Hormones and Behavior xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Hormones and Behavior

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

Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis

Shawn N. Geniole [a,b], Brian M. Bird [c,e], Erika L. Ruddick [d], Justin M. Carré [d,*]

[a] Department of Psychology, Brock University, St. Catharines, Ontario, Canada
[b] Neuropsychopharmacology and Biopsychology Unit, Faculty of Psychology, University of Vienna, Vienna, Austria
[c] Department of Psychology, Laurentian University, Sudbury, Ontario, Canada
[d] Department of Psychology, Nipissing University, North Bay, Ontario, Canada
[e] Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada

ARTICLE INFO

Article history:
Received 12 January 2016
Revised 3 October 2016
Accepted 5 October 2016
Available online xxxx

Keywords:
Testosterone
Meta-analysis
Competition
Biosocial model of status
Challenge hypothesis

ABSTRACT

A contribution to a special issue on Hormones and Human Competition. Since Archer's (2006) influential meta-analysis, there has been a major increase in the number of studies investigating the effect of competition outcome on testosterone reactivity patterns in humans. Despite this increased research output, there remains debate as to whether competition outcome modulates testosterone concentrations. The present paper examines this question using a meta-analytic approach including papers published over the last 35 years. Moreover, it provides the first meta-analytic estimate of the effect of competition outcome on testosterone concentrations in women. Results from a meta-analysis involving 60 effect sizes and >2500 participants indicated that winners of a competition demonstrated a larger increase in testosterone concentrations relative to losers ($D = 0.20$)—an effect that was highly heterogeneous. This 'winner-loser' effect was most robust in studies conducted outside the lab (e.g., in sport venues) ($D = 0.43$); for studies conducted in the lab, the effect of competition outcome on testosterone reactivity patterns was relatively weak ($D = 0.08$), and only found in studies of men ($D = 0.15$; in women: $D = -0.04$). Further, the 'winner-loser' effect was stronger among studies in which pre-competition testosterone was sampled earlier than ($D = 0.38$, after trim and fill correction) rather than within ($D = 0.09$) 10 min of the start of the competition. Therefore, these results also provide important insight regarding study design and methodology, and will be a valuable resource for researchers conducting subsequent studies on the 'winner loser' effect.

Contents

* Corresponding author.
E-mail address: justinca@nipissingu.ca (J.M. Carré).

http://dx.doi.org/10.1016/j.yhbeh.2016.10.002
0018-506X/

1. Introduction

Competitive behavior is ubiquitous in the animal kingdom, and the outcome of competition (victory vs. defeat) may have important fitness consequences. For instance, success in competition enables the animal to obtain preferential access to valued resources such as food, shelter, and mating opportunities. Testosterone, a steroid hormone produced primarily by the gonads, is believed to play a key role in modulating physiological and behavioral processes critical to survival and reproduction (Ketterson & Nolan, 1992). Notably, testosterone concentrations are not static, but rather, fluctuate rapidly within the context of social interactions (Archer, 2006; Oliveira, 2009; Wingfield et al., 1990). In this paper, we first provide a brief review of the literature on testosterone and human competition, with an introduction to the Challenge Hypothesis and Biosocial Model of Status—two of the main theoretical models guiding current research on testosterone and competitive behavior. Next, we perform formal meta-analyses examining the following questions: (1) Does the outcome of competition modulate testosterone reactivity patterns in humans? (2) Is the effect of competition outcome on testosterone reactivity patterns similar in men and women? In addition to these key questions, we also investigate other factors that may moderate the effect of competition on testosterone reactivity patterns (e.g., time of day, type of competition, time of sampling).

1.1. Challenge hypothesis

Wingfield et al. (1990) developed the 'Challenge Hypothesis' in an attempt to explain intra- and inter-species variation in testosterone concentrations in birds. In this model, testosterone concentrations fluctuate around three levels during the season: Level A = low level baseline; Level B = breeding baseline; and Level C = physiological maximum. In monogamous male birds that provide paternal care, testosterone concentrations are relatively low during the non-breeding season (Level A) and increase (Level B) at the onset of the breeding season, a change functionally linked to the initiation of spermatogenesis, expression of secondary sex characteristics, and the full display of male reproductive behavior. Testosterone concentrations increase further (Level C) during male-to-male competitive interactions, changes posited to facilitate territorial and aggressive behavior. At the end of the mating season and with the corresponding decrease in male-to-male competition, testosterone levels return to the non-breeding baseline (Level A). Wingfield et al. (2001) have proposed that the costs associated with maintaining elevated testosterone concentrations throughout the season (e.g., decreased paternal care, increased risk for physical injury/death, depressed immune function, increased energetic demands) may have led to a highly flexible endocrine system capable of rapidly modulating testosterone concentrations in response to changes in the social environment. The Challenge Hypothesis was originally proposed to account for the trade-off between mating and parental efforts in birds. Nevertheless, support for the basic tenets of the model has now been obtained in numerous taxa including fish (Oliveira, 2009), non-human primates (Bernstein et al., 1974; Sobolewski et al., 2013), insects (Tibbetts & Crocker, 2014), and humans (Archer, 2006).

1.2. Biosocial model of status

The Biosocial Model of Status (Mazur, 1976; 1985) is another theoretical model that has mainly been adopted by researchers studying the social endocrinology of human competitive behavior. The model posits that fluctuations in testosterone concentrations will be dependent on the outcome of competitions, increasing after victories and decreasing after defeats. Although this model has mainly been used in the context of human competition, the primary predictions of the model are based on findings from research involving male rhesus monkeys. Specifically, researchers reported that male rhesus monkeys emerging victorious after aggressive interactions experienced marked elevations in testosterone, whereas losers experienced a decrease in testosterone (Rose et al., 1972; 1975). Mazur and Lamb (1980) were the first to extend these findings to humans, reporting that male tennis players experienced a rise in serum testosterone concentrations after a decisive victory compared to a defeat. Shortly after this publication, another small-scale study examined serum testosterone reactivity patterns in male varsity wrestlers and found that winners had elevated post-competition testosterone concentrations relative to losers (Elias, 1981). These studies provided initial support for the idea that competition outcome plays a key role in modulating testosterone reactivity patterns in humans. However, one limitation of this work is that physical exercise can potentiate testosterone release independent of competition (see Vingren et al., 2010, for review), and thus, differences in endocrine reactivity patterns between winners and losers may be due to outcome-related differences in physical activity expended. One clear way to address this potential confound is to examine testosterone reactivity patterns in competitive interactions that are performed without physical exertion. In the first such study, Gladue et al. (1989) measured salivary testosterone concentrations in men randomly assigned to win or lose in a laboratory reaction time task. Results indicated that male winners had elevated testosterone concentrations relative to losers, suggesting that physical exertion cannot fully account for the divergent testosterone response patterns observed in earlier studies.

1.3. Archer's (2006) meta-analysis on the effects of competition on testosterone

A decade ago, Archer (2006) conducted a meta-analysis of studies examining the effects of competition outcome on testosterone reactivity patterns ($k = 12$ effect sizes). Specifically, Archer (2006) evaluated the 'winner-loser' effect—that is, the extent to which winners and losers

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

differed in terms of testosterone responses to competition. Results indicated that winners had elevated testosterone concentrations relative to losers ($D = 0.31$). Further analyses indicated that the 'winner-loser' effect was influenced by the type of competition performed. Specifically, winners had elevated testosterone concentrations compared to losers for studies involving contrived laboratory competitions ($D = 0.38$), but not studies involving sport competitions ($D = 0.05$).

Archer's (2006) meta-analysis has been highly cited (832 citations according to Google Scholar, October 11, 2016), which speaks to the increased interest in examining the effect of competition on testosterone reactivity patterns in humans. At the time of Archer's (2006) publication, only one of the studies involved women (Bateup et al., 2002). Over the past decade, however, dozens of additional studies have been conducted, several of which involved examining women's testosterone responses to competition. In this paper, we provide an updated meta-analysis on the effect of competition outcome on testosterone reactivity patterns in men and women using a relatively large number of effect sizes ($k = 62$; 45 for men and 17 for women).

2. Methods

Studies were identified for inclusion by first obtaining manuscripts analyzed in Archer's (2006) meta-analysis and included in a recent review paper by Carré and Olmstead (2015). In addition, articles were identified by a search in Google Scholar from 2005 to 2015 by using the search terms "competition testosterone humans" and "competition and testosterone change humans", and in the PSYC INFO database by using the combinations of (1) Testosterone AND winner (and loser), and (2) Testosterone AND winning (and losing) AND competition. These combined methods produced 57 manuscripts relevant to our current study. Four additional manuscripts not captured by our search terms were identified either in the references of another article (Edwards & Casto, 2013; Edwards & Kurlander, 2010) or by one of the anonymous reviewers of this manuscript (Jones & Josephs, 2006; Oliveira et al., 2014), and added to this list. Excluded from our analyses were studies from which we were unable to extract the effect sizes (Bateup et al., 2002; Mazur & Lamb, 1980; Oxford et al., 2010, within-group tournament first; Stanton & Schultheiss, 2007; van Anders & Watson, 2007, study 2), that did not have both a winning and a losing condition (Carré & Putnam, 2010, study 2; Gonzalez-Bono et al., 2000; Mehta et al., 2015), or, if they did have both conditions, involved conditions that were not matched on other important variables (e.g., the type of competition or sport differed between the winning and losing groups; Edwards & Kurlander, 2010). With these search terms and the inclusion and exclusion criteria, our initial analysis of the 'winner-loser' effect involved 62 effect sizes (two of which were subsequently identified as outliers and removed, leaving 60 effect sizes total), which were extracted from 49 different manuscripts (see Table 1). Note that when manuscripts involved multiple studies or reported results separately for men and for women or for different groups of participants (e.g., ingroup vs outgroup in Flinn et al., 2012), we were able to extract more than one effect size, thus accounting for the greater number of effect sizes than manuscripts included in the current meta-analysis.

2.1. Coding of study information and moderators

2.1.1. Competition duration, and relative timing of the pre- and post-competition testosterone measurements used to calculate the effect size

When possible, we extracted the length of the competition and the timing of the pre- and the post-competition testosterone measures relative to the onset and completion of the competition. When this information was not provided, it was estimated based on figures showing the experimental timelines, and other studies that utilized similar competitions. When there were multiple times in which testosterone was sampled, we preferentially calculated and extracted effect sizes based on the sampling times immediately before and immediately after the competitions.

2.1.2. Mean age

We extracted the mean age of the sample whenever this information was reported in the corresponding manuscripts. If men and women were analyzed separately, but age was only reported for the combined sample, we used the same mean age for both sexes. If an age range rather than a mean was reported, we used the mid-point of the range in our analyses.

2.1.3. Sample size

Whenever possible, we aimed to extract the sample size included in the analysis of the 'winner-loser' effect. Despite using a within-subject design in which the same participants experienced both a win and a loss, some authors analyzed the data as if each competition experience was a separate participant (e.g., Booth et al., 1989). In such cases, we used the same n as used in the authors' analyses. If the n within each condition (win vs loss) was not reported, but the total n was reported, we assumed an equal number of winners and losers in each condition (which sometimes led to non-discrete values for the n within winners and within losers).

2.1.4. Country of study

Studies were also coded for the country in which they were conducted. If this information was not directly reported in the text, we assumed the country to be the same as that of the corresponding authors' institution.

2.1.5. Physical activity

We coded whether the competition used in the study involved physical activity (hockey, rugby, tennis, soccer or football, wrestling, judo, basketball, badminton, hunting, karate, kickboxing, video games in which players control avatars with full body movements, dog agility competition) or not (e.g., rock-paper-scissors, poker, watching a sporting event or election outcomes, laboratory tasks, dominoes, chess, video games involving button pressing, coin tosses).

2.1.6. Watching versus playing

We coded each study for whether participants competed themselves or, instead, watched other individuals compete (e.g., watched the outcomes of elections or sporting events).

2.1.7. Location of testing

We coded each study for whether it was conducted in the lab or not (e.g., in the field, at the participant's home, at a sporting arena, pub or bar).

2.1.8. Outcome determination method

We coded each study for whether the outcome was determined naturally, through the skills of each player involved in the competition, or if it was manipulated experimentally by rigging the contest.

2.1.9. Method for determining testosterone concentrations

We coded each study for whether testosterone was determined through plasma/serum versus saliva.

2.1.10. Time of testing

We also coded the time of day during which the study was conducted (a time range was often reported). In some cases, the time of testing was not directly reported but the time of saliva sampling was reported, in which case we used these times as estimates of the experimental testing times in our analysis. We then grouped the studies into those involving mornings (as well as afternoons) versus those involving afternoons only.

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

2.2. Extracting and calculating effect sizes and sampling variance

Effect sizes were calculated using the means and standard deviations of pre- and post-competition testosterone values, change scores (post- minus pre-competition testosterone values), and percent change scores (post- minus pre-competition testosterone values, divided by pre-competition testosterone values) using formulas in Table 1 of Morris & DeShon (2002). In some cases, the means and standard deviations were not reported in text but the authors included a figure from which these values could be estimated; two of the authors estimated each of these values and the average of these two estimates were used in all analyses (preliminary analysis suggested the estimates were highly reliable: $r = 0.99$; $t_{166} = 0.28$, $p = 0.78$). If these values were not available, we used the F, t, z, w^2, or p values from other relevant analyses that tested the winner-loser effect, and then converted these to D values using the effect size calculator provided in Wilson's (2001) effect size

Table 1
Study details used in the meta-analysis.

	Testosterone changes during competition			Sample size (n)			Moderator variables											
Manuscript	Losers (D)[a]	Winners (D)[a]	Winner Effect (D)[b]	Losers	Winners	Combined[c]	A	B	C	D	E	F	G	H	I	J	K	L
Aguilar et al., 2013[d]	−1.150	0.593	1.744	7	7	7	1	1	0	1	1	1	0	1	–	1	1	1
Apicella et al., 2014	0.303	0.777	0.474	24	25	49	1	1	1	0	0	1	0	1	1	0[o]	1	1
Bernhardt et al., 1998, Study 1	–	–	1.544	4	4	8	1	1	1	1	0	0	0	1	–	1	0	1
Bernhardt et al., 1998, Study 2	–	–	1.102	12	9	21	1	1	1	1	0	0	0	1	–	1[p]	1[q]	1
Booth et al., 1989	−0.269	0.402	0.671	16[e]	20[e]	36[e]	1	–	1	1	1	1	0	1	–	1	0	1
Carré & Putnam, 2010[m]	0.275	0.687	0.413	15[f]	15[f]	15[f]	1	0	1	0	0	0	–	1	0	0	0	1
Carré et al., 2013[m]	−0.385	0.075	0.460	52.5	52.5	105	1	0	1	0	1	1	1	1	1	0	0	1
Carré et al., 2013[m]	−0.133	−0.356	−0.223	56.5	56.5	113	0	0	1	0	1	1	1	1	1	0	0	1
Carré et al., 2009[m]	−0.421	−0.220	0.201	14	13	27	1	0	1	0	0	1	1	1	0	0	0	0
Carré et al., 2009[m]	−0.096	−0.214	−0.118	25	25	50	0	0	1	0	0	1	1	1	0	0	0	0
Costa & Salvador, 2012	−0.231	0.328	0.559	15.5	15.5	31	0	0	0	0	0	1	0	1	0	0	0	0
Denson et al., 2013	−0.257	0.344	0.601	23	30	53	0	0	0	0	0	1	1	1	1	1[r]	0	1
Edwards et al., 2006[h]	1.225	1.068	−0.157	10[g]	10[g]	10	0	0	1	1	1	1	0	1	0	1	1	1
Elias, 1981[s]	0.086	0.451	0.365	6	7	13	1	0	1	1	1	1	0	0	–	0	0	0
Filaire et al., 2001	0.189	−0.404	−0.593	9	9	18	1	0	0	1	1	1	0	1	0	0	0	0
Flinn et al., 2012, Outgroup	–	–	0.481	8	8	16	1	1	0	1	0	1	0	1	0	0	0	–
Flinn et al., 2012, Ingroup	–	–	0.165	8	8	16	1	1	0	1	0	1	0	1	0	0	0	–
Fry et al., 2011	0.826	0.972	0.146	31[e]	34[e]	65[e]	1	0	1	1	1	1	0	0	1	–[t]	0[u]	0
Gladue et al., 1989[i]	–	–	0.583	19.5	19.5	39	1	0	1	0	0	1	0	1	0	1[v]	1[w]	1
Gonzalez-Bono et al., 1999	−0.370	0.253	0.624	8	7	15	1	0	0	1	1	1	0	1	1	1	1	1
Hamilton et al., 2009	–	–	0.109	13	13	13	0	0	1	1	1	1	0	1	–	1	0	–
Hasegawa et al., 2008	2.271	2.840	0.569	26	15	41	1	0	0	1	0	1	0	1	–	0	0	–
Jiménez et al., 2012	−1.215	1.562	2.778	17	10	27	1	0	0	1	1	1	0	1	1	1	1	1
Jiménez et al., 2012	−0.597	0.898	1.495	17	6	23	0	0	0	1	1	1	0	1	1	1	1	1
Jones & Josephs, 2006	−0.203[m]	0.391[m]	0.594[m]	40	43	83	1	1	1	1	1	1	0	1	0	1	1	–
Maner et al., 2008	–	–	0.242[ii]	11.5	11.5	23	1	0	1	0	0	1	1	1	0	0	1	–
Maner et al., 2008	–	–	−0.044[ii]	17.5	17.5	35	0	0	1	0	0	1	1	1	0	0	1	–
Mazur et al., 1992, Regional Tournament[jj]	–	–	1.552	7	4	11	1	–	1	1	0	1	0	1	1	–[n]	1	1
Mazur et al., 1997[ee]	0.000	0.000	0.000	14	14	28	1	0	1	0	0	1	0	1	0	0	0	1
Mazur et al., 1997[ee]	−0.008	−0.006	0.001	16	16	32	0	0	1	0	0	1	0	1	0	0	0	1
McCaul et al., 1992, Study 1	0.084	0.088	0.004	14	14	28	1	1	1	0	0	1	1	1	0	0	0	1
McCaul et al., 1992, Study 2	−0.090	0.172	0.261	35	35	70	1	1	1	0	0	1	1	1	0	0	0	1
Mehta & Josephs, 2006	−0.075	−0.118	−0.043	27	23	50	1	–	1	0	0	1	1	1	0	0	1	0
Oliveira et al., 2009	−0.738	1.104	1.843	16	13	29	0	0	0	1	1	1	0	1	0	1	1	1
Oliveira et al., 2013	1.787	−0.163	−1.950	17	17	34	0	0	0	0	0	1	1	1	0	0	1	1
Oliveira et al., 2014	0.481	0.191	−0.290	18	18	36	1	0	0	0	0	1	1	1	0	–[x]	1	1
Oxford et al., 2010, Between Group First	–	–	−0.546[j]	9.5	9.5	19	1	0	1	0	0	1	0	1	0	–[kk]	1	1
Parmigiani et al., 2006, Competition[y]	0.030	0.407	0.377	11	11	22	1	1	0	1	1	1	0	0	0	0	0	0
Parmigiani et al., 2009[y]	0.257	0.259	0.002	12	12	24	1	1	0	1	1	1	0	0	0	0	0	0
Pesce et al., 2015[hh]	−1.999	−2.297	−0.298	13	12	25	1	1	0	1	1	1	0	1	0	1	1	0
Pound et al., 2009[mm]	0.011	0.358	0.348	10	47	57	1	0	0	0	0	1	1	1	0	0	0	0
Salvador et al., 1987	−0.126	−0.151	−0.025	6	6	12	1	0	0	1	1	1	0	0	1	0	1	0
Schultheiss & Rohde, 2002[ff]	−0.285	−0.117	0.169	32[dd]	34[dd]	66	1	0	0	0	0	1	1	1	0	0	0	0
Schultheiss et al., 1999	–	–	0.277	21	21	42	1	0	1	0	0	1	1	1	0	–[n]	0	0
Schultheiss et al., 2005, Study 1	0.084	−0.099	−0.184	46	41	87	1	0	1	0	0	1	1	1	1	0	0	0
Schultheiss et al., 2005, Study 2	0.038	−0.038	−0.076	36	38	74	0	0	1	0	0	1	1	1	1	0	0	0
Serrano et al., 2000	0.000	0.200	0.200	7	5	12	1	0	0	1	1	1	0	1	1	1	0	1
Stanton et al., 2009[k]	−0.126	0.016	0.142	28.5	28.5	57	1	0	1	1	0	0	0	1	0	–[z]	0	–
Stanton et al., 2009[k]	−0.069	−0.070	−0.001	53	53	106	0	0	1	1	0	0	0	1	0	–[z]	0	–
Steiner et al., 2010	0.300	0.293	−0.006	16	16	32	1	0	1	0	0	1	0	1	0	0	0	1
Suay et al., 1999	0.349	0.550	0.218	13	14	27	1	0	0	1	1	1	0	1	1	0	0	–
Trumble et al., 2012[m]	0.642	0.579	−0.063	41	41	82	1	1	0	1	1	1	0	1	1	1	0	–
Trumble et al., 2014[gg]	−0.125	0.334	0.459	13	18	31	1	1	0	1	1	1	0	1	1	–[aa]	–[ll]	1
van Anders & Watson, 2007, Study 1	−0.366	−0.198	0.168	19	18	37	1	0	1	0	0	1	0	1	0	0	1[bb]	0
van Anders & Watson, 2007, Study 1	−0.268	−0.231	0.037	19	19	38	0	0	1	0	0	1	0	1	0	0	1[bb]	0
van Anders & Watson, 2007, Study 2	–	–	−0.366	21	22	43	0	0	1	0	0	1	1	1	0	0	1[bb]	0
van der Meij et al., 2010	0.491	0.377	−0.114	42	42	84	1	0	0	0	0	1	1	1	0	1[cc]	0	1
Welker & Carré, 2014[m]	−0.010	−0.119	−0.109	33	39	72	1	0	1	0	0	1	1	1	1	0	0	0
Zilioli & Watson, 2012	−0.232	−0.013	0.219	29	30	59	1	0	1	0	0	1	1	1	0	0	1	0
Zilioli & Watson, 2014, First Competition[l]	−0.253	−0.019	0.234	40	40	80	1	0	1	0	0	1	1	1	0	0	1	0
Zilioli et al., 2014, Study 1	0.078	−0.205	−0.283	33	32	65	0	0	1	0	0	1	1	1	1	0	1	0
Zilioli et al., 2014, Study 2	−0.238	−0.491	−0.253	27	26	53	0	0	1	0	0	1	1	1	0	0	1	0

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

determination program or other conversion formulas. Whenever possible, we calculated the effect sizes in the raw score metric such that a *D* value of 1 indicates that the increase in testosterone among winners was 1 *SD* greater than the increase in losers or, similarly, that the decrease in losers was 1 *SD* greater than that of winners. Effect sizes with negative values indicate that the increase in losers was greater than that of winners or that the decrease in winners was greater than that of losers. For other studies, we first calculated the effect size in the change score metric and then transformed it to the raw score metric using Formula 11 in Morris and DeShon (2002) (we find no differences between effect sizes that were initially calculated as change scores, and then converted to raw scores, versus those that were initially calculated as raw scores, see Results section). To calculate the sampling variance for each effect size, we used sampling variance formulas for raw effect sizes provided in Table 2 of Morris and Deshon (2002). Specifically, for effect sizes that were calculated based on post-competition testosterone values only (rather than on both pre- and post-competition, change, or percent change scores), we used the "independent-groups posttest" formula. For effect sizes that were calculated based on pre- and post-competition testosterone values (including change scores), we instead used the "single-group pretest-posttest" formula for winners and losers separately, and then summed these two values (as recommended in Morris

Notes to Table 1:

[a] Positive values indicate increases in testosterone from pre- to post-competition.

[b] Positive values indicate greater increases (or lesser decreases) for winners compared to losers.

[c] In some cases the *n* for the combined sample will not equal the sum of the winner *n* and the loser *n* because the study involved a within-subject design in which the same participants won one competition and lost another. A = Sex (0 = women, 1 = men). B = Age (0 = younger than 25, 1 = older than 25 years). C = Country of Study (0 = country outside of North America, 1 = country within North America). D = Testing Location (0 = laboratory, 1 = not in the laboratory). E = Physical Activity (0 = no physical activity, 1 = physical activity). F = Watching vs Playing (0 = watched the competition, 1 = played in the competition). G = Method for Determining Outcome (0 = outcome determined naturally, through the competition, 1 = competition was rigged, outcome was manipulated experimentally). H = Method of Measuring Testosterone (0 = plasma, 1 = saliva). I = Time of Testing (0 = afternoon only, 1 = morning or morning and afternoon). J = Time of Pre-Competition Testosterone Sample (0 = within 10 min of start of competition, 1 = earlier than 10 min before start of competition). K = Time of Post-Competition Testosterone Sample (0 = within 10 min of conclusion of competition, 1 = later than 10 min of conclusion of competition). L = Duration of Competition (0 = 15 min or less, 1 = longer than 15 min).

[d] The effect sizes for the two wins were averaged, and compared to the effect size for the loss.

[e] The *n* for this study represents the number of observations rather than the number of participants given the authors conducted the analysis at the level of the observations.

[f] Although there were >15 participants in this study, the design was such that only 15 players were included in both the "winner" and "loser" video conditions; thus we used $n = 15$ in the analysis.

[g] Although there were >10 participants in this study, the design was such that only 10 women played in both the winning and in the losing match; thus, we used $n = 10$ in the analysis.

[h] We excluded men from this study given that they only experienced a victory (and thus could not be compared to a corresponding group of men that experienced a defeat).

[i] The effect size for this study was derived from the results of an ANCOVA in which pre-competition testosterone was the covariate and outcome (win vs loss) and decisiveness of outcome (close vs decisive) were the between-subject factors. Thus, the D value for the winner effect represents the difference in changes in testosterone between winners and losers, collapsed across (or controlling for) the decisiveness of the outcome.

[j] The authors also analyzed changes in testosterone as a function of the player's individual contribution. Here, for consistency with the other effect sizes included in the meta-analysis, we only extracted the effect size related to winning and losing the competition.

[k] Although participants went to the laboratory from 10 am–5 pm prior to the election, actual sampling was done at participants' homes at night, averaging from 8:08 pm–12:20 am; thus time of testing was coded as afternoon.

[l] This study also included a second competition in which the same participants competed again, on a subsequent day. To avoid non-independence, and potential carry over effects from the first to the second competition, we only included the effect sizes from the first competition.

[m] These effect sizes were calculated using values provided by the authors of the corresponding manuscripts.

[n] Although the authors collected pre-competition testosterone samples, there was insufficient information to calculate the effect size based on this information (instead, the effect size was calculated based on only the testosterone concentrations measured after the competition). Therefore, the moderator is not coded for this effect size.

[o] The authors reported that the pre-competition testosterone sample was collected before instructions were given; here, we assume that instructions took <10 min.

[p] This moderator coding is based on the author's reported range of pre-competition sample timing of 30–10 min.

[q] This moderator coding is based on the author's reported range of post-competition sample timing of 15–25 min.

[r] The authors reported that pre-competition testosterone samples were collected prior to a provocation procedure, which took 10 min, and a rating activity, both of which occurred before the start of the competition. Therefore, we coded this testosterone sample as being collected earlier than 10 min before the competition.

[s] Means and standard deviations for calculating the effect sizes were derived from raw data provided in text of this paper.

[t] There was no specification of the timing of pre-competition sample collection, and thus a moderator code is not included.

[u] The coding for this moderator was based on the article's abstract which indicated that post-competition samples were collected immediately after competition.

[v] Pre-competition sample time was coded based on the range (10–15 min) provided by the authors.

[w] Effect size was calculated from an ANCOVA with multiple post-competition time points, and thus the moderator coding here represents the last available sample included in the ANCOVA, which was >10 min post-competition.

[x] Not enough information was provided to determine the timing of the pre-competition sample.

[y] Although there was also a highly ritualized fight in this paper ("Kata"), it was not competitive in nature and was thus excluded from the analyses.

[z] Not enough information was provided to determine the timing of the pre-competition testosterone sample.

[aa] Not enough information was provided to estimate the amount of time prior to, or following the competition (i.e., gun shots with kill or no kill) that a saliva sample was collected.

[bb] Participants in this study provided the post-competition saliva sample after 25 min had passed or after participants had completed all questionnaires. We assume that most participants took longer than 10 min to complete the questionnaires.

[cc] Coding for the pre-competition sample time was estimated from other information provided (e.g., length of task, post-competition sampling interval, total time testing).

[dd] The *n*s here are based on the number of data points depicted in graphs, rather than the *n*s reported in the text of the corresponding manuscript.

[ee] Calculations are based on the normalized testosterone levels, as these were the data available in graphical form in the manuscript. Standard deviations were presented as a range; for effect size calculations, the mid-point of this range was used.

[ff] Authors from this study depicted data for low inhibition and high inhibition individuals. Calculations were based on means of the high and low groups combined.

ggWinners from this study are those who shot their guns and made a kill, while losers are those who shot their gun but did not make a kill. Post -competition values were those taken after the hunters returned home.

hh This study also included information on fight simulations from the training season, but effect sizes here are based only on official competition results.

[ii] These values were determined using the *p* values from the regression coefficient representing the effect of competition outcome (win vs loss) on testosterone change. The authors reported $df = 50$ for the first test of the regression coefficient; we therefore used this *df* for each of the effect size calculations for this manuscript. The authors reported tests of the coefficients for low and for high anxious groups. We calculated an effect for each group within men and within women, and then calculated and used the average effect across the two groups within men and within women.

[jj] The authors also report the results of a city tournament; to avoid issues of non-independence, however, we excluded the effect size from this tournament because it involved a subset of the participants from the regional tournament, which we do include.

[kk] Although the authors measured pre-competition testosterone, we could not extract an effect size that incorporated this information.

[ll] The authors reported collecting a testosterone sample 10 min after the participant returned home from a hunt. Nevertheless, it is difficult to determine whether the hunt ended well before or right before the hunter returned home. Because of this ambiguity, we did not include a value for this moderator.

[mm] The authors examine percent change scores from 10 min preceding the competition to 5-min pre-competition, 5 min after, and 20 min after the competition. The effect size reflects percent change from the baseline sample to 5-min post-competition.

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

Table 2
Associations between the moderators.

	A	B	C	D	E	F	G	H	I	J	K
B	**0.36**										
C	−0.15	−0.22									
D	0.17	**0.40**	**−0.42**								
E	0.05	0.18	**−0.38**	**0.67**							
F	−0.05	−0.11	−0.23	−0.21	0.22						
G	−0.14	**−0.30**	**0.26**	**−0.73**	**−0.45**	0.22					
H	−0.18	−0.11	0.15	**−0.33**	**−0.41**	−0.09	0.24				
I	−0.02	−0.05	−0.13	0.22	**0.39**	0.17	−0.04	−0.11			
J	−0.04	0.21	−0.24	**0.53**	**0.45**	−0.19	**−0.39**	0.19	0.18		
K	−0.12	−0.01	0.11	−0.04	−0.01	0.13	−0.04	0.13	−0.07	0.22	
L	0.06	0.21	−0.03	0.15	0.11	−0.25	**−0.29**	**0.34**	0.18	**0.57**	0.04

Notes. Bold font indicates a significant association, $p < 0.05$. Because the number of studies differs across moderator analyses, the threshold for significance changes across the cells. A = Sex (0 = female, 1 = male). B = Age (0 = younger than 25, 1 = older than 25 years). C = Country of Study (0 = country outside of North America, 1 = country within North America). D = Testing Location (0 = laboratory, 1 = not in the laboratory). E = Physical Activity (0 = no physical activity, 1 = physical activity). F = Watching vs Playing (0 = watched the competition, 1 = played in the competition). G = Method for Determining Outcome (0 = outcome determined naturally, through the competition, 1 = competition was rigged, outcome was manipulated experimentally). H = Method of Measuring Testosterone (0 = plasma, 1 = saliva). I = Time of Testing (0 = afternoon only, 1 = morning or morning and afternoon). J = Time of Pre-Competition Testosterone Sample (0 = within 10 min of start of competition, 1 = earlier than 10 min before start of competition). K = Time of Post-Competition Testosterone Sample (0 = within 10 min of conclusion of competition, 1 = later than 10 min of conclusion of competition). L = Duration of Competition (0 = 15 min or less, 1 = longer than 15 min).

& DeShon, 2002), unless outcome (win vs loss) was analyzed in the corresponding manuscripts as a within-subjects factor (as in Aguilar et al., 2013; Carré & Putnam, 2010; Edwards et al., 2006; Hamilton et al., 2009), in which case we averaged the two values. The dataset used to calculate effect sizes for the current meta-analysis is available at the following link: http://carrelab.nipissingu.ca/datasets/.

2.3. Conducting the meta-analysis

We analyzed the extracted effect sizes and variance estimates using Comprehensive Meta-Analysis, Version 3.3.070. Meta-analyses can be conducted using either fixed-effect or random-effects models. Whereas fixed-effect models assume that there is one true population effect size and that any variation in effect size from study to study represents sampling error, random-effects models assume that there are likely many different true effect sizes that are specific to different subpopulations and contexts (e.g., men versus women, field versus lab-based studies, Borenstein et al., 2009). We use random-effects models for all analyses reported here. *Q* tests are reported for each analysis, which indicate whether or not a significant proportion of the variability in the distribution of effect sizes is because of true heterogeneity (between-study differences) rather than sampling error (within-study variability). I^2 values are also reported; these values represent the proportion of variability in effect sizes that can be attributed to between-study differences rather than sampling error (Higgins & Thompson, 2002). Moderator analyses were conducted to determine if any of the heterogeneity could be attributed to specific study characteristics (e.g., sex of participants, study location). When a significant moderator was identified, we conducted separate analyses within each subgroup to generate the subgroup's summary effect size. We report an R^2 analog (herein, R^2) for each moderator, which provides an estimate of the amount of between-sample heterogeneity in effect sizes that is explained by the moderator. When multiple moderators were identified as significant, we examined the association between the moderator variables. If there was substantial overlap between the moderators (i.e., they were correlated, $ps < 0.05$), we ran a meta-regression with both moderators entered into the model simultaneously, which allowed us to examine the unique effect of one moderator over and above the effect of the other. Although the coefficients from some of these analyses were non-significant, the magnitude of the coefficients offered some insight into the extent to which the moderators contributed uniquely or independently (i.e., over and above the other, related, moderators) to the heterogeneity in the effect sizes. The *B* weights from these analyses represent the difference in *D* values between the two groups, controlling statistically for the other moderators in the model.

2.4. Publication bias

For each analysis in which the summary effect size was significant, we calculated a fail-safe *n* (Rosenthal, 1979), which represents an estimate of the number of null effect sizes ($Ds = 0.00$) that would need to be added to the meta-analysis to nullify the summary effect ($p > 0.05$). To identify potential publication bias, we visually inspected funnel plots and conducted Egger's test (one-tailed) of the regression intercept (Egger et al., 1997). When there was evidence of publication bias in an analysis involving 10 or more effect sizes, we provided additional estimates of the effects after adjustments using random-effects trim and fill techniques (Duval & Tweedie, 2000). Nevertheless, we caution that these asymmetry tests can be unreliable and increase the likelihood of false positives when a given distribution of effect sizes is heterogeneous, which was the case for most of our analyses (for more on the conditions under which these asymmetry tests should not be employed, see Ioannidis & Trikalinos, 2007a). Further, asymmetry may result from factors other than publication bias (e.g., chance, true heterogeneity, reviewed in Egger et al., 1997). We also performed the Test for Excess Significance (TES; Ioannidis & Trikalinos, 2007b). TES evaluates whether there is an excess of statistically significant results in the literature by comparing the observed number of statistically significant findings to the expected number based on estimates of statistical power. The cause of the discrepancies between observed and expected findings may be due to publication bias, but may also be due to excessive researcher degrees of freedom in the primary literature, fabrication of data, or randomness. The estimated power of each study data set depends on the "true" plausible effect size. In the current meta-analysis, we considered the true effect size to be equal to (a) the observed effect size from individual studies or (b) the observed effect size from the random-effects meta-analysis. Given the large degree of heterogeneity in the 'winner-loser' effect (see results), we performed TES on the following subsamples: male participants tested in the lab, male participants tested in a non-lab environment, female participants tested in the lab, and female participants tested in a non-lab environment. In each case, the observed number of studies with statistically significant results is compared with the expected number of statistically significant studies using a binomial test. Finally, we also investigated the extent to which variability in sample size was associated with variability in effect size (Kuhberger et al., 2014). Negative associations between sample size and effect size can be suggestive of publication bias.

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

3. Results

3.1. Preliminary analysis: What is the correlation between pre- and post-competition testosterone concentrations?

One challenge in conducting a meta-analysis on effect sizes that capture changes across time is that the correlation between the measures at the first and second sampling time is needed for calculating the sampling variance (see formulas in Table 2 of Morris & DeShon, 2002). This value, however, is rarely reported. When these correlations are missing for many of the effect sizes, researchers recommend conducting a meta-analysis on the available correlations and then using the mean weighted effect size as an estimate of the strength of the correlation for studies that do not report this value (Lipsey & Wilson, 2001; Morris & DeShon, 2002). Preliminary analysis on these pre-post correlations indicated that the strength of associations did not differ for winners and losers ($k = 19$, $Q_1 = 0.40$, $p = 0.53$). Thus, when authors reported the correlations for the total sample (involving winners and losers combined) in addition to the winners and losers separately, we used the correlations from the total, combined sample.

The mean weighted correlation between pre- and post-competition testosterone was $r = 0.76$ (95% *CIs* = 0.66 to 0.84, $p < 0.001$; fail-safe $n = 2946$, Egger's intercept = 0.02, $p = 0.50$). Despite heterogeneity in the distribution of effect sizes ($Q_{18} = 126.50$, $p < 0.001$, $I^2 = 85.77$), neither the sex of the participants ($k = 19$), the mean age of the participants ($k = 18$), nor the estimated time that elapsed between the two sampling time points ($k = 15$) moderated the strength of the association ($ps > 0.28$). Thus, when calculating the sampling variance of each effect size, we used a pre-post correlation of $r = 0.76$.

3.2. Do winners demonstrate a larger increase (or smaller decrease) in testosterone concentrations relative to losers?

We identified two outlying effect sizes (male sample of Jiménez et al., 2012, >3 *SDs* above the mean; Oliveira et al., 2013, <3 *SDs* below the mean) in an initial forest plot (see Fig. 1), which were removed from all analyses.[1] Winners showed larger pre- to post-competition increases in testosterone (or smaller decreases) than did losers ($k = 60$, $D = 0.20$, 95% *CIs* = 0.10 to 0.31, $p < 0.001$; fail-safe $n = 540$), an effect that was heterogeneous with 65% of the variability ($I^2 = 64.74$) related to between-sample differences rather than to sampling error ($Q_{60} = 167.33$, $p < 0.001$).[2] The distribution of effect sizes was asymmetrical (Egger's intercept = 1.46, $p = 0.01$; see funnel plot in Supplementary Figures), but random-effects trim and fill produced a trivial increase in the effect size ($D = 0.22$, 95% *CIs* = 0.12 to 0.33). Sex did not moderate the magnitude of the 'winner-loser' effect ($p = 0.30$), but the effect was significant only in men ($k = 44$, $D = 0.23$, 95% *CIs* = 0.13 to 0.34, $p < 0.001$; $Q_{43} = 86.25$, $p < 0.001$, $I^2 = 50.15$; fail-safe $n = 382$, Egger's intercept = 0.74, $p = 0.11$) (in women: $k = 16$, $D = 0.14$, 95% *CIs* = −0.10 to 0.38, $p = 0.24$; $Q_{15} = 73.15$, $p < 0.001$, $I^2 = 79.50$). However, the distribution of effect sizes in women was asymmetrical (Egger's intercept = 4.50, $p = 0.01$) and random-effects trim and fill increased the effect size ($D = 0.22$, 95% *CIs* = −0.02 to 0.45).

To better characterize the differences between winners and losers, we conducted follow-up analyses involving every study that reported the pre- and post-competition means and *SDs* or change scores for the winners and for the losers separately (85% of studies; see studies that included a separate effect for winners and for losers in Table 1). Rather than analyzing the difference scores representing the changes in winners relative to losers (winners' pre-post change minus losers' pre-post change) as in the main analysis above, we instead conducted the analysis on each of the pre-post changes within winners and within losers separately. Four additional outlying effect sizes (>3 *SDs* of mean) were excluded from this specific analysis (winners and losers from Pesce et al., 2015; winners and losers from Hasegawa et al., 2008). For winners, testosterone concentrations increased significantly from pre- to post-competition ($k = 46$, $D = 0.18$, 95% *CIs* = 0.07 to 0.29, $p = 0.001$; $Q_{45} = 244.85$, $p < 0.001$, $I^2 = 81.62$; fail-safe, $n = 504$; Egger's intercept = 1.42, $p = 0.08$). For losers, testosterone concentrations decreased, non-significantly, from pre- to post-competition ($k = 46$, $D = -0.02$, 95% *CIs* = −0.13 to 0.08, $p = 0.65$; $Q_{45} = 221.30$, $p < 0.001$, $I^2 = 79.67$; Egger's intercept = −0.23, $p = 0.41$). Thus, our follow-up analyses indicated that the 'winner-loser' effect was driven by winners experiencing a rise in testosterone and losers experiencing no change in testosterone. We also performed similar analyses split by sex. For male winners, testosterone concentrations increased significantly from pre- to post-competition ($k = 33$, $D = 0.21$, 95% *CIs* = 0.10 to 0.33, $p < 0.001$; $Q_{32} = 131.78$, $p < 0.001$, $I^2 = 75.72$; fail-safe, $n = 440$; Egger's intercept = −0.07, $p = 0.47$). For male losers, testosterone concentrations did not change significantly from pre- to post-competition ($k = 33$, $D = 0.01$, 95% *CIs* = −0.12 to 0.15, $p = 0.83$; $Q_{32} = 162.12$, $p < 0.001$, $I^2 = 80.26$; Egger's intercept = −0.50, $p = 0.34$). Thus, for men, our follow-up analyses indicated that the 'winner-loser' effect was driven by winners experiencing a significant rise in testosterone and losers experiencing no change in testosterone. For female winners, testosterone concentrations did not change significantly from pre- to post-competition ($k = 13$, $D = 0.10$, 95% *CIs* = −0.12 to 0.33, $p = 0.36$). Examination of the funnel plot (see Supplementary Figures) indicated that the distribution was highly heterogeneous, ($Q_{12} = 85.22$, $p < 0.001$, $I^2 = 85.92$) and asymmetrical (Egger's intercept = 5.21, $p = 0.008$); random-effects trim and fill did not, however, change the summary estimate. For female losers, testosterone concentrations did not change significantly from pre- to post-competition ($k = 13$, $D = -0.12$, 95% *CIs* = −0.29 to 0.05, $p = 0.17$; $Q_{12} = 50.19$, $p < 0.001$, $I^2 = 76.09$; Egger's intercept = 0.12, $p = 0.48$). Therefore, in women, although testosterone changes were divergent for winners and losers, with winners slightly increasing and losers slightly decreasing from pre- to post-competition, neither the increase in winners nor the decrease in losers was statistically significant. This divergent pattern, however, likely accounted for an overall winner-loser effect in women ($D = 0.22$, after trim and fill) that was of a comparable magnitude to that found in men ($D = 0.23$).

To determine the corresponding percent change in testosterone associated with these effect size values, we calculated percent change for each effect using pre- and post-competition means [((post-competition testosterone − pre-competition testosterone) / pre-competition testosterone)) * 100], or we used the percent change means reported by the authors of the corresponding manuscripts. We then regressed these percent change scores onto the computed effect sizes (*D* values) in a linear regression analysis.[3] The model [percent change score = 0.05 + (43.63 * *D*)] accounted for 85.6% of the variability in percent change scores ($F_{1,80} = 470.17$, $p < 0.001$). Using this model, the effect sizes obtained above

[1] The same pattern of results emerged when these two outlying effect sizes were winsorized (reduced to ±3 *SDs* of the mean) and included in the meta-analysis rather than excluded: the overall winner-loser effect was of a similar magnitude ($D = 0.21$, 95% *CIs* = 0.09 to 0.34) and was moderated by location of testing ($B = 0.46$) and by the timing of the pre-competition testosterone sample ($B = 0.62$) ($ps < 0.01$).

[2] Effect sizes that were initially calculated as change scores but then converted to raw scores ($k = 17$) did not differ from those that were initially calculated as raw scores ($k = 43$) ($Q_1 = 0.20$, $p = 0.66$), supporting our combining of these effects for meta-analyses. Effect sizes that were extracted from studies in which outcome was treated as a within-subject factor ($k = 4$) did not differ from those extracted from studies in which outcome was treated as a between-subject factor ($k = 56$) ($Q_1 = 1.50$, $p = 0.22$), again supporting our decision to combine effect sizes from both types of studies in our analyses.

[3] We were unable to calculate percent change scores for 11 (11%) of the samples. The regression analysis also excluded Oliveira et al. (2013), Pesce et al. (2015), Hasegawa et al. (2008), and the male sample of Jiménez et al. (2012), because they were identified as outliers in the preceding analyses, and the winner effect from Aguilar et al. (2013), because it involved an average of two rather than a single winning competition.

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

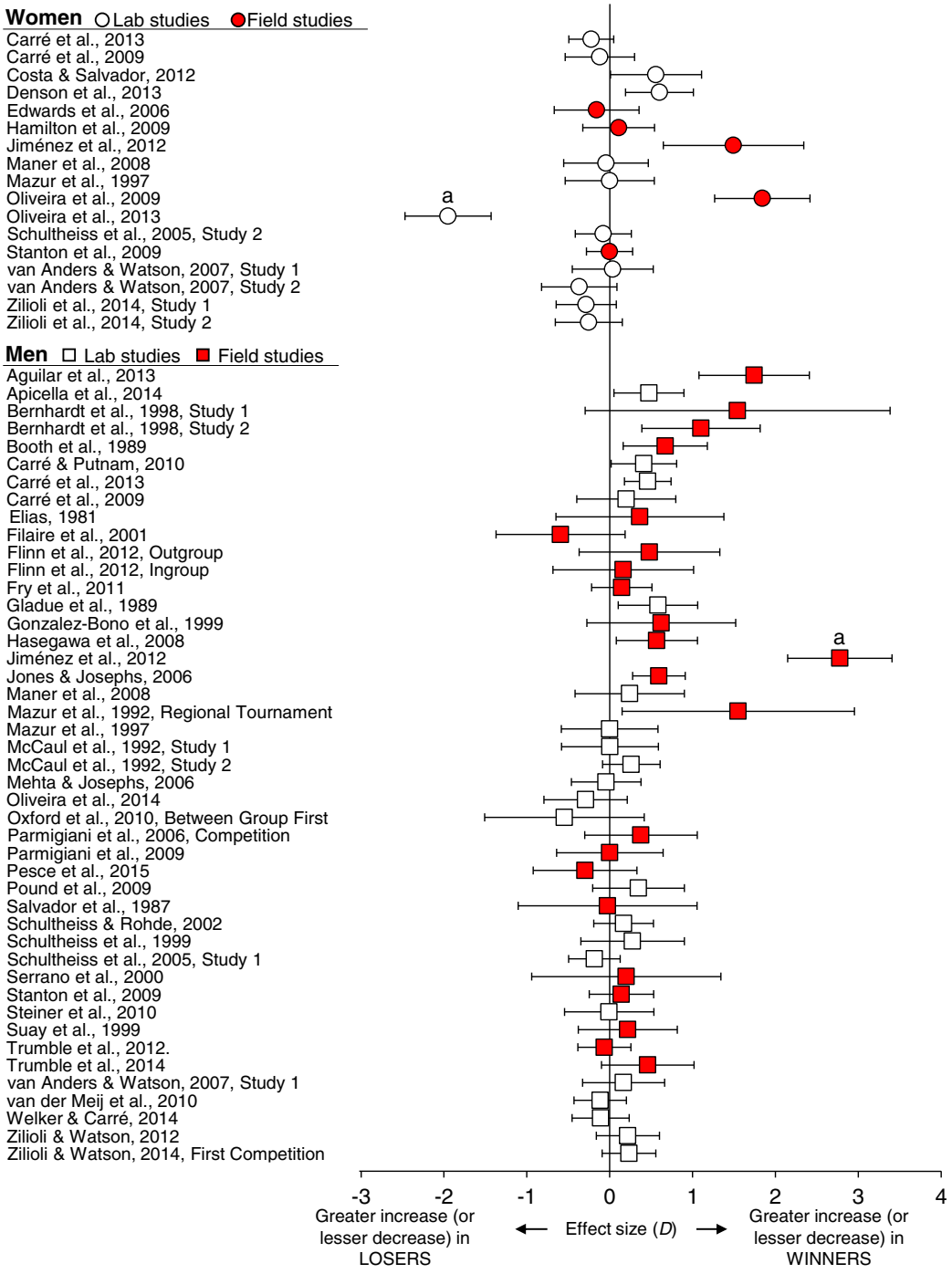

Fig. 1. Forest plot depicting effect size from each sample. Error bars represent 95% confidence intervals. [a]Outlying effect sizes (±3 *SDs* from the mean) that were removed from the main analyses.

corresponded to a 7.90% increase in testosterone among winners and a 0.82% decrease in testosterone among losers. Using the same percent change method split by sex, these effect sizes correspond to a 9.21% and a 4.41% increase for winning men and women, and to a 0.49% increase and a 5.19% decrease in losing men and women, respectively.

3.3. Moderators of the winner-loser effect

Age ($k = 57$, $Q_1 = 4.07$, $p = 0.04$, $R^2 = 0.07$), country of study ($k = 60$, $Q_1 = 4.20$, $p = 0.04$, $R^2 = 0.04$), location of testing ($k = 60$, $Q_1 = 8.86$, $p = 0.003$, $R^2 = 0.11$), method of determining contest outcome ($k = 59$, $Q_1 = 8.87$, $p = 0.003$, $R^2 = 0.15$), timing of the pre-competition testosterone sample ($k = 52$, $Q_1 = 12.25$, $p < 0.001$, $R^2 = 0.16$), and the duration of the competition ($k = 49$, $Q_1 = 9.83$, $p = 0.002$, $R^2 = 0.15$) were all significant moderators. All other moderators were non-significant (watching vs playing: $k = 60$, $Q_1 = 0.81$, $p = 0.37$; testosterone determination method: $k = 60$, $Q_1 = 0.03$, $p = 0.87$; time of testing: $k = 53$, $Q_1 = 0.15$, $p = 0.69$; timing of post-competition testosterone sample: $k = 59$, $Q_1 = 1.45$, $p = 0.23$); and physical activity: $k = 60$, $Q_1 = 3.64$, $p = 0.06$, $R^2 = 0.04$). Because the location of testing was correlated

with the other significant moderators (see Table 2), we conducted several meta-regressions, pitting testing location against one of the other correlated moderators in each model. Testing location emerged as the stronger and/or significant predictor when pitted against age ($k = 57$; testing location: $B = 0.24$, $p = 0.05$; age: $B = 0.15$, $p = 0.27$), country of study ($k = 60$; testing location: $B = 0.27$, $p = 0.02$; country of study: $B = -0.13$, $p = 0.29$), and, albeit slightly, the method of determining the contest outcome ($k = 59$; location of testing: $B = 0.21$, $p = 0.19$; method of determining outcome: $B = -0.16$, $p = 0.30$). When pitted against the timing of the pre-competition testosterone sample, however, testing location was a weaker, non-significant moderator ($k = 52$; location of testing: $B = 0.18$, $p = 0.22$; timing of pre-competition sample: $B = 0.34$, $p = 0.03$). Therefore, testing location appears to moderate the magnitude of the 'winner-loser' effect, with winners showing larger testosterone increases than losers, especially when the study is conducted outside of, compared to, within the lab. This difference between studies conducted in the lab versus outside of the lab is partly because studies conducted outside of the lab are more likely to involve outcomes that are determined naturally, through competition (rather than determined experimentally, through contest rigging), and because such studies often involve testosterone samples collected before (rather than within) 10 min of the start of the competition.

The timing of the pre-competition testosterone sample was also correlated with the duration of the competition such that studies involving longer competitions were more likely to involve testosterone samples collected earlier than (rather than within) 10 min of the start of the competition. When both of these variables were entered simultaneously as moderators in a meta-regression ($k = 43$), the timing of the pre-competition testosterone sample was significant ($B = 0.44$, $p = 0.009$) and the duration of the competition was marginally significant ($B = 0.24$, $p = 0.10$). Therefore, the 'winner-loser' effect is exaggerated among studies in which testosterone is collected earlier than (vs within) 10 min of the start of the competition and studies in which the competition duration is longer than (vs shorter than or equivalent to) 15 min. In Table 3, we report the results of separate meta-analyses conducted within the subgroups of the location, pre-competition sampling time, and duration of competition variables. For the interested reader, we also provide similar analyses split by sex (see Table 4). Notably, the effects of all significant moderators (except for country of study) show very similar patterns of effects in men and women—both in terms of magnitude of effect, and direction of effect.

3.4. Publication bias

Examination of funnel plots and Egger's intercept revealed some evidence for publication bias. The trim and fill technique (Duval & Tweedie, 2000) increased effect sizes in some case (full sample, $D = 0.20$ to $D = 0.22$; lab studies, $D = 0.43$ to $D = 0.46$; women, $D = 0.14$ to $D = 0.20$). In two cases, the trim and fill technique reduced the overall effect size, though it remained statistically significant in both cases (pre competition samples obtained >10 min before the competition, $D = 0.59$ to $D = 0.38$; competition lasting >15 min, $D = 0.43$ to $D = 0.29$). For the subsample of women for whom pre-competition samples were obtained within 10 min of the start of the competition, the effect size became statistically significant after trim and fill ($D = -0.13$ to $D = -0.20$).

3.5. Test of excess significance

When TES was calculated using the effect size from each individual sample, there was no evidence for bias (i.e., excess significance). However, when TES was calculated using the effect size estimate from the random-effects meta-analysis, there was evidence for bias in most of the subsamples tested (see Table 5). Thus, the results suggest that there are more significant effects reported in the literature than expected based upon the statistical power of the studies. However, it should be noted that there was a high degree of heterogeneity (even within-

Table 3
Results of follow-up meta-analysis within each subgroup of the significant moderators.

		Summary estimates			Heterogeneity statistics		Publication Bias statistics	
			CIs					
Moderators	*k*	*D*	L	H	*Q*	I^2	Fail-safe *n*	Egger's intercept
Age								
<25 years of age	43	**0.13**	**0.02**	**0.25**	**109.11**	**61.51**	100	1.14
>25 years of age	14	**0.41**	**0.15**	**0.66**	**38.40**	**66.64**	91	1.20
Country of study								
Country within North America	38	**0.12**	**0.02**	**0.22**	**74.61**	**50.41**	87	1.01
Country outside of North America	22	**0.38**	**0.13**	**0.62**	**85.60**	**75.47**	155	1.67
Testing location								
Lab	33	0.08	−0.02	0.18	**57.48**	**44.33**		0.37
Non-lab	27	**0.43**	**0.21**	**0.64**	**94.71**	**72.55**	312	**1.53**[a]
Method of determining contest outcome								
Outcome determined naturally, through competition	36	**0.35**	**0.19**	**0.52**	**105.52**	**66.83**	431	1.17
Rigged competition, outcome manipulated experimentally	23	0.03	−0.08	0.14	**40.68**	**45.91**		0.36
Timing of pre-competition testosterone sample								
>10 min pre-competition	16	**0.59**	**0.27**	**0.90**	**84.99**	**82.35**	228	**2.77**[b]
≤10 min pre-competition	36	0.09	−0.01	0.18	**50.48**	**30.66**		0.48
Competition duration								
>15 min	25	**0.43**	**0.21**	**0.65**	**108.12**	**77.80**	324	**2.10**[c]
≤15 min	24	0.01	−0.09	0.11	25.26	8.94		0.56

Notes. CI = confidence intervals, L = low, H = high. Values in bold font are significant ($p < 0.05$).
[a] Random-effects trim and fill increased the effect size ($D = 0.46$, 95% *CIs* = 0.25 to 0.68).
[b] Random-effects trim and fill decreased the estimate ($D = 0.38$, 95% *CIs* = 0.04 to 0.72).
[c] Random-effects trim and fill decreased the estimate ($D = 0.29$, 95% *CIs* = 0.05 to 0.54).

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

Table 4
Results of follow-up meta-analysis within each subgroup of the testing location, pre-competition testosterone sampling, and competition duration variables, split by sex.

		Summary Estimates			Heterogeneity Statistics		Publication Bias Statistics	
			CIs					
Moderators	*k*	*D*	L	H	*Q*	I^2	Fail-safe *n*	Egger's intercept
Age								
Men								
<25 years of age	27	**0.15**	**0.05**	**0.25**	33.17	21.62	39	−0.10
>25 years of age	14	**0.41**	**0.15**	**0.66**	**38.97**	**66.64**	91	1.20
Women								
<25 years of age	16	0.14	−0.10	0.38	**73.15**	**79.50**		**4.50**[a]
>25 years of age*								
Country of study								
Men								
Country within North America	26	**0.26**	**0.13**	**0.38**	**42.78**	**41.57**	193	0.72
Country outside of North America	18	0.20	−0.01	0.42	**41.68**	**59.22**		1.18
Women								
Country within North America	12	**−0.12**	**−0.24**	**−0.01**	5.40	0.00	2	0.37
Country outside of North America	4	**1.09**	**0.44**	**1.74**	**15.55**	**80.70**	52	4.95
Testing location								
Men								
Lab	22	**0.15**	**0.04**	**0.26**	29.15	27.95	29	−0.43
Non-lab	22	**0.38**	**0.17**	**0.59**	**51.15**	**58.95**	161	0.87
Women								
Lab	11	−0.04	−0.22	0.14	**20.63**	**51.52**		2.62
Non-lab	5	0.61	−0.09	1.31	**43.40**	**90.78**		5.81
Method of determining contest outcome								
Men								
Outcome determined naturally, through competition	28	**0.33**	**0.16**	**0.50**	**58.89**	**54.15**	216	0.50
Rigged competition, outcome manipulated experimentally	15	0.11	−0.01	0.23	18.89	25.87		−0.11
Women								
Outcome determined naturally, through competition	8	0.44	−0.01	0.88	**46.31**	**84.88**		4.79
Rigged competition, outcome manipulated experimentally	8	−0.10	−0.30	0.10	**14.91**	**53.06**		1.50
Timing of pre-competition testosterone sample								
Men								
>10 min pre-competition	11	**0.51**	**0.16**	**0.86**	**46.64**	**78.56**	75	2.18
≤10 min pre-competition	26	**0.18**	**0.09**	**0.28**	25.55	2.17	63	−0.23
Women								
>10 min pre-competition	5	**0.74**	**0.05**	**1.44**	**35.48**	**88.73**	36	6.51
≤10 min pre-competition	10	−0.13	−0.26	0.004	9.47	4.94		**2.37**[b]
Competition duration								
Men								
>15 min	19	**0.39**	**0.17**	**0.60**	**51.97**	**65.36**	160	1.15
≤15 min	17	0.07	−0.04	0.18	12.92	0.00		0.12
Women								
>15 min	6	0.55	−0.09	1.20	**55.44**	**90.98**		6.02
≤15 min	7	−0.11	−0.30	0.08	8.66	30.69		4.14

Notes. CI = confidence intervals, L = low, H = high. Values in bold font are significant ($p < 0.05$).
[a] Random-effects trim and fill increased the effect size ($D = 0.22$, 95% *CIs* = −0.02 to 0.45).
[b] Random-effects trim and fill decreased the effect size ($D = -0.20$, 95% *CIs* = −0.35 to −0.04).
* There were no studies that involved female participants who had a mean age >25 years.

subsamples), which undermines the usefulness of TES (Ioannidis & Trikalinos, 2007a). Moreover, some have also questioned the validity of this measure (Simonsohn, 2012). Nevertheless, one clear finding from this analysis is that statistical power for assessing the 'winner-loser' effect is very low, regardless of which estimate is used to approximate the true effect size (see Table 5).

Table 5
Test of excessive significance (TES).

		Effect size estimate used for power calculation	
Sample	Measure	Individual studies	Random-effects
Women, non-lab	Average power	0.53	0.53
	TES	0.73	0.73
Women, lab	Average power	0.31	0.07
	TES	0.30	<**0.001**
Men, non-lab	Average power	0.38	0.26
	TES	0.13	**0.01**
Men, lab	Average power	0.24	0.13
	TES	0.22	**0.02**

Note: Significant values are in bold. TES = Test of excess significance.

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

4. Discussion

The current meta-analysis provides a much-needed quantitative synthesis of research examining the extent to which competition outcome modulates testosterone concentrations in humans. The overall analysis included 60 effect sizes, which included over 2500 research participants. In addition, the current meta-analysis extends previous work (Archer, 2006) by examining whether competition outcome modulates testosterone reactivity patterns in women. Results indicated that winners had elevated testosterone concentrations relative to losers (i.e., presence of the 'winner-loser' effect), an effect that was of similar magnitude in men ($D = 0.23$) and in women ($D = 0.22$, after trim and fill correction). Nevertheless, the overall 'winner-loser' effect was heterogeneous, with several moderating variables influencing its magnitude. The strength of the 'winner-loser' effect depended on the location of the competition, whereby the effect was much stronger in studies conducted outside the lab ($D = 0.46$, after trim and fill correction) compared to studies conducted in the lab ($D = 0.08$). Also, the effect size for studies conducted outside the lab was significant in men ($D = 0.38$, $k = 22$) and approached significance in women ($D = 0.61$, $k = 5$).

What may underlie the larger 'winner-loser' effect for studies conducted outside of the lab? One possibility is that the larger effect may be related to participants' investment in the competitive interaction. Participants engaging in contrived laboratory situations may not be as invested in the competitive interactions to the same degree as participants engaging in competitive interactions occurring outside the lab. Often, the studies conducted outside the lab consisted of athletes engaged in sport competitions (e.g., soccer, basketball, field hockey). The outcome of such interactions are highly salient to athletes, and thus, may be partly responsible for the more robust 'winner-loser' effect observed in studies conducted outside the lab. Another possibility is that some other factors (i.e., third variables) may underlie the 'winner-loser' effect observed in studies conducted outside the lab. Physical activity on its own, for example, is known to influence testosterone concentrations (see Vingren et al., 2010, for review), and studies conducted outside of the lab were more likely to involve physical activity than were studies conducted in the lab (see Table 1). Thus, winner-loser differences in testosterone responses might in part be explained by variation in physical activity experienced by winners and losers. Nevertheless, physical activity does not appear to fully account for differences between lab and non-lab studies: when controlling statistically for whether or not the studies involved physical activity, or when restricting the meta-analysis to studies that did not involve physical activity, the effect of testing location persisted (see supplementary analyses). Therefore, although differences in physical activity might explain, to some extent, why studies conducted outside versus inside the lab produce larger differences in testosterone responses between winners and losers, the data presented here suggest that such effects do not depend on physical activity.

Another possible reason as to why studies conducted outside the lab produce larger differences between winners' and losers' testosterone responses is that such studies typically involve spectators, whereas lab-based studies of competition rarely (if ever) involve audiences. The presence of spectators may influence the degree to which testosterone concentrations change in response to competition. Indeed, Miller et al. (2012) reported that the ratio of male-to-female spectators influenced the degree to which testosterone concentrations increased in response to Ultimate Frisbee competition. Specifically, men demonstrated a larger increase in testosterone when there were relatively more female spectators, whereas women demonstrated a larger increase in testosterone when there were relatively more male spectators (Miller et al., 2012). Finally, there is also the possibility that individual difference factors (physical and/or psychological) may influence the probability of winning—and that such individual difference factors (and not the outcome, per se) is what produces differences in testosterone reactivity patterns between winners and losers tested outside of the lab. Thus, in studies conducted outside the lab, there are numerous factors (besides competition outcome) that may contribute to the differential pattern of testosterone responses in winners relative to losers.

Another important moderator of the 'winner-loser' effect that emerged was the timing of the pre-competition hormone sampling. Specifically, studies in which the pre-competition sample was obtained >10 min prior to the competitive interaction produced larger 'winner-loser' effects than studies in which the samples were collected immediately prior to competition. One potential reason for this finding (especially in sport competitions) is that athletes typically have 'warm-ups' prior to competition, which have been shown to increase testosterone concentrations (Casto et al., 2014; Edwards & Kurlander, 2010). Supplementary analyses of data here, however, suggest that such warm-up effects associated with competitions involving physical activity cannot fully account for the effect of pre-competition testosterone sampling: the effect of the pre-competition sampling time remained significant when controlling statistically for physical activity (and thus warm-up effects associated with competitions involving physical activity) and when restricting the analysis to studies that did not involve physical activity. A related explanation may be that testosterone pulses in anticipation of competition, more generally (not specific to physical warm-ups), elevate testosterone values (relative to a true baseline) immediately prior to competition. Such pulses may create ceiling effects and thus interfere with the ability to detect differential changes in winners' and losers' testosterone concentrations across the competitive interaction. Therefore, researchers investigating the 'winner-loser' effect may benefit from collecting pre-competition testosterone samples well before (>10 min) the start of the competition.

Notably, it is not possible to draw causal claims concerning the role of competition outcome in modulating testosterone reactivity patterns from studies conducted outside the lab. Specifically, all of the studies conducted outside of the lab that were included in the current meta-analysis involved naturally determined (rather than experimentally manipulated) contest outcomes. Causal evidence for the role of competition outcome in modulating testosterone concentrations can only be found in studies that experimentally manipulate the outcome of competitive interactions. In the current meta-analysis, there was no support for an overall effect of competition outcome on testosterone responses from lab-based studies. However, secondary analyses split by sex indicated that competition outcome had a relatively small (but significant) effect on testosterone reactivity patterns in men ($D = 0.15$, $k = 22$), but not women ($D = -0.04$, $k = 11$). Therefore, there was some, albeit limited, evidence in the current meta-analysis that the outcome of a competition drives divergent testosterone responses for winners and losers.

It is possible that the 'winner-loser' effect is only found in certain individuals and/or under specific social contexts. For instance, Schultheiss et al. (2005) have found that individual differences in one's implicit need for power/dominance influenced the degree to which testosterone concentrations changed in response to a rigged laboratory competition. Specifically, male winners had elevated testosterone concentrations relative to losers, but only to the extent that they scored high on a measure of implicit power motive (Schultheiss et al., 2005). Other work has examined individual differences in trait anxiety as a personality trait that may moderate the effect of competition outcome on testosterone reactivity patterns. Maner et al. (2008) found that male losers of a rigged laboratory competition showed a decrease in testosterone concentrations compared to winners, but only to the extent that they scored high on a measure of trait social anxiety. However, in a larger sample of men and women, individual differences in trait anxiety did not moderate the effect of competition outcome on testosterone reactivity patterns (Norman et al., 2015). Other researchers have examined whether contextual factors moderate the effect of competition on testosterone responses. Carré (2009) found that male hockey players demonstrated a larger increase in testosterone after a victory that occurred in the team's home venue, compared to a similar victory that occurred in the opponents' venue. Also, other work conducted in a rural

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

Dominican community suggests that the status of one's opponent may play an important role in modulating testosterone responses to victory and defeat. In this work, Flinn et al. (2012) examined testosterone reactivity patterns during competition played among members of the same village (within-group competition), or between members of different villages (between-group competition). Results indicated that winners had elevated testosterone concentrations relative to losers, but only during between-group competition (Flinn et al., 2012).

In addition, recent work suggests that the degree to which the outcome of competition is close or decisive may influence testosterone responses to competition measured in a lab context. The social status instability hypothesis posits that competitive interactions in which the outcome is close may lead to a reverse 'winner-loser' effect whereby losers of close competitions may demonstrate a rise in testosterone concentrations relative to winners of close competition—which may ultimately promote status-seeking behavior among losers of close competitions. In contrast, winning a status contest under uncertain circumstances (i.e., close victory) might be associated with decrements in testosterone—which may promote avoidance of further status contests (see Mehta et al., 2015). In two studies, Zilioli et al. (2014) reported that testosterone concentrations were higher in women who experienced a close defeat relative to women who experienced a close victory, providing some support for the idea that a reversal of the typical 'winner-loser' effect can occur under circumstances of social instability (i.e., close victory or close defeat). Expanding upon these findings, a recent experiment manipulated both competition outcome and status stability in male participants (Wu et al., in press). The authors reported that competition outcome did not interact with status stability to predict testosterone reactivity patterns. However, a more complex three-way interaction emerged in which baseline cortisol concentrations interacted with competition outcome and status stability to predict testosterone concentrations: Competition outcome and status stability interacted to predict testosterone concentrations, but only in men with relatively high baseline cortisol concentrations. For men with high cortisol concentrations, narrow wins were associated with a decrease in testosterone concentrations relative to narrow losses (Wu et al., in press)—a finding that is consistent with previous evidence in women (Zilioli et al., 2014). Collectively, these findings suggest that it will be important for future research to examine the role of individual difference factors and social-contextual variables in moderating the effect of competition outcome on testosterone reactivity patterns.

The reciprocal component of Mazur's (1976; 1985) Biosocial Model of Status posits that changes in testosterone in response to victory or defeat will feedback to influence future dominance-related behaviors. For instance, a rise in testosterone may facilitate competitive behavior aimed at defending one's status in the face of threat. In contrast, the decrease in testosterone observed among losers may promote submissive behaviors aimed at avoiding further status decrements and/or physical injury. In support of this reciprocal component of Mazur's model (1976; 1985) a number of studies have now reported that acute changes in testosterone concentrations within the context of competition positively predict competitive motivation (Mehta & Josephs, 2006; Carré & McCormick, 2008), and aggressive/antagonistic behavior (Carré et al., 2010, 2014; Cote et al., 2013; Geniole et al., 2011, 2013; also see Carré et al., 2011 and Carré & Olmstead, 2015 for reviews). Notably, studies that have included men and women (e.g., Carré et al., 2009; Carré et al., 2013) have failed to document any relationship between testosterone reactivity to competition and subsequent aggression in women. It is possible that the aggression task used in these studies may tap into a form of aggression that is more salient to men, which may explain the lack of association between testosterone reactivity and aggressive behavior in women. However, it may also be that additional measurement error associated with assaying testosterone in women (because of variability related to menstrual cycle and to hormonal contraceptives, for example, Arslan et al., 2008; Sowers et al., 2001) may in part explain the lack of association between testosterone dynamics and these behavioral phenotypes in women. Also, it is possible that competition-induced changes in testosterone modulate other forms of social behavior in women. Indeed, Casto and Edwards (2016) recently reported that changes in testosterone following competition predicted subsequent willingness to reconcile with their competitive opponent. Despite potential sex differences in associations between neuroendocrine fluctuations and human social behavior, the data reviewed in this section are consistent with the idea that acute fluctuations in testosterone within the context of human competition may have important downstream effects on social behavior.

Despite recent evidence for the reciprocal component of Mazur's (1976; 1985) Biosocial Model of Status, we cannot make strong causal claims concerning the role of testosterone dynamics in modulating human social behavior. For example, rather than causing an increase in one's competitive and aggressive behaviors, testosterone reactivity may simply be a correlate of one's propensity to engage in such behaviors (see Edwards, 2006). Addressing this third-variable problem requires the careful experimental manipulation of testosterone concentrations via pharmacological challenge. In studies involving the pharmacological manipulation of testosterone, it is still important to consider potential moderators of this effect. Recently, Mehta et al. (2015, 2015) have experimentally manipulated testosterone concentrations in young women within the context of competitive interactions. Results indicated that testosterone's causal effect on willingness to approach competitive interactions was influenced by context (whether they won or lost a previous interaction) and personality (whether they scored high or low on a trait measure of dominance). In winners of a competitive interaction, testosterone caused women to approach a subsequent competitive interaction, but this effect was only found for those women scoring relatively high in trait dominance. For losers of a competition, testosterone caused women to avoid a subsequent competitive interaction. Another recent experiment indicates that a single dose of testosterone rapidly increases aggressive behavior in men, but only to the extent that they scored high on trait dominance and/or low in trait self-control (Carré et al., in press). These studies highlight the importance of considering individual difference factors (e.g., trait dominance) and social context (win vs. loss) when examining the causal role of testosterone in promoting dominance-related behaviors.

4.1. Limitations and future directions

One limitation of our meta-analysis is that we restricted our search to published studies and thus may have overestimated the true effect size by excluding unpublished studies. Although we implemented a number of strategies to test for publication bias and to mitigate the influence of such bias (e.g., visual inspection of funnel plots, fail-safe *N*, Egger's regression, trim and fill method), we caution readers that these approaches are not without limitations, especially in the presence of effect size heterogeneity (Ioannidis & Trikalinos, 2007a). Also, although our moderator analyses revealed several factors that influence the magnitude of the 'winner-loser' effect, several of these moderators were highly correlated with each other, and thus it remains unclear which of these moderators is most critical to influencing testosterone responses to competitive interactions. Finally, to the extent that our effect size estimate from the random-effects model represents a reasonable estimate of the true effect size in the population, our results indicate that studies examining the 'winner-loser' effect in humans are woefully underpowered (average power = 0.30, see Table 5). The effect size estimate for men tested in the lab was $D = 0.15$, and thus, a sample size of $n = 1102$ (551 winners, 551 losers) would be required to achieve 80% power to detect group differences at $p < 0.05$ (one-tailed). For studies conducted outside the lab, the effect size estimate was $D = 0.43$ (combining studies of both men and women), and thus, a sample size of $n = 136$ (68 winners, 68 losers) would be required to achieve 80% power to detect group differences at $p < 0.05$ (one-tailed). We must note that effect size estimates even within subsamples were

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

heterogeneous (Mean I^2 from subsamples in Table 3 = 57%), highlighting the importance of examining factors (e.g., psychological, contextual, hormonal) which may underlie some of the variability in effect sizes from studies examining the 'winner-loser' effect. Identifying these moderating factors may enable researchers to achieve the desired statistical power without having to obtain these relatively large sample sizes (e.g., *Ns* > 1000).

In summary, combining data from more than 2500 participants that cover 35 years of research, our findings are consistent with the idea that the outcome of competitive interactions differentially influences testosterone concentrations such that winners experience greater increases in testosterone relative to losers. This effect was particularly robust in studies conducted outside the confines of the laboratory, but was also statistically significant (though small in magnitude) in men tested in the lab. However, the effect of competition outcome on testosterone reactivity patterns was highly heterogeneous, and although some moderator variables were identified (e.g., location of competition, timing of pre-competition sample), there still remains significant unexplained heterogeneity. Thus, future research will be required to identify the source of such heterogeneity in effect size. Moreover, our results indicate that previous studies examining the effect of competition outcome on testosterone reactivity were underpowered, and thus, future research will need to substantially increase sample sizes in order to have enough statistical power to detect the small-to-moderate effect sizes reported here.

Acknowledgment

This project was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (to JMC) and by a Northern Ontario Heritage Fund Corporation (NOHFC) grant (to JMC). SNG holds a Banting Postdoctoral Fellowship from the Social Sciences and Humanities Research Council (SSHRC). BMB holds a Canada Graduate Scholarship from SSHRC, NSERC, SSHRC, and NOHFC played no further role in data collection, analysis, and interpretation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.yhbeh.2016.10.002.

References

Aguilar, R., Jiménez, M., Alvero-Cruz, J., 2013. Testosterone, cortisol and anxiety in elite field hockey players. Physiol. Behav. 119, 38–42.
van Anders, S.M., Watson, N.V., 2007. Effects of ability- and chance-determined competition outcome on testosterone. Physiol. Behav. 90, 634–642.
Apicella, C.L., Dreber, A., Mollerstrom, J., 2014. Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. Psychoneuroendocrinology 39, 58–64.
Archer, J., 2006. Testosterone and human aggression: an evaluation of the challenge hypothesis. Neurosci. Biobehav. Rev. 30, 319–345.
Arslan, A.A., Gu, Y., Zeleniuch-Jacquotte, A., Koenig, K.L., Liu, M., Velikokhatnaya, L., Shore, R.E., Toniolo, P., Linkov, F., Lokshin, A.E., 2008. Reproducibility of serum pituitary hormones in women. Cancer Epidemiol. Biomark. Prev. 17, 1880–1883.
Bateup, H., Booth, A., Shirtcliff, E., Granger, D., 2002. Testosterone, cortisol, and women's competition. Evol. Hum. Behav. 23, 181–192.
Bernhardt, P.C., Dabbs, J.M., Fielden, J.A., Lutter, C.D., 1998. Testosterone changes during vicarious experiences of winning and losing among fans at sporting events. Physiol. Behav. 65, 59–62.
Bernstein, I.S., Rose, R.M., Gordon, T.P., 1974. Behavioral and environmental events influencing primate testosterone levels. J. Hum. Evol. 3, 517–525.
Booth, A., Shelley, G., Mazur, A., Tharp, G., Kittok, R., 1989. Testosterone, and winning and losing in human competition. Horm. Behav. 23, 556–571.
Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2009. Introduction to Meta-Analysis. John Wiley & Sons, Chichester, U.K.
Carré, J.M., Geniole, S.N., Ortiz, T.L., Bird, B.M., Videto, A., Bonin, P.L., 2016. Exogenous testosterone rapidly increases aggressive behavior in dominant and impulsive men. Biol. Psychiat. (in press).
Carré, J.M., Gilchrist, J.D., Morrissey, M.D., McCormick, C.M., 2010. Motivational and situational factors and the relationship between testosterone dynamics and human aggression during competition. Biol. Psychol. 84, 346–353.
Carré, J.M., McCormick, C.M., 2008. Aggressive behavior and change in salivary testosterone concentrations predict willingness to engage in a competitive task. Horm. Behav. 54, 403–409.
Carré, J.M., 2009. No place like home: testosterone responses to victory depend on game location. Am. J. Hum. Biol. 21, 392–394.
Carré, J.M., Olmstead, N.A., 2015. Social neuroendocrinology of human aggression: examining the role of competition-induced testosterone dynamics. Neuroscience 286, 171–186.
Carré, J.M., Putnam, S.K., 2010. Watching a previous victory produces an increase in testosterone among elite hockey players. Psychoneuroendocrinology 35, 475–479.
Carré, J.M., Campbell, J.A., Lozoya, E., Goetz, S.M., Welker, K.M., 2013. Changes in testosterone mediate the effect of winning on subsequent aggressive behaviour. Psychoneuroendocrinology 38, 2034–2041.
Carré, J.M., Iselin, A.M.R., Welker, K.M., Hariri, A.R., Dodge, K.A., 2014. Testosterone reactivity to provocation mediates the effect of early intervention on aggressive behavior. Psychol. Sci. 25, 1140–1146.
Carré, J.M., McCormick, C.M., Hariri, A.R., 2011. The social neuroendocrinology of human aggression. Psychoneuroendocrinology 36, 935–944.
Carré, J.M., Putnam, S.K., McCormick, C.M., 2009. Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. Psychoneuroendocrinology 343, 561–570.
Casto, K.V., Edwards, D.A., 2016. Testosterone and Reconciliation among Women: After-Competition Testosterone Predicts Prosocial Attitudes towards Opponents. Adapt. Hum. Behav. Physiol. 2, 220–223.
Casto, K.V., Elliott, C.M., Edwards, D.A., 2014. Intercollegiate cross country competition: effects of warm-up and racing on salivary levels of cortisol and testosterone. Int J Exer Sci 7, 318–328.
Costa, R., Salvador, A., 2012. Associations between success and failure in a face-to-face competition and psychobiological parameters in young women. Psychoneuroendocrinology 37, 1780–1790.
Cote, K.A., McCormick, C.M., Geniole, S.N., Renn, R.P., MacAulay, S.D., 2013. Sleep deprivation lowers reactive aggression and testosterone in men. Biol. Psychol. 92, 249–256.
Denson, T., Mehta, P., Ho Tan, D., 2013. Endogenous testosterone and cortisol jointly influence reactive aggression in women. Psychoneuroendocrinology 38, 416–424.
Duval, S., Tweedie, R., 2000. Trim and fill: a simple funnel- plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56 (2), 455–463.
Edwards, D., Wetzel, K., Wyner, D., 2006. Intercollegiate soccer: saliva cortisol and testosterone are elevated during competition, and testosterone is related to status and social connectedness with teammates. Physiol. Behav. 87, 135–143.
Edwards, D.A., 2006. Competition and testosterone. Horm. Behav. 50, 681–683.
Edwards, D.A., Casto, K.V., 2013. Women's intercollegiate athletic competition: cortisol, testosterone, and the dual-hormone hypothesis as it relates to status among teammates. Horm. Behav. 64, 153–160.
Edwards, D.A., Kurlander, L.S., 2010. Women's Intercollegiate Volleyball and Tennis:effects of warm-up, competition, and practice on saliva levels of cortisol and testosterone. Horm. Behav. 58, 606–613.
Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ (Clinical Research Ed.). 315 (7109), 629–634.
Elias, M., 1981. Serum cortisol, testosterone, and testosterone-binding globulin responses to competitive fighting in human males. Aggress. Behav. 7, 215–224.
Filaire, E., Maso, F., Sagnol, M., Ferrand, C., Lac, G., 2001. Anxiety, hormonal responses, and coping during a judo competition. Aggress. Behav. 27, 55–63.
Flinn, M.V., Ponzi, D., Muehlenbein, M.P., 2012. Hormonal mechanisms for regulation of aggression in human coalitions. Hum. Nat. 23, 68–88.
Fry, A., Schilling, B., Fleck, S., Kraemer, W., 2011. Relationships between competitive wrestling success and neuroendocrine responses. J. Strength. Cond. Res. 25, 40–45.
Geniole, S.N., Busseri, M.A., McCormick, C.M., 2013. Testosterone dynamics and psychopathic personality traits independently predict antagonistic behavior toward the perceived loser of a competitive interaction. Horm. Behav. 63, 790–798.
Geniole, S.N., Carré, J.M., McCormick, C.M., 2011. State, not trait, neuroendocrine function predicts costly reactive aggression in men after social exclusion and inclusion. Biol. Psychol. 87, 137–145.
Gladue, B., Boechler, M., McCaul, K., 1989. Hormonal response to competition in human Males. Aggress. Behav. 409–422.
Gonzalez-Bono, E., Salvador, A., Ricarte, J., Serrano, M.A., Arnedo, M., 2000. Testosterone and attribution of successful competition. Aggress. Behav. 26, 235–240.
Gonzalez-Bono, E., Salvador, A., Serrano, M.A., Ricarte, J., 1999. Testosterone, cortisol, and mood in a sports team competition. Horm. Behav. 35, 55–62.
Hamilton, L.D., van Anders, S.M., Cox, D., Watson, N.V., 2009. The effect of competition on salivary testosterone in elite female athletes. Int. J. Sports Phys. Perf. 4, 538–542.
Hasegawa, M., Toda, M., Morimoto, K., 2008. Changes in salivary physiological stress markers associated with winning and losing. Biomed. Res. 29, 43–46.
Higgins, J.P.T., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21 (11), 1539–1558.
Ioannidis, J.P.A., Trikalinos, T.A., 2007a. The appropriateness of asymmetry tests for publication bias in meta-analyses: A large survey. CMAJ 176 (8), 1091–1096.
Ioannidis, J.P.A., Trikalinos, T.A., 2007b. An exploratory test for an excess of significant findings. Clin Trials. 4, 245–253.
Jiménez, M., Aguilar, R., Alvero-Cruz, J., 2012. Effects of victory and defeat on testosterone and cortisol response to competition: Evidence for same response patterns in men and women. Psychoneuroendocrinology 37, 1577–1581.
Jones, A.C., Josephs, R.A., 2006. Interspecies hormonal interactions between man and the domestic dog (*Canis familiaris*). Horm. Behav. 50, 393–400.
Ketterson, E.D., Nolan, V., 1992. Hormones and life histories: an integrative approach. Am. Nat. 140, S33–S62.

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002

Kuhberger, A., Fritz, A., Scherndl, T., 2014. Publication bias in psychology: a diagnosis based on the correlation between effect size and sample size. PLoS One 99, e105825.
Lipsey, M.W., Wilson, D.B., 2001. Practical meta-analysis. 49. Sage Publications, Thousand Oaks, CA.
Maner, J.K., Miller, S.L., Schmidt, N.B., Eckel, L.A., 2008. Submitting to defeat: social anxiety, dominance threat, and decrements in testosterone. Psychol. Sci. 19, 764–768.
Mazur, A., 1976. Effects of testosterone on status in primate groups. Folia Primatol. 26, 214–226.
Mazur, A., 1985. A biosocial model of status in face-to-face primate groups. Soc. Forces. 64, 377–402.
Mazur, A., Lamb, T., 1980. Testosterone, status, and mood in human males. Horm. Behav. 14, 236–246.
Mazur, A., Booth, A., Dabbs, J., 1992. Testosterone and chess competition. Soc. Psychol. Q. 55, 70–77.
Mazur, A., Susman, E., Edelbrock, S., 1997. Sex differences in testosterone response to a video game contest. Evol. Hum. Behav. 18, 317–326.
McCaul, K., Gladue, B., Joppa, M., 1992. Winning, losing, mood, and testosterone. Horm. Behav. 26, 486–504.
Mehta, P.H., Snyder, N.A., Knight, E.L., Lassetter, B., 2015. Close versus decisive victory moderates the effect of testosterone change on competitive decisions and task enjoyment. Adapt. Hum. Behav. Phys. 3, 291–311.
Mehta, P.H., Josephs, R.A., 2006. Testosterone change after losing predicts the decision to compete again. Horm. Behav. 50, 684–692.
Mehta, P.H., van Son, V., Welker, K.M., Prasad, S., Sanfey, A.G., Smidts, A., Roelofs, K., 2015. Exogenous testosterone in women enhances and inhibits competitive decision-making depending on victory-defeat experience and trait dominance. Psychoneuroendocrinology 60, 224–236.
van der Meij, L., Buunk, A.P., Almela, M., Salvador, A., 2010. Testosterone responses to competition: the opponent's psychological state makes it challenging. Biol. Psychol. 84, 330–335.
Miller, S.L., Maner, J.K., McNulty, J.K., 2012. Adaptive attunement to the sex of individuals at a competition: the ratio of opposite- to same-sex individuals correlates with changes in competitor's testosterone levels. Evol. Hum. Behav. 33, 57–63.
Morris, S.B., DeShon, R.P., 2002. Combining effect size estimates in meta-analysis with repeated mea- sures and independent-groups designs. Psychol. Methods 7, 105–125.
Norman, R.E., Moreau, B.J.P., Welker, K.M., Carré, J.M., 2015. Trait anxiety moderates the relationship between testosterone responses to competition and aggressive behavior. Adapt Hum Behav. Physiol. 1, 312–324.
Oliveira, G.A., Uceda, S., Oliveira, T., Fernandes, A., Garcia-Marques, T., Oliveira, R.F., 2013. Threat perception and familiarity moderate the androgen response to competition in women. Front. Psychol. 4.
Oliveira, G.A., Uceda, S., Oliveira, T.F., Fernandes, A.C., Garcia-Marques, T., Oliveira, R.F., 2014. Testosterone response to competition in males is unrelated to opponent familiarity or threat appraisal. Front. Psychol. 5.
Oliveira, R.F., 2009. Social behavior in context: hormonal modulation of behavioral plasticity and social competence. Integr. Comp. Biol. 49, 423–440.
Oliveira, T., Gouveia, M.J., Oliveira, R.F., 2009. Testosterone responsiveness to winning and losing experiences in female soccer players. Psychoneuroendocrinology 34, 1056–1064.
Oxford, J., Ponzi, D., Geary, D., 2010. Hormonal responses differ when playing violent video games against an ingroup and outgroup. Evol. Hum. Behav. 31, 201–209.
Parmigiani, S., Bartolomucci, A., Palanza, P., Galli, P., Rizzi, N., Brain, P.F., Volpi, R., 2006. In Judo, Randori (Free Fight) and Kata (Highly Ritualized Fight) differentially change plasma cortisol, testosterone, and interleukin levels in male participants. Aggress. Behav. 32, 481–489.
Parmigiani, S., Dadomo, H., Bartolomucci, A., Brain, P.F., Carbucicchio, A., Costantino, C., Ferrari, P.F., Palanza, P., Volpi, R., 2009. Personality traits and endocrine response as possible asymmetry factors of agonistic outcome in karate athletes. Aggress. Behav. 35, 324–333.
Pesce, M., La Fratta, I., Ialenti, V., Patruno, A., Ferrone, A., Franceschelli, S., Rizzuto, A., Tatangelo, R., Campagna, G., Speranza, L., Felaco, M., Grilli, A., 2015. Emotions, immunity and sport: Winner and loser athlete's profile of fighting sport. Brain Behav. Immun. 46, 261–269.
Pound, N., Penton-Voak, I., Surridge, A., 2009. Testosterone responses to competition in men are related to facial masculinity. Proc. R. Soc. B 276, 153–159.
Rose, R.M., Bernstein, I.S., Gordon, T.P., 1975. Consequences of social conflict on plasma testosterone levels in rhesus monkeys. Psychosom. Med. 37, 50–61.
Rose, R.M., Gordon, T.P., Bernstein, I.S., 1972. Plasma testosterone levels in the male rhesus: influences of sexual and social stimuli. Science 178, 643–645.
Rosenthal, R., 1979. The file drawer problem and tolerance for null results. Psychol. Bull. 86, 638–641.
Salvador, A., Simon, V., Suay, F., Llorens, L., 1987. Testosterone and cortisol responses to competitive fighting in human males: A pilot study. Aggress. Behav. 13, 9–13.
Schultheiss, O.C., Campbell, K.L., McClelland, D.C., 1999. Implicit power motivation moderates men's testosterone responses to imagined and real dominance success. Horm. Behav. 36, 234–241.
Schultheiss, O.C., Rohde, W., 2002. Implicit power motivation predicts men's testosterone changes in implicit learning in a contest situation. Horm. Behav. 41, 195–202.
Schultheiss, O.C., Wirth, M.M., Torges, C.M., Pang, J.S., Villacorta, M.A., Welsh, K.M., 2005. Effects of implicit power motivation on men's and women's implicit learning and testosterone changes after social victory or defeat. J. Pers. Soc. Psychol. 88, 174–188.
Serrano, M., Salvador, A., Gonzalez-Bono, E., Sanchis, C., Suay, F., 2000. Hormonal responses to competition. Psicológica 12, 440–444.
Simonsohn, U., 2012. It does not follow: Evaluating the one-off publication bias critiques by Francis (2012a, 2012b, 2012c, 2012d, 2012e, in press). Perspect. Psychol. Sci. 7, 597–599.
Sobolewski, M.E., Brown, J.L., Mitani, J.C., 2013. Female parity, male aggression, and the Challenge Hypothesis. Primates 54, 81–88.
Sowers, M.F., Beebe, J.L., McConnell, D., Randolph, J., Jannausch, M., 2001. Testosterone concentrations in women aged 25–50 years: associations with lifestyle, body composition, and ovarian status. Am. J. Epidemiol. 153, 256–264.
Stanton, S.J., Beehner, J.C., Saini, E.K., Kuhn, C.M., LaBar, K.S., 2009. Dominance, politics, and physiology: voters' testosterone changes on the night of the 2008 United States Presidential Election. PLoS One 4, e7543.
Stanton, S.J., Schultheiss, O.C., 2007. Basal and dynamic relationships between implicit power motivation and estradiol in women. Horm. Behav. 52, 571–580.
Steiner, E.T., Barchard, K.A., Meana, M., Hadi, F., Gray, P.B., 2010. The deal on testosterone responses to poker competition. Curr. Psychol. 29, 45–51.
Suay, F., Salvador, A., Gonzalez-Bono, E., Sanchis, C., Martinez-Sanchis, S., Simon, V.M., Montoro, J.B., 1999. Effects of competition and its outcome on serum testosterone, cortisol, and prolactin. Psychoneuroendocrinology 24, 551–566.
Tibbetts, E.A., Crocker, K.C., 2014. The challenge hypothesis across taxa: Social modulation of hormone titres in vertebrates and insects. Anim. Behav. 92, 281–290.
Trumble, B.C., Cummings, D., Von Rueden, C., O' Connor, K.A., Smith, E.A., Gurven, M., Kaplan, H., 2012. Physical competition increases testosterone among Amazonian forager-horticulturalists: a test of the 'challenge hypothesis'. Proc. R. Soc. B 279, 2907–2912.
Trumble, B.C., Smith, E.A., O' Connor, K.A., Kaplan, H.S., Gurven, M.D., 2014. Successful hunting increases testosterone and cortisol in a subsistence population. Proc. R. Soc. B 281, 20132876.
Vingren, J.L., Kraemer, W.J., Ratamess, N.A., Anderson, J.M., Volek, J.S., Maresh, C.M., 2010. Testosterone physiology in resistance exercise and training. Sports Med. 40, 1037s–11053.
Welker, K.M., Carré, J.M., 2014. Individual differences in testosterone predict persistence in men. Eur. J. Personal. 29, 83–89.
Wilson, D.B., 2001. Effect size determination program. University of Maryland, College Park, MD .
Wingfield, J.C., Hegner, R.E., Dufty Jr., A.M., Ball, G.F., 1990. The "Challenge Hypothesis": theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. Amer. Nat. l36, 829–846.
Wingfield, J.C., Lynn, S.E., Soma, K.K., 2001. Avoiding the 'costs' of testosterone: ecological bases of hormone–behavior interactions. Brain Behav. Evol. 57, 239–251.
Wu, Y., Eisenegger, C., Zilioli, S., Watson, N.V., Clark, L., 2016. Comparison of clear and narrow outcomes on testosterone levels in social competition. Horm, Behav. (in press).
Zilioli, S., Mehta, P.H., Watson, N.V., 2014. Losing the battle but winning the war: uncertain outcomes reverse the usual effect of winning on testosterone. Biol. Psychol. 103, 54–62.
Zilioli, S., Watson, N.V., 2012. The hidden dimensions of the competition effect: basal cortisol and basal testosterone jointly predict changes in salivary testosterone after social victory in men. Psychoneuroendocrinology 37, 1855–1865.
Zilioli, S., Watson, N.V., 2014. Testosterone across successive competitions: evidence for a 'winner effect' in humans? Psychoneuroendocrinology 47, 1–9.

Please cite this article as: Geniole, S.N., et al., Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis, Horm. Behav. (2016), http://dx.doi.org/10.1016/j.yhbeh.2016.10.002