmm package (Hadfield, 2010); note that the residuals incorporating random effects could only be extracted from the models using MCMCglmm but not metafor. We used 130,000 iterations, 100 thinning, 30,000 burn-in, and inverse gamma prior for all four residual extractions. We ran Egger’s regression test (Egger et al., 1997) on the residuals using the regtest function in metafor. Egger’s test regresses the standardised residuals on precision. Publication bias is indicated by an intercept that is significantly different from zero.

We ran a trim-and-fill analysis (Duval & Tweedie, 2000) on the residuals using the trimfill function in metafor to test for funnel plot asymmetry and identify missing studies. The analysis assumes that the funnel plot is symmetric and attempts to remove (trimming) the smaller studies that are causing asymmetry while filling the
distribution with missing studies to symmetrize the distribution. We also used the $I^2$ statistic reported in the trimfill function to look at the amount of heterogeneity left after accounting for the random and fixed factors that were retained in the final averaged model.

III. RESULTS

(1) Testosterone

The results (parameter estimates) of the meta-analytic and meta-regression models for both experimental and correlational studies are presented in Fig. 2 and Table 1.

(a) Experimental studies

Both study identity (likelihood ratio test: $\chi^2_1 = 486.17, P < 0.0001$) and species identity (likelihood ratio test: $\chi^2_1 = 110.83, P < 0.0001$) significantly improved the model when entered individually into the model, but species identity did not have a significant effect over and above the effect of study identity (likelihood ratio test: $\chi^2_1 = 0$) when both random variables were entered into the model simultaneously. Therefore, we only included study identity as the random variable in subsequent analyses. Overall, there was a medium significant immunosuppressive effect of testosterone after controlling for study identity ($r_{\text{overall}} = -0.28, 95\% \text{ CI} [-0.39, -0.17], P < 0.0001$; Fig. 2A). Controlling for similarity due to common phylogenetic descent did not have a significant effect on the effect size (likelihood ratio test: $\chi^2_1 = 0$). Therefore we ran all subsequent analyses without controlling for phylogeny.
There was large heterogeneity in this data set ($I^2 = 89.16\%$). Therefore we conducted moderator analyses. The final averaged model from the AICc model selection retained the factors castration and immune measure (see Table S3 for model results). There was a significant effect of castration ($Q_1 = 5.55, P = 0.02$). The immunosuppressive effect was stronger in castrated animals, which showed a significant medium-to-large negative effect ($r_{\text{castrated}} = -0.41$, 95% CI $[-0.55, -0.24]$, $P < 0.0001$), compared to non-castrated animals, which showed a small-to-medium negative effect, ($r_{\text{non-castrated}} = -0.22$, 95% CI $[-0.35, -0.07]$, $P = 0.004$) (Fig. 2A). The effect of immune measure was non-significant ($Q_4 = 6.59, P = 0.16$).

There was a small and significant positive relationship between year of publication and effect size in the single-factor model (slope estimate = 0.03, 95% CI [0.01, 0.04], $P = 0.001$) (Fig. 3). Year of publication was also retained in the final AICc model, but the relationship between year of publication and effect size became non-significant after controlling for immune measure and castration ($Q_1 = 3.65, P = 0.06$). Therefore, the temporal trend appears to be due to changes in moderators over time.

For the two funnel plot analyses, Egger’s regression test did not indicate significant asymmetry in the funnel plot of the residuals ($t_{154} = -0.83, P = 0.41$). The trim-and-fill analysis estimated a total of 34 effect sizes missing from the right side of the distribution and that the effect-size estimate should be adjusted to $r = -0.15$ (95% CI $[-0.26, -0.04]$) (Fig. 4A), indicating that the actual effect of testosterone might be smaller than our initial estimate of $r = -0.28$. The $I^2$ statistic indicated that considerable heterogeneity still remained in the residuals (81.11%), suggesting that the effect of testosterone might be moderated by other variables.
(b) Correlational studies

Both study identity (likelihood ratio test: $\chi^2_1 = 117.27, P < 0.0001$) and species identity (likelihood ratio test: $\chi^2_1 = 42.84, P < 0.0001$) significantly improved the model when entered individually into the model, but species identity did not have a significant effect over and above the effect of study identity (likelihood ratio test: $\chi^2_1 = 0$) when both random variables were entered into the model simultaneously.

Therefore, we only included study identity as the random variable in subsequent analyses. Overall, the correlation between circulating testosterone level and immune function after controlling for study identity was small and non-significant (Fig. 2B; Table 1). Controlling for similarity due to common phylogenetic descent did not have a significant effect on the effect size (likelihood ratio test: $\chi^2_1 = 0$). Therefore we ran all subsequent analyses without controlling for phylogeny.

There was large heterogeneity in this data set ($I^2 = 95.41\%$). Therefore we conducted moderator analyses. The final averaged AICc model retained the moderators immune measure, immune challenge, and mating system (see Table S4 for model results). There was a significant effect of immune measure ($Q_1 = 8.56, P = 0.04$). There was a medium positive relationship between cell-mediated immune function and testosterone that was significant ($r_{\text{cell-mediated}} = 0.26, 95\% \text{ CI}[0.007, 0.48], P = 0.04$) (Fig. 2B). The relationship between immune function and testosterone was non-significant for the other immune measures ($r_{\text{humoral-mediated}} = 0.06, 95\% \text{ CI}[-0.16, 0.27], P = 0.59$; $r_{\text{parasite load}} = -0.17, 95\% \text{ CI}[-0.50, 0.21], P = 0.38$; $r_{\text{white blood cells}} = 0.08, 95\% \text{ CI}[-0.14, 0.29], P = 0.47$). The effect of immune challenge ($Q_1 = 2.46, P = 0.12$) and mating system ($Q_1 = 0.40, P = 0.53$) were non-significant.

No indirect evidence of publication bias was detected. The relationship between year of publication and effect size was non-significant in the single factor model.
(slope estimate = –0.02, 95% CI [–0.05, 0.01], P = 0.17). Year of publication was retained in the final averaged AICc model, but the relationship between year and effect size remained non-significant after controlling for immune measure, immune challenge, and mating system (Q₁ = 2.62, P = 0.11). Egger’s regression test indicated no significant asymmetry in the funnel plot of the residuals (t₁₀₇ = 0.18, P = 0.85). Trim and fill analysis estimated no missing effect sizes (Fig. 4B). The I² statistic indicated that only a small amount of heterogeneity remained in the residuals (20.84%).

(2) Oestrogen

The results (parameter estimates) of the meta-analytic and meta-regression models for both experimental and correlational studies are shown in Fig. 5 and Table 2.

(a) Experimental studies

When entered individually into the model, both study identity (likelihood ratio test: χ²₁ = 911.63, P < 0.0001) and species identity (likelihood ratio test: χ²₁ = 11.11, P = 0.0009) significantly improved the model, but species identity did not have a significant effect over and above the effect of study identity (likelihood ratio test: χ²₁ = 0) when both random variables were entered into the model simultaneously. Therefore we only included study identity as the random variable in subsequent analyses. Overall, the effect of oestrogen on immune function after controlling for study identity was small and non-significant (Fig. 5A; Table 2). Controlling for similarity due to common phylogenetic descent did not have a significant effect on
There was large heterogeneity in this data set ($I^2 = 94.60\%$). Therefore, we conducted moderator analyses. The final averaged AICc model retained the following moderators: immune measure, immune challenge, ovariectomy, and dosage (see Table S5 for model results). There was a significant effect of immune measure ($Q_4 = 41.28, P < 0.001$). Oestrogen had a medium-to-large immunosuppressive effect on cell-mediated immune function that was significant ($r_{\text{cell-mediated}} = 0.41, 95\% \text{ CI } [-0.65, -0.09], P = 0.01$). The effects of oestrogen on parasite load, humoral-mediated immune function, and cytokine levels were in the opposite direction. Oestrogen had a medium-to-large immunoenhancing effect on parasite load (i.e. reducing parasite load) that was significant ($r_{\text{parasite load}} = 0.46, 95\% \text{ CI } [0.09, 0.72], P = 0.02$). Oestrogen also had a medium but non-significant immunoenhancing effect on humoral-mediated immune function and cytokine level ($r_{\text{humoral-mediated}} = 0.30, 95\% \text{ CI } [-0.02, 0.56], P = 0.07$; $r_{\text{cytokine}} = 0.29, 95\% \text{ CI } [-0.04, 0.56], P = 0.08$) (Fig. 5A; Table 2). There was a significant effect of immune challenge ($Q_1 = 40.61, P < 0.001$). Studies using measures of baseline immunity showed a large positive effect that was significant ($r_{\text{baseline immunity}} = 0.60, 95\% \text{ CI } [0.11, 0.85], P = 0.02$) while studies using measures of immune reactivity showed a small and non-significant effect ($r_{\text{immune reactivity}} = 0.10, 95\% \text{ CI } [-0.22, 0.40], P = 0.54$) (Fig. 5A; Table 2). There was a significant effect of dosage ($Q_1 = 10.36, P = 0.001$). Studies using supraphysiological oestrogen dosages showed a medium-to-large positive effect that was significant ($r_{\text{supraphysiological}} = 0.48, 95\% \text{ CI } [0.14, 0.72], P = 0.008$) while studies using physiological dosages showed a small and non-significant effect ($r_{\text{physiological}} = -0.05$, without controlling for phylogeny.
95% CI [-0.40, 0.31], P = 0.78) (Fig. 5A; Table 2). The effect of ovariectomy ($Q_1 = 1.82, P = 0.18$) was non-significant.

There was a small but significant positive relationship between year of publication and effect size in the single-factor model (estimate = 0.04, 95% CI [0.004, 0.08], $P = 0.03$), but the relationship was highly influenced by one large negative effect size in 1974 (Fig. 6). We re-ran the time-lag bias analysis a second time excluding that particular effect size. The relationship became non-significant (estimate = 0.02, 95% CI [-0.02, 0.06], $P = 0.30$). Year of publication was retained in the final averaged AICc model, but the relationship between year and effect size remained non-significant after controlling for immune measure, immune challenge, ovariectomy, and dosage ($Q_1 = 3.70, P = 0.054$). Therefore, there was no evidence of a significant temporal trend.

Egger’s regression test indicated no significant asymmetry in the funnel plot ($t_{126} = 0.51, P = 0.61$). The trim-and-fill analysis estimated a total of 31 effect sizes missing from the left side of the distribution. However, the missing effect sizes did not qualitatively influence the results. The overall effect size estimate remained small and non-significant after adjusting for the missing effect sizes ($-0.10$, 95% CI [$-0.41$, 0.18]) (Fig. 7A). $I^2$ indicated considerable heterogeneity in the residuals (89.23%).

(b) Correlational studies

Out of the 64 effect sizes in this data set, 60 effect sizes came from the same species and a single study while the remaining four effect sizes came from two species and three other studies. It was therefore impossible to distinguish between the variances of study identity and species identity. Therefore we only tested study identity as a random factor. Study identity significantly improved the model.
(likelihood ratio test: $\chi^2_1 = 7.99, P < 0.0001$) and was included as a random variable for subsequent analyses. Overall, there was a significant medium-to-large positive relationship between circulating oestrogen level and immune function after controlling for study identity ($r_{\text{overall}} = 0.43$, 95% CI [0.11, 0.66], $P = 0.01$) (Fig. 5B; Table 2). Controlling for similarity due to common phylogenetic descent did not have a significant effect on the effect size (likelihood ratio test: $\chi^2_1 = 0.57, P = 0.45$).

However, the overall effect size became non-significant after controlling for phylogeny ($r_{\text{phylogenetic}} = 0.42$, 95% CI [−0.17, 0.79], $P = 0.16$) (Fig. 5B; Table 2). We therefore ran all subsequent analyses with phylogeny included as a random variable.

There was moderate heterogeneity in this data set ($I^2 = 60.52\%$). Therefore, we conducted moderator analyses. We could not perform an AICc model selection because immune measure was confounded with immune challenge: all of the cell-mediated and cytokine measures were reactivity measures and all of the humoral-mediated and white blood cell measures were baseline measures. Therefore, we ran single-factor moderator analyses for the two factors separately. Neither the effects of immune measure ($Q_3 = 1.16, P = 0.76$) nor immune challenge ($Q_1 = 0.06, P = 0.80$) were significant (Fig. 5B; Table 2).

Year of publication was not significantly related to effect size in the single-factor model (estimate = −0.05, 95% CI [−0.13, 0.04], $P = 0.27$). Since none of the moderator effects were significant, we extracted the residuals from the intercept-only model with study identity and phylogeny entered as random factors. Egger’s regression test indicated no significant asymmetry in the residual funnel plot ($t_{62} = –0.19, P = 0.85$). Trim-and-fill analysis estimated no missing effect sizes (Fig. 7B). $I^2$ indicated moderate heterogeneity in the residuals (56.22%).
IV. DISCUSSION

(1) The relationship between testosterone and immune function in males

The results from the experimental studies support the hypothesis that testosterone suppresses immune function (Ansar Ahmed et al., 1985; Bouman et al., 2005; Grossman, 1985; Klein, 2004; Schuurs & Verheul, 1990). Overall, testosterone manipulation had a medium immunosuppressive effect \( r = -0.28 \) that was significant. Controlling for similarity due to common phylogenetic descent did not influence our results, which suggests that our results may be generally applicable across the species studied.

Castrated animals showed a stronger immunosuppressive effect than non-castrated animals. One possible explanation is that in the non-castrated animals, testosterone manipulation in the treatment group triggered a compensatory reduction in endogenous testosterone via a feedback loop, thus reducing the difference in testosterone levels between the treatment and control groups.

Our results provide important support for a critical assumption of the ICHH (Folstad & Karter, 1992), namely that testosterone suppresses immune function. According to the ICHH, testosterone-based male ornaments are honest signals of immune function because the immunosuppressive effect of testosterone makes it impossible for males with poor immune function to sustain high levels of testosterone for ornamentation. In their meta-analysis, Roberts et al. (2004) found that the overall effect of testosterone on immune function in males was non-significant and that a significant immunosuppressive effect was found only in reptiles and not in mammals or birds. They concluded that there was little support for the ICHH. Thus, our results differ from those of Roberts et al. (2004). They found an overall effect of \( d = -0.32 \), which transforms to \( r = -0.16 \). Our overall effect size \( r = -0.28 \) was almost twice
that. The difference in findings is probably due to the accumulation of studies since 2004 and the inclusion of studies from fields other than evolutionary biology, such as biomedical sciences, in our meta-analysis. Based on our results, there is good evidence that testosterone suppresses immune function.

We assumed that the effect of testosterone would be stronger in polygamous species compared to monogamous species. Polygamous species face more intense sexual selection and are thus expected to evolve stronger condition dependence of ornamentation (Andersson, 1994). However, mating system did not influence the size of the effect of testosterone on immune suppression. One possible reason for a lack of a mating-system effect might be that, while in polygamous systems a trade-off between testosterone and immune function is driven by male expenditure on ornaments, in monogamous systems the same trade-off is driven by male expenditure on parental care. Indeed, there is evidence that male parental care is associated with a reduction in circulating testosterone level (Gray, Yang & Pope, 2006; Wingfield et al., 1990). It is possible then that we did not observe a significant effect of mating system because trade-offs between mating and parental care may balance the overall reproductive effort among polygamous and monogamous species.

In contrast to the experimental studies, we did not find a significant overall correlation between testosterone and immune function. Our results are consistent with arguments that correlational studies are not ideal for testing the effect of testosterone on immune function (Getty, 2006).

(2) The relationship between oestrogen and immune function in females

The effect of oestrogen manipulation on immune function depended on the immune measure used. Oestrogen had a significant medium-to-large
immunosuppressive effect on cell-mediated immune function \( (r_{\text{cell-mediated}} = -0.41) \) but had a significant medium-to-large immunoenhancing effect on parasite loads \( (r_{\text{parasite load}} = 0.46) \). Oestrogen also had a medium but non-significant immunoenhancing effect on humoral-mediated immune function and cytokine level. Although these effects on humoral-mediated immune function and cytokine level were non-significant, the effect sizes \( (r_{\text{humoral-mediated}} = 0.30 \) and \( r_{\text{cytokine}} = 0.29) \) were medium in magnitude (Cohen, 1988). Moreover, the effect sizes were slightly larger than those typically found in biological studies, which range from \( r = 0.16 \) to 0.25 (Møller & Jennions, 2002). Therefore, the immunoenhancing effect of oestrogen on humoral-mediated immune function and cytokine level may prove to be biologically important despite the lack of statistical significance.

Our results are consistent with observations by researchers that oestrogen suppresses cell-mediated immune function and enhances humoral-mediated immune function (Cutolo et al., 1996; Klein, 2004). However, little is known about why oestrogen would have different effects on different immune components. This diversity in effects across immune components may reflect life-history trade-offs in females based on the costs and benefits of different immune components (Lee, 2006). Cell-mediated responses are energetically and nutritionally costly because they are associated with the activation of the systemic inflammatory response (Halloran et al., 1992; Janeway et al., 1999). In comparison, humoral-mediated responses are less costly because they are associated with the activation of the anti-inflammatory system (Janeway et al., 1999). Lee (2006) argued that because females invest more energy and resources in their offspring compared to males (Trivers, 1972), females are expected to adopt an immune profile that is less cell-mediated and more humoral-mediated. Doing so allows females to reduce the cost of maintaining a healthy
immune system while diverting energetic resources towards reproduction and parenting. Our results suggest that oestrogen, the major female sex hormone, influences the immune profile females adopt by suppressing cell-mediated immune function and enhancing humoral-mediated immune function.

We also found that the effect of oestrogen depended on whether measures of baseline immunity or immune reactivity were used. Given that measures of immune reactivity are considered more rigorous measures of immune function compared to baseline immunity (Demas et al., 2011; Norris & Evans, 2000), it is surprising to find that immune reactivity showed a small and non-significant effect ($r_{\text{immune reactivity}} = 0.10$) while baseline immunity showed a large positive and significant effect ($r_{\text{baseline immunity}} = 0.60$). However, it should be noted that the different measures of immune function were not equally distributed between immune reactivity and baseline immunity. Baseline immunity consisted of only white blood cell and humoral-mediated measures (19 effect sizes in total), while immune reactivity consisted of all five immune measures (127 effect sizes in total). As discussed above, the effect of oestrogen on immune function depends on the immune measure. Therefore, the non-significant results for immune reactivity may be due to the effects of the different immune measures cancelling each other out. On the other hand, the large effect size for baseline immunity may be due to a relatively small sample of effect sizes that consists of immune measures on which oestrogen had a positive effect. Nonetheless, it remains possible that the effect of oestrogen on immune function could differ for baseline and reactivity measures of immune function.

Besides moderators associated with the measurement of immune function, there was also an effect of dosage. Larger dosages showed larger effects. Specifically, supraphysiological dosages led to a significant medium-to-large immunoenhancing
effect ($r_{\text{supraphysiological}} = 0.48$) while physiological dosages showed a non-significant
effect that was close to zero ($r_{\text{physiological}} = -0.05$).

Overall, there was a medium-to-large positive relationship in correlational
studies between circulating oestrogen level and immune function ($r = 0.43$), but this
relationship became non-significant after we controlled for the similarity due to
common phylogenetic descent between species. This change was due largely to the
widening of the confidence intervals after accounting for phylogeny. This data set
consists of 64 effect sizes, 60 of which belonged to one single study of one species.

Therefore, the widening of the confidence interval probably reflects the over-
representation of a single species in the data set and the meta-analytic mean may not
be general. Furthermore, unlike the results from the experimental studies, the
correlation between oestrogen and immune function did not depend on immune
measure or immune challenge. Like the testosterone results, the oestrogen results
suggest that correlational designs are unsuitable for testing the effects of sex
hormones on immune function.

(3) Publication bias

We found a significant positive relationship between year of publication and
effect size for experimental studies of testosterone. Decreases in the magnitude of
effect sizes over time have been reported in numerous meta-analyses in evolutionary
biology (Jennions & Møller, 2002). It should, however, be noted that the analyses
used in this study are indirect tests of publication bias. A direct test of publication bias
requires a comparison of the effect sizes between published and unpublished studies
(Song et al., 2000; Møller & Jennions, 2001). A significant result from indirect tests
may not always indicate publication bias (Jennions & Møller, 2002; Jennions et al.,
For example, the temporal trend we found for experimental studies of testosterone appears to be due to changes in moderators across time. We found that the significant trend disappeared after we controlled for castration and immune measure. We graphically explored the moderators that changed across time by plotting the relationship between year of publication and effect size using different colours for each moderator level (Figs S5 and S6). We found that in the later years, more studies were conducted using cell-mediated and cytokine measures. The effect of testosterone on cytokine levels is smaller than the overall effect size (Fig. 2A; Table 1). The effect size for cell-mediated measures seems to be comparable to the overall effect size (Fig. 2A; Table 1). Therefore, the significant temporal trend seems likely to have been due to the increase in number of effect sizes assessing cytokines in recent years. Castration was fairly equally distributed across year of publication. Our finding suggests that the significant temporal trend was not due to a publication bias.

The trim-and-fill analysis also detected a substantial number of missing effect sizes in the same data set. Although the missing effect sizes did not change the results qualitatively, they did reduce the overall effect size by almost half from a medium effect size ($r = -0.28$) to a small effect size ($r = -0.15$). Even though the immunosuppressive effect remained significant, our result suggests that the effect of testosterone on immune function might not be as strong as initially indicated.

However, like the temporal trend findings, caution must be exercised when interpreting the results from the trim-and-fill analysis because the findings might reflect causes other than publication bias (Thornhill, Møller & Gangestad, 1999; Jennions et al., 2013). Heterogeneity in effect sizes can also lead to funnel plot asymmetry. We tried to control for the effects of heterogeneity by running the trim-
and-fill analysis on the residuals extracted from the final AICc model. However, we
found that the $I^2$ value for experimental studies remained high even after we
controlled for the random and moderator variables. Therefore, the funnel plot
asymmetry we observed in the experimental studies might have been caused by
unidentified moderators and not by publication bias. We therefore believe that the
initial estimate of $r = -0.28$ is more reflective of the actual effect size.

We detected little indirect evidence of publication bias for the other three data
sets. The trim-and-fill analysis estimated 32 missing effect sizes from the oestrogen
experimental studies, but the missing effects did not influence the results. Overall, our
results seem fairly robust to publication bias.

(4) Sex differences in immune function

Females tend to have better immune function compared to males (i.e. lower and
less-intense parasitism and stronger immune responses) (Klein, 2004; Schuurs &
Verheul, 1990; Zuk & McKean, 1996). Our results suggest that these sex differences
might be due to the combined effects of testosterone in males and oestrogen in
females. Our results also showed that the effect for oestrogen depends on the immune
measure. Therefore, it would be interesting to examine studies looking at sex
differences in immune function and test whether the effect sizes differ depending on
the immune measure.

(5) Heterogeneity in the effects of sex hormones on immune function

In meta-analysis, it is important to examine both the mean effect size and the
variance of the effect sizes (i.e. heterogeneity). The main tenet of life-history theory is
that trade-offs between fitness components occur due to limited resource availability.
One implication of this theory is that trade-offs between fitness components could vary across individuals. Therefore, one would predict that the effect of sex hormones, which mediate trade-offs between immune function and reproductive functions, would show significant heterogeneity. Indeed, we found large heterogeneity in the effect sizes across all four data sets. We ran moderator analyses to examine the factors that account for the variation in effect sizes. However, the heterogeneity remained moderate to large for three of the analyses, apart from correlational studies of testosterone, even after we accounted for the random and moderator effects from the final averaged AICc models.

Increasing evidence suggests that the effects of sex hormones on immune function can be dependent on individual condition. The amount of resources available to individuals varies substantially. It has been predicted that trade-offs between fitness components occur only when resources are limiting (van Noordwijk & de Jong, 1986; McDade, 2003). For example, the effect of testosterone on immune function in Sceloporus graciosus lizards depends on the quality of food available to the lizards (Ruiz et al., 2010). Testosterone enhanced immune function in lizards that were given extra vitamins on top of their usual diet, but decreased immune function in lizards that did not receive extra vitamins. The effect of testosterone might also depend on the effect of leptin, a hormone that functions as a signal of energetic resource level. In a study on zebra finches (Taeniopygia guttata), leptin increased immune function and prevented the immunosuppressive effect of testosterone (Alonso-Alvarez et al., 2007b). The effect of testosterone on immune function may also depend on stress levels [Rantala et al., 2012, but see Roberts et al. (2009) and Roberts et al. (2007a) for contradictory findings].
In relation to the issue of looking at variance, recent advances have applied meta-analytic techniques to analysing the variance instead of the mean data of experimental studies involving two groups (Nakagawa et al., 2015). For example, instead of asking whether testosterone suppresses immune function, we could ask whether testosterone increases or decreases the variance in immune function across individuals relative to controls. We did not run such analyses because the theoretical predictions were focused on the mean effect and not the variance.

(6) Limitations and future directions

Sex hormone levels change across time in response to life-history changes. For example, testosterone in males peaks during the breeding season and drops when the breeding season ends (Wingfield et al., 1990; Nelson, 2005). Similarly, oestrogen in females varies across the fertility cycle (Abraham et al., 1972). One might wonder how relevant the results in this meta-analysis are for understanding the effects of sex hormones during different life-history stages. In this review, for the testosterone studies, all studies except five that were unclassified reported using physiological dosages. For the oestrogen studies, we found a significant effect of immune measure even after controlling for the effect of dosage. Therefore, we were able to conclude that our results were not just an artefact of using dosages that were in excess of what is normally found in the body. However, we were unable to look at the seasonal or life-history relevance of the dosage levels because most studies did not provide such information. Future studies examining the effects of sex hormones in relation to different life-history stages will provide us with a better understanding of the effects of sex hormones on the immune system.
The studies reviewed herein have focused on the strength of the immune response. Navarro et al. (2003) found a positive correlation between immune-response strength and latency to maximum immune response in *Passer domesticus*. Their results suggest a trade-off between the strength and rapidity of the immune response. Therefore, future studies should measure both the strength and time course of the immune response to gain a better picture of the effects of sex hormones on immune function.

V. CONCLUSIONS

(1) We found meta-analytic evidence that testosterone has a medium-sized suppressive effect on immune function. This effect was generalizable across the species studied. Castrated animals showed a greater immunosuppressive effect than non-castrated animals, but the immunosuppressive effect was significant in both cases. Our overall effect size for experimental studies of testosterone was almost twice that of a previous meta-analysis (Roberts et al., 2004).

(2) We also found meta-analytic evidence that oestrogen has a medium-to-large suppressive effect on cell-mediated immune function while having a medium-to-large effect in reducing parasite loads and a medium but non-significant enhancing effect on humoral-mediated immune function and cytokine level. Oestrogen also had a significant immune-enhancing effect in studies using supraphysiological dosages and studies using baseline measures of immune function.

(3) When effect sizes were derived from correlational studies, the relationships between circulating sex hormone levels and immune function measurements were small and non-significant for both testosterone and oestrogen, suggesting that
correlational studies are unsuitable for testing the effects of sex hormones on immune function. Thus, an experimental approach is imperative to study the effects of sex hormones on immune function.

(4) We found little evidence of publication bias using indirect tests. There was a small and positive relationship between year of publication and effect size for experimental studies of testosterone that became non-significant after we controlled for castration and immune measure, suggesting that the temporal trend was due to changes in moderators over time. The trim-and-fill analysis for experimental studies of testosterone estimated a total of 34 missing effect sizes and that the immunosuppressive effect of testosterone should be reduced from −0.28 to −0.15. However, due to the substantial heterogeneity in the residuals after accounting for the random and fixed factors, we cannot rule out the possibility that the asymmetry in the funnel plot was due to heterogeneity. Overall, our results seem to be fairly robust to publication bias.

(5) We found substantial heterogeneity in the effect sizes for all four meta-analyses. The amount of heterogeneity in three of the meta-analyses, apart from correlational studies of testosterone, remained substantial even after we accounted for the relevant random and fixed factors, suggesting that there are other factors that moderate the effects of sex hormones on immune function.

VI. ACKNOWLEDGEMENTS

We would like to thank L. Lagisz for her help with creating the phylogenies. This work was supported by the ARC Centre of Excellence in Cognition and its Disorders (CE110001021), ARC Professorial Fellowships to L.W.S. (DP110104594) and G.R. (DP0877379) and an ARC Discovery Outstanding Researcher Award to G.R.
(DP130102300). S.N. was supported by a Rutherford Discovery Fellowship, New Zealand, and an ARC Future Fellowship (FT130100268).

VII. REFERENCES


944 immunity in two species of Mediterranean lacertid lizards. Journal of
946 BESTER-MEREDITH, J. K., & MARLER, C. A. (2001). Vasopressin and aggression in
947 cross-fostered californian mice (Peromyscus californicus) and white-footed mice
948 (Peromyscus leucopus). Hormones and Behavior 40, 51–64.
949 BILBO, S. D. & NELSON, R. J. (2001). Sex steroid hormones enhance immune function
950 in male and female Siberian hamsters. American Journal of Physiology:
951 Regulatory, Integrative and Comparative Physiology 280, R207–R213.
953 efficiency of bulls during pasture mating. Theriogenology 6, 393–401.
955 Selection and Reproductive Competition in Insects (eds M. S. Blum & N. A.
957 BORTOLOTTI, G. R., MOUGEOT, F., MARTINEZ-PADILLA, J., WEBSTER, L. M. I. &
958 PIERTNEY, S. B. (2009). Physiological stress mediates the honesty of social
962 BRYJA, J., PATZENHAUEROVÁ, H., ALBRECHT, T., MoŠANSKY, L., STANKO, M., &
963 STOPKA, P. (2008). Varying levels of female promiscuity in four Apodemus mice
966 dominance signalling and immunosuppression in the house sparrow, Passer
969 Inference: A Practical Information-Theoretic Approach. NY: Springer.
970 BUTTENM, W. A. & ASTHEIMER, L. B. (2000). Testosterone does not affect basal
971 metabolic rate or blood parasite load in captive male White-plumed Honeyeaters
973 CALIPPE, B., DOUIN-ECHINARD, V., LAFFARGUE, M., LAURELL, H., RANA-POUSSINE,
975 Chronic estradiol administration in vivo promotes the proinflammatoty response
976 of macrophages to TLR4 activation: Involvement of the phosphatidylinositol 3-
977 kinase pathway. The Journal of Immunology 180, 7980-7988.
981 encoding of oestriadiol suppression of delayed-type hypersensitivity in (NZB X
982 NZW) F1 mice. Immunology 73, 186-190.
984 mediated and testosterone-mediated effects on the immune system in normal and
985 autoimmune mice are genetically linked and inherited as dominant traits. Immunology
987 CASAGRANDE, S., DIJKSTRA, C., TAGLIAVINI, J., GOERLICH, V. C., & GROOTHUIS, T. G.
988 G. (2010). Differential effects of testosterone, dihydrotestosterone and estradiol
989 on carotenoid deposition in an avian sexually selected signal. Journal of
Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology 197, 1–13.


Testosterone is positively related to the output of nematode eggs in male Alpine ibex (Capra ibex) faeces. Evolutionary Ecology Research 9, 1277-1292.


Functionally opposing effects of testosterone on two different types of parasite:


4. Experimental study investigating the relationship between testosterone and parasite infection in the wild. Evolutionary Biology 26, 123-133.


7. Raloxifene and estradiol-mediated effects on uterus, bone and B lymphocytes in mice. Journal of Endocrinology 175, 319-327.


10. Evolutionary Biology 12, 88. doi:10.1186/1471-2148-12-88


Inconsistency in meta-analysis.

Differential association between circulating testosterone and infection risk by several viruses in natural cat populations: A behavioural-mediated effect?

Parasitology 140, 521-529.


Proceedings of the Royal Society  

sheep, (2005). Fecal counts of lungworm larvae and reproductive effort in bighorn  

Behaviour are related to a male ornamental trait in the common yellowthroat.  

of Experimental Biology  

damage to the immunocompetence handicap hypothesis: Unraveling direct and indirect  

International Journal for Parasitology  

growth of  

Zeitschrift Für Parasitenkunde  

growth of  

International Journal of Parasitology  

sterone are more immunocompetent.  

Applied Animal Behaviour Science  

oral testosterone is related to male success in the Azorean rock-pool blenny: The  

Biology  

American Naturalist  

max testosterone in intact and orchiectomized mice.  

Biology and Sociobiology  

Testosterone, ticks and travels: A test of the immunocompetence-handicap hypothesis in free-ranging male sand lizards.  

Biological Sciences  

Physiological and Reproductive Ecology  

Biology  

and travels: A test of the immunocompetence handicap hypothesis: Unraveling direct and indirect  

The American Naturalist  

activation of female sex pheromone in male red-sided garter snakes.  

Journal of Experimental Biology  


Animal Behaviour  


Proceedings of the Royal Society of London B: Biological Sciences  

Peters, A. (2000). Testosterone treatment is immunosuppressive in superb fairy-wrens, yet free-living males with high testosterone are more immunocompetent.
surface TLR4 expression on murine macrophages and regulate sepsis

macrophage expression in the mouse of toll-like receptor 4, a trigger for

Rivero, J. C., Ioue, Y., Murakami, N. & Horii, Y. (2002). Androgen- and
estrogen-dependent sex differences in host resistance to *Strongyloides


dependent manner? An experimental test in blue tits. *The Journal of


immunocompetence handicap hypothesis: A review of the evidence. *Animal

(2009). The effects of testosterone on immune function in quail selected for

role of testosterone in bib size determination in the male house sparrow *Passer

(2007b). Physiological, morphological and behavioural effects of selecting zebra
finches for divergent levels of corticosterone. *The Journal of Experimental

of testosterone and corticosterone on immunocompetence in the zebra finch.
*Hormones and Behavior* **51**, 126-134.

Robinson, D. P., Lorenzo, M. E., Jian, W. & Klein, S. L. (2011). Elevated 17β-
estriadiol protects females from influenza A virus pathogenesis by suppressing

male reproductive tactics and the immunocompetence handicap in the Azorean
rock-pool blenny, *Parablennius parvicornis*. *Proceedings of the Royal Society of

immune response correlates with decreased androgen levels in male peafowl,

Ros, A. F. H., Ferreira, C., Santos, R. S. & Oliveira, R. F. (2006b). Regulation of
immunocompetence by different androgen metabolites in a blenny with


VIII. SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.

Table S1. Excluded studies with reasons for this exclusion.

Table S2. References used for mating system classification.

Fig. S1. Phylogeny for experimental studies of testosterone.

Fig. S2. Phylogeny for correlational studies of testosterone.

Fig. S3. Phylogeny for experimental studies of oestrogen.

Fig. S4. Phylogeny for correlational studies of oestrogen.

Table S3. Details of the summary statistics provided by the final averaged model from the AICc model selection for experimental studies of testosterone, including the Q statistics for the test of main effect significance for each moderator.

Table S4. Details of the summary statistics provided by the final averaged model from the AICc model selection for correlational studies of testosterone, including the Q statistics for the test of main effect significance for each moderator.
Table S5. Details of the summary statistics provided by the final averaged model from the AICc model selection for experimental studies of oestrogen, including the Q statistics for the test of main effect significance for each moderator.

Fig. S5. The relationship between year of publication and effect size for experimental studies of testosterone. Effect sizes are separated by castration: black indicates no castration; blue indicates castration.

Fig. S6. The relationship between year of publication and effect size for experimental studies of testosterone. Effect sizes are separated by immune measure: blue indicates cell-mediated measures; light blue indicates cytokine levels; black indicates humoral-mediated measures, grey indicates parasite loads; purple indicates white blood cell counts.
Table 1. Parameter estimates and \( P \)-values for the effect of testosterone manipulation on immune function and the correlation between circulating testosterone level and immune function in males. \( M \) is the mean effect size. CI.lb and CI.ub are the lower and upper bounds of the 95% confidence interval respectively. * indicates moderators that were retained in final AICc models.

<table>
<thead>
<tr>
<th></th>
<th>Experimental studies</th>
<th>Correlational studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>CI.lb</td>
</tr>
<tr>
<td>Meta-analytic mean</td>
<td>-0.28</td>
<td>-0.39</td>
</tr>
<tr>
<td>Phylogenetic mean</td>
<td>-0.28</td>
<td>-0.39</td>
</tr>
<tr>
<td>Mating system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polygamous</td>
<td>-0.30</td>
<td>-0.43</td>
</tr>
<tr>
<td>Monogamous</td>
<td>-0.17</td>
<td>-0.33</td>
</tr>
<tr>
<td>Natural vs Lab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural</td>
<td>-0.25</td>
<td>-0.37</td>
</tr>
<tr>
<td>Lab</td>
<td>-0.41</td>
<td>-0.60</td>
</tr>
<tr>
<td>Immune measure*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell-mediated</td>
<td>-0.29</td>
<td>-0.42</td>
</tr>
<tr>
<td>Cytokines</td>
<td>-0.03</td>
<td>-0.34</td>
</tr>
<tr>
<td>Humoral-mediated</td>
<td>-0.32</td>
<td>-0.45</td>
</tr>
<tr>
<td>Parasite load</td>
<td>-0.33</td>
<td>-0.45</td>
</tr>
<tr>
<td>White blood cells</td>
<td>-0.24</td>
<td>-0.40</td>
</tr>
<tr>
<td>Immune-challenged</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.29</td>
<td>-0.40</td>
</tr>
<tr>
<td>No</td>
<td>-0.28</td>
<td>-0.42</td>
</tr>
<tr>
<td>Castrated*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.41</td>
<td>-0.55</td>
</tr>
<tr>
<td>No</td>
<td>-0.22</td>
<td>-0.35</td>
</tr>
</tbody>
</table>
Table 2. Parameter estimates for the effect of oestrogen manipulation on immune function and the correlation between circulating oestrogen level and immune function in females. \( M \) is the mean effect size. CI.lb and CI.ub are the lower and upper bounds of the 95% confidence interval, respectively. * indicates moderators that were retained in final AICc models.

<table>
<thead>
<tr>
<th></th>
<th>Experimental studies</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Correlational studies</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>CI.lb</td>
<td>CI.ub</td>
<td>( P )</td>
<td>( M )</td>
<td>CI.lb</td>
<td>CI.ub</td>
<td>( P )</td>
<td></td>
</tr>
<tr>
<td><strong>Meta-analytic mean</strong></td>
<td>0.16</td>
<td>−0.15</td>
<td>0.44</td>
<td>0.30</td>
<td>0.43</td>
<td>0.11</td>
<td>0.66</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td><strong>Phylogenetic mean</strong></td>
<td>0.16</td>
<td>−0.15</td>
<td>0.44</td>
<td>0.30</td>
<td>0.42</td>
<td>−0.17</td>
<td>0.79</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td><strong>Immune measure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell-mediated</td>
<td>−0.41</td>
<td>−0.65</td>
<td>−0.09</td>
<td>0.01</td>
<td>Cell-mediated</td>
<td>0.42</td>
<td>−0.18</td>
<td>0.79</td>
<td>0.16</td>
</tr>
<tr>
<td>Cytokines</td>
<td>0.29</td>
<td>−0.04</td>
<td>0.56</td>
<td>0.08</td>
<td>Cytokines</td>
<td>0.48</td>
<td>−0.45</td>
<td>0.91</td>
<td>0.31</td>
</tr>
<tr>
<td>Humoral-mediated</td>
<td>0.30</td>
<td>−0.02</td>
<td>0.56</td>
<td>0.07</td>
<td>Humoral-mediated</td>
<td>0.5</td>
<td>−0.08</td>
<td>0.83</td>
<td>0.08</td>
</tr>
<tr>
<td>Parasite load</td>
<td>0.46</td>
<td>0.09</td>
<td>0.72</td>
<td>0.02</td>
<td>White blood cells</td>
<td>0.42</td>
<td>−0.18</td>
<td>0.79</td>
<td>0.16</td>
</tr>
<tr>
<td>White blood cells</td>
<td>0.10</td>
<td>−0.27</td>
<td>0.44</td>
<td>0.61</td>
<td>Immune-challenged</td>
<td>Yes</td>
<td>0.42</td>
<td>−0.21</td>
<td>0.80</td>
</tr>
<tr>
<td><strong>Immune-challenged</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>0.43</td>
<td>−0.18</td>
<td>0.81</td>
</tr>
<tr>
<td>Yes</td>
<td>0.10</td>
<td>−0.22</td>
<td>0.40</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.60</td>
<td>0.11</td>
<td>0.85</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ovariectomized</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.09</td>
<td>−0.24</td>
<td>0.40</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.41</td>
<td>−0.05</td>
<td>0.72</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dosage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physiological</td>
<td>−0.05</td>
<td>−0.40</td>
<td>0.31</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supraphysiological</td>
<td>0.48</td>
<td>0.14</td>
<td>0.72</td>
<td>0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. Prisma flow chart (Moher et al., 2009) depicting the process and outcome of the literature search.
Fig. 2. Parameter estimates for (A) studies investigating the effect of testosterone manipulation on immune function and (B) studies investigating the correlation between circulating testosterone level and immune function in males. Diamonds represent the mean and error bars represent 95% confidence intervals. \( N \) refers to the number of effect sizes. White, light-grey, medium-grey and dark-grey spaces represent the regions for small, small-to-medium, medium-to-large, and large effect sizes, respectively, based on Cohen (1988).
Fig. 3. Relationship between effect size and year of publication (indicated by the bold line) for studies investigating the effect of testosterone manipulation on immune function. Size of each point indicates the sample size of that effect size.
Fig. 4. Funnel plot of the residuals plotted against precision for (A) studies investigating the effect of testosterone manipulation on immune function in males and (B) studies investigating the correlation between circulating testosterone level and immune function in males. Filled circles are actual effect sizes and empty circles are missing effect sizes estimated from the trim-and-fill analyses. Dashed lines indicate the zero line.
Fig. 5. Parameter estimates for (A) studies investigating the effect of oestrogen manipulation on immune function and (B) studies investigating the correlation between circulating oestrogen level and immune function in females. Diamonds represent the mean and error bars represent 95% confidence intervals. $N$ refers to the number of effect sizes. White, light-grey, medium-grey and dark-grey spaces represent the regions for small, small-to-medium, medium-to-large, and large effect sizes, respectively, based on Cohen (1988).
**Fig. 6.** Relationship between effect size and year of publication for studies investigating the effect of oestrogen manipulation on immune function. Solid line indicates the relationship when all effect sizes were included. Dashed line indicates the non-significant relationship after removing the effect size from 1974 that appeared to be driving the significant relationship between effect size and year of publication. Size of each point indicates the sample size of that effect size.
Fig. 7. Funnel plot of the residuals plotted against precision for (A) studies investigating the effect of oestrogen manipulation on immune function in females and (B) studies investigating the correlation between circulating oestrogen level and immune function in females. Filled circles are actual effect sizes and empty circles are missing effect sizes estimated from the trim-and-fill analyses. Dashed lines indicate the zero line.