Hormones and Behavior 170 (2025) 105719

Contents lists available at ScienceDirect

Hormones and Behavior

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

Testing the congruency hypothesis using meta-analysis: Are changes in oral contraceptive use correlated with partnered women's sexual satisfaction?

Victor Shiramizu [a,*], Julia Stern [b], František Bartoš [c], Yasaman Rafiee [a], Thomas V. Pollet [d], Benedict C. Jones [a]

[a] Deparment of Psychological Sciences and Health, University of Strathclyde, UK
[b] Department of Psychology, University of Bremen, Germany
[c] Department of Psychological Methods, University of Amsterdam, Netherlands
[d] Department of Psychology, Northumbria University, UK

ARTICLE INFO

Keywords:
Oral contraceptives
Sexual satisfaction
Congruency hypothesis
meta-analysis

ABSTRACT

Based on claims that changes in women's hormone levels influence their mating psychology, the Congruency Hypothesis proposes that women in relationships who change their hormonal contraceptive use after meeting their partner will report lower sexual satisfaction with their partner than women who do not change their oral contraceptive use. However, findings from studies testing this hypothesis have reported mixed results. Consequently, we conducted a meta-analysis of published studies on this topic. For the thirteen effects sizes from studies using between-subjects designs, the pooled correlation between congruency and sexual satisfaction was not statistically significant ($r = 0.04$, $p = .091$) and a Robust Bayesian meta-analysis found that the null hypothesis was moderately favoured over the Congruency Hypothesis ($BF_{10} = 0.123$, i.e., $BF_{01} = 8.13$, $r = 0.003$). For the four effect sizes from studies using within-subjects designs, the pooled correlation between congruency and sexual satisfaction was statistically significant ($r = 0.18$, $p = .001$) and a Robust Bayesian meta-analysis found weak evidence in favour of the Congruency Hypothesis ($BF_{10} = 1.55$, $r = 0.09$). Although the effect of congruency on sexual satisfaction may be statistically significant in some analyses, across all analyses, results indicated that the magnitude of the congruency effect was small. Thus, we suggest that it is unlikely that changes in oral contraceptive use have a substantial (i.e., large) effect on women's sexual satisfaction on average. Still, some women might experience congruency effects and the overall evidence remains uncertain. More work, ideally in the form of randomized controlled trials, is needed to find a definite answer for research questions relying on the Congruency Hypothesis.

1. Introduction

>150 million women aged 15 to 49 years worldwide use oral contraceptives, with over 114 million of them being married (United Nations Department of Economic and Social Affairs, Population Division, 2022). Despite their introduction in the 1960s, most research has focused on the physical effects of oral contraceptives, but little is known about their effects on women's behavior and cognition (Gurvich et al., 2023; Montoya and Bos, 2017). Worryingly, misinformation and biased media coverage regarding the side effects of oral contraceptives on women's behavior are widespread on social media, impacting the contraceptive decision-making of adolescents and young women (Niemann et al., 2024; Shackleford et al., 2024). Thus, scientific investigations into the possible effects of oral contraceptives on women's behavior and cognition are urgently needed and are also of interest to the general public.

Among the most prominent findings of potential effects of hormonal contraceptives is that women's mate preferences have been reported to be influenced by changes in their conception probability, and, consequently, by hormonal contraceptives (see Gildersleeve et al., 2014 for a review). For example, studies of attractiveness judgments of men's faces, voices, bodies, and behavioral displays have reported that women's preferences for masculine men were increased during the ovulatory (i.e., high fertility) phase of the menstrual cycle (e.g., Feinberg et al., 2006; Gangestad et al., 2004; Little and Jones, 2012; Little et al., 2007; Penton-Voak et al., 1999; Puts, 2006). Similarly, studies have reported that

* Corresponding author.
E-mail address: victor.shiramizu@strath.ac.uk (V. Shiramizu).

https://doi.org/10.1016/j.yhbeh.2025.105719
Received 21 October 2024; Received in revised form 5 February 2025; Accepted 8 March 2025
Available online 15 March 2025
0018-506X/

women using oral contraceptives (which suppress ovulation) showed weaker preferences for masculine men than did women not using oral contraceptives (e.g., Feinberg et al., 2008; Little et al., 2002, 2013). Findings such as these have been widely interpreted as evidence for a positive effect of conception probability on women's preferences for masculine men, and it has been argued that hormonal contraceptives alter mate choice by interfering with natural mate preferences by removing the mate-preference shifts across the cycle (e.g., Alvergne and Lummaa, 2010).

Based on the findings described above, Roberts et al. (2014) predicted that women would report lower sexual satisfaction with their romantic partner when they had changed their use of oral contraceptives between initial partnership formation and testing (incongruent state) than when they had not changed their use of oral contraceptives between initial partnership formation and testing (congruent state). A test of this prediction in a sample of 365 women showed evidence for this Congruency Hypothesis (Roberts et al., 2014). This early work on the Congruency Hypothesis suggested that there may be risks associated with changing use of oral contraceptives during a romantic relationship. Potential risks that have been predicted were lower (sexual) satisfaction, or not finding the partner attractive anymore after changing the contraceptive method, potentially leading to separation (Alvergne and Lummaa, 2010; Roberts et al., 2012; Russell et al., 2014). It has even been proposed that incongruent contraceptive use could have an impact on women's health and reproduction, including their offspring's health and survival, through decreased genetic compatibility between both partners (Alvergne and Lummaa, 2010; Birnbaum et al., 2017), although there have been no direct longitudinal tests of this prediction.

However, two potentially serious problems with the Congruency Hypothesis have recently emerged in the literature. First, evidence from subsequent tests of the Congruency Hypothesis have produced mixed results. Although some studies have replicated the pattern of results Roberts et al. (2014) reported (Cobey et al., 2016; French and Meltzer, 2020), other large-scale studies have reported non-significant associations between sexual satisfaction and congruency of oral contraceptive use between partnership formation and testing (Botzet et al., 2021; Jern et al., 2018). Second, many recent studies testing for putative changes in women's preferences for masculine men either as a function of the menstrual cycle (Jones et al., 2018; Jünger et al., 2018a, 2018b; Marcinkowska et al., 2018; Stern et al., 2020, 2021; van Stein et al., 2019) or because of oral contraceptive use (Jones et al., 2018; Marcinkowska et al., 2019) have reported null results. These null results are noteworthy because they undermine a potential theoretical basis for the Congruency Hypothesis. Importantly, these studies reporting null results have typically addressed methodological concerns raised about the earlier work reporting significant effects of conception probability on masculinity preferences, such as small sample sizes, between-subjects designs, or imprecise methods to estimate fertility (for discussions of these and other methodological concerns see Blake et al., 2016, Gangestad et al., 2016, and Jones et al., 2019).

We note here, however, that there may be additional reasons to predict the congruency effect, not all of which are necessarily subject to all the concerns outlined above. For example, it has been argued that oral contraceptives have disruptive effects on olfactory cues to the assortment on the major histocompatibility complex (MHC). More precisely, women, as well as females in other species, usually prefer the odour of men that are dissimilar in genes in the MHC, potentially to increase the heterozygosity of offspring. This preference seems to be disrupted by oral contraceptives, as it has been reported that women taking the pill rather prefer MHC similar men, suggesting that oral contraceptives change preferences for MHC (Roberts et al., 2008). Consequently, oral contraceptives could change women's preferences for their partner's odour.

Because of the issues outlined above, the robustness and magnitude of the putative effect of changing oral contraceptive use during a relationship on women's sexual satisfaction with their partner is unclear. Consequently, we conducted a systematic literature search for all studies that investigated the Congruency Hypothesis. We then conducted a meta-analysis to test the hypothesized effect of the congruency between current use of oral contraceptives and their use during relationship formation on women's sexual satisfaction.

2. Methods

2.1. Literature search protocol

The effect of interest in our analysis was the association between women's sexual satisfaction and the congruency of oral contraceptive use between when they first met their current partner and at the time of study participation. Criteria for inclusion in our analyses were studies that tested the congruency of oral contraceptive use and sexual satisfaction and were available in the English language. Studies were identified through a literature search on Google Scholar, Web of Knowledge Core Collection and EBSCO Host (APA PsyArticles and APA PsycInfo databases). We also sent a call for any unpublished studies to the mailing lists of the Human Behavior and Evolution Society (HBES, December 14, 2023), the International Academy of Sex Research (IASR, December 14, 2023), the International Association for Relationship Research (December 14, 2023), and we also posted on X (formerly known as Twitter, December 14, 2023). No data were received before February 2, 2024. Full texts were searched for combinations of: ("hormonal contraception" OR "oral contraception") AND "sexual satisfaction" AND "congruency hypothesis". Searches included studies from inception of the database to September 25, 2023. A total of 51 records were extracted to ENDNOTE; 10 duplicates were removed; the first (VS) and last (BCJ) authors then conducted a screening process based on the titles and abstracts and 21 records were judged to be irrelevant (i.e., did not test the congruency hypothesis). Of the remaining 20 records, 7 were included in the analyses after conducting full-text screening (the excluded studies focused on relationship satisfaction or jealousy rather than sexual satisfaction). These 7 records reported a total of 17 effects (13 between-subjects effects and 4 within-subjects effects) with a total of 3834 participants. Fig. 1 presents a PRISMA flow diagram with the inclusion/exclusion of studies. Table 1 lists these studies and gives an overview of

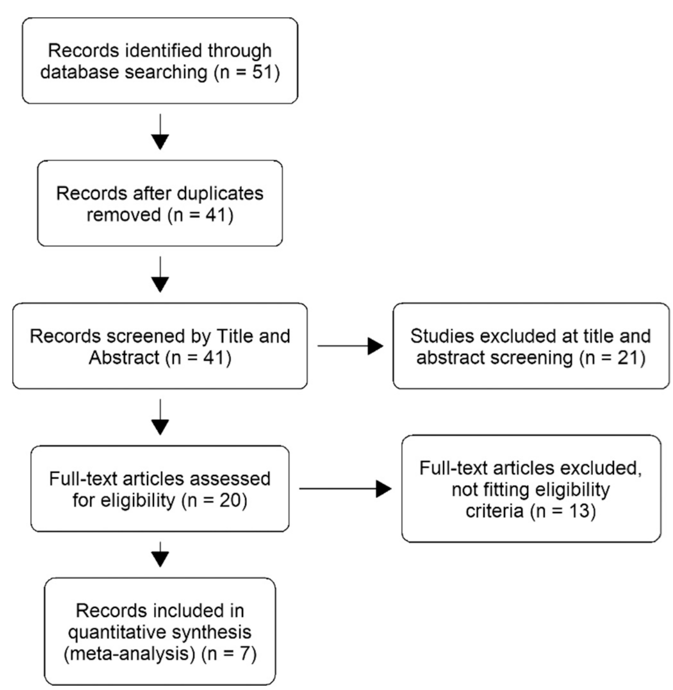

Fig. 1. PRISMA diagram summarising the flow of studies.

Table 1
Studies included in our meta-analyses, their sample sizes, proportion of congruent/incongruent hormonal contraceptive users, whether the sample included pregnant women, the type of sexual satisfaction measure used, whether the study was preregistered, the between-subjects effect size, and the within-subjects effect size.

Study	n	Proportion of congruent/ incongruent users	Included Pregnant Women?	Sexual satisfaction measure	Preregistered?	Between-subjects effect (*r* coefficient)	Within-subjects effect (*r* coefficient)
Roberts et al., 2014	365	CNHC = 30.7 %; CHC = 28.2 %; INHC = 31.5 %; IHC = 9.6 %	No	Single item	No	0.13	N/A
Russell et al., 2014 (Study 1)	48	Frequencies for all possible combinations not reported[a]	No	Standardized questionnaire (Index of Sexual Satisfaction)	No	−0.09	0.07
Russell et al., 2014 (Study 2)	70	Frequencies for all possible combinations not reported[b]	No	Single item	No	0.30	0.14
Cobey et al., 2016	84	44 % of women were using HC at relationship formation[c]	Yes	Single item	No	0.25	N/A
Jern et al., 2018 (Effect 1)	948	CNHC = 22.9 %; CHC = 35.2 %; INHC = 8.1 %; IHC = 33.8 %	No	Single item	No	0.02	N/A
Jern et al., 2018 (Effect 2)	948	CNHC = 22.9 %; CHC = 35.2 %; INHC = 8.1 %; IHC = 33.8 %	No	Standardized questionnaire (Female Sexual Function Inventory)	No	0.05	N/A
French and Meltzer, 2020 (Study 1)	109	Frequencies for all possible combinations not reported[d]	No	Standardized questionnaire (Index of Sexual Satisfaction)	No	−0.09	0.19
French and Meltzer, 2020 (Study 2)	94	Frequencies for all possible combinations not reported[d]	No	Standardized questionnaire (Index of Sexual Satisfaction)	No	0.02	0.27
Botzet et al., 2021	774	CNHC = 32 %; CHC = 31 %; INHC = 19 %; IHC = 17 %	No	Single item	No	0.05	N/A
Fiurašková et al., 2022 (Sample 1)	172	46 % of women using HC at relationship formation[c]	Yes	Standardized questionnaire (New Sexual Satisfaction Scale)	No	−0.03	N/A
Fiurašková et al., 2022 (Sample 2)	323	36 % of women were using HC at relationship formation[c]	Yes	Standardized questionnaire (New Sexual Satisfaction Scale)	No	0.12	N/A
Fiurašková et al., 2022 (Sample 3)	660	67 % of women were not using HC at relationship formation[e]	No	Standardized questionnaire (New Sexual Satisfaction Scale)	No	−0.06	N/A
Fiurašková et al., 2022 (Sample 4)	187	63 % of women were not using HC at relationship formation[e]	No	Standardized questionnaire (New Sexual Satisfaction Scale)	No	−0.06	N/A

Note: Within-subject effect sizes could only be calculated for the four studies that included a longitudinal component. CNHC: congruent non-hormonal contraception users; CHC: congruent contraception users; INHC: incongruent non-hormonal contraception users; IHC: incongruent hormonal contraception users.
[a] 33 % of women reported using HC at relationship formation, and 63 % of them discontinued HC at some point during the study.
[b] 53 % of women reported using HC at relationship formation, and 46 % of them discontinued HC at some point during the study.
[c] These women correspond to the congruent users as HCs levels mimic the hormonal state of pregnancy (Alvergne and Lummaa, 2010).
[d] Across both studies, 36 % of women became HC incongruent at some point, while 64 % remained HC congruent status throughout the study. Among them, 31 % reported using HC at relationship formation and at each assessment, while 79 % consistently did not use HC.
[e] These women correspond to the congruent users, as none of the participants were using HC during testing.

the study characteristics.

2.2. Variable coding

The sample characteristics coded for each study were (a) citation information, (b) sample size, and (c) relevant statistics and effect size information for analysis. If studies did not include sufficient information to compute effect sizes, authors were contacted, or the effect sizes were calculated based on publicly available data. All authors responded to our queries.

2.3. Analysis

All analyses were conducted in R 4.3.1 (R Core Team, 2023). Raw effect sizes for the effect of congruency on sexual satisfaction were calculated for each study and converted to *r* coefficients using the R package effectsize 0.8.3 (Ben-Shachar et al., 2020). For Roberts et al. (2014) and Cobey et al. (2016), raw effect sizes could be extracted directly from statistics reported in the papers. For Jern et al. (2018) and French and Meltzer (2020), we calculated raw effect sizes from their open datasets. For Russell et al. (2014), Botzet et al. (2021), and Fiurašková et al. (2022), raw effect sizes were provided by the authors of the original study because statistics reported in these papers did not include raw effect sizes and the datasets were not publicly available.

All effects were coded such that positive effects indicated that sexual satisfaction was higher in the congruent group (i.e., in women whose use of oral contraceptives was the same at relationship formation and testing) than in the incongruent group (i.e., in women whose use of oral contraceptives was different at relationship formation and testing). Prior to conducting the analysis, we applied a Fisher's *r*-to-*z* transformation to the extracted effect sizes, but the estimates were back-transformed to *r* when reporting in the text. Sampling variances and standard errors for the effect sizes were calculated using the R package esc 0.5.1 (Lüdecke, 2022).

To assess the relationship between congruency and sexual satisfaction, we conducted a meta-analysis with random-effects to account for other sources of variance not attributed to sampling error (Hedges and Vevea, 1998). The between-study variance component was estimated using the restricted maximum likelihood estimator, as simulation studies have shown that this particular estimator offers better properties for estimating such variances (Veroniki et al., 2016). Random-effects meta-analyses were conducted using the R package metafor 4.4.0 (Viechtbauer, 2023). Heterogeneity was assessed by Cochran's Q, I^2, τ, τ^2, and prediction interval statistics. Indications of publication bias were assessed via Egger's regression test (Egger et al., 1997), and also visually via a power-enhanced funnel plot, which further show the power of primary studies to detect an underlying overall effect estimate as a true effect (Kossmeier et al., 2020). We calculated the median statistical power across all included studies for the observed summary effect size using the R package metameta 0.2 (Quintana, 2023).

In addition, we carried out a Robust Bayesian Meta-Analysis (RoBMA). This method employs Bayesian model averaging to account for model uncertainty by simultaneously estimating 36 different models that can be categorized as (1) presence versus absence of the effect, (2) homogeneity versus heterogeneity, and (3) presence versus absence of publication bias (Bartoš et al., 2023; Maier et al., 2023). To quantify and interpret evidence from the Bayes factor, we used labels proposed by Lee and Wagenmakers (2013) (i.e., $1 < BF_{10} < 3$, $3 < BF_{10} < 10$, $10 < BF_{10}$ corresponds to weak, moderate, and strong evidence in favour of the effect respectively, with evidence in favour of the absence of the effect is obtained by inverting the BF, $BF_{01} = 1/BF_{10}$). RoBMA was conducted using the R package RoBMA 3.1.0 (Bartoš and Maier, 2020). All data, full outputs, and analysis code are publicly available on the Open Science Framework (https://osf.io/28pku/). The analyses document also includes additional analyses not reported fully here such as influence diagnostics analysis and additional robustness checks.

3. Results

Analyses included a total of 17 effects (13 between-subjects effects and 4 within-subjects effects) from 7 lab-groups with a total of 3834 participants.

3.1. Analyses of between-subjects effects

A random-effects meta-analytic model including only the 13 between-subjects effects showed a statistically non-significant pooled correlation ($r = 0.04$; 95 % CI [−0.01–0.09]; 95 % prediction interval [−0.08–0.17], $p = .091$). Fig. 2 shows a forest plot of the individual effect sizes and the pooled effect size. Moreover, there was statistically significant heterogeneity in the effects ($Q(12) = 26.53$, $p = .009$, $\tau_z = 0.06$). The proportion of heterogeneity not attributable to sampling error was 54.88 %. Due to the limited number of studies, we did not pursue moderator analyses to attempt to explain this heterogeneity.

A power-enhanced funnel plot revealed no asymmetry in the distribution of effects (Fig. 3). The Egger's regression test did not find evidence of publication bias ($z = 0.480$, $p = .631$). The median power of the included studies to detect the underlying effect ($r = 0.04$) was 8.7 %. Even when considering a true effect of $r = 0.13$ as reported by Roberts et al. (2014), the median power of included studies to detect this effect was only 42 %. Influence case diagnostics did not identify any potential outlier (see analysis output OSF).

RoBMA found moderate evidence against the congruency effect ($BF_{10} = 0.123$, $r = 0.003$, 95 % CI [< 0.001, 0.046]). Alternatively put, the null hypothesis is favoured over the alternative hypothesis of a congruency effect by a factor of 8.13, i.e. $BF_{01} = 1/0.123$. There was weak evidence in favour of heterogeneity ($BF_{rf} = 1.27$, $\tau_z = 0.034$, 95 % CI [< 0.001, 0.114]), and moderate evidence in favour of publication bias ($BF_{pb} = 3.20$).

Overall, not only did the frequentist meta-analysis fail to support a statistically significant effect, but the Bayesian meta-analyses showed moderate evidence in favour of the absence of the congruency effect within the 13 between-subjects estimates.

Three studies included pregnant women (Cobey et al., 2016; Fiurašková et al., 2022, Study 1; Fiurašková et al., 2022, Study 2). We repeated the analyses described above, this time excluding these three studies. Here, the random-effects meta-analysis showed a statistically non-significant pooled correlation ($r = 0.03$; 95 % CI [−0.02–0.08]; 95 % prediction interval [−0.09–0.14, $p = .267$]) and there was statistically significant heterogeneity across effect sizes ($Q(9) = 19.178$, $p = .024$, $\tau_z = 0.05$). Fig. 4 shows a forest plot of the individual effect sizes and the pooled effect size. The proportion of heterogeneity not attributable to sampling error was 51.86 %. The Egger's regression test showed no evidence of publication bias ($z = 0.0002$, $p = .999$). RoBMA found strong evidence against the congruency effect ($BF_{10} = 0.097$, $r = 0.002$, 95 % CI [< 0.001, 0.035]), weak evidence in favour of heterogeneity ($BF_{rf} = 1.42$, $\tau_z = 0.034$, 95 % CI [< 0.001, 0.113]), and weak evidence against publication bias ($BF_{pb} = 0.937$). The pattern of results in these analyses was the same as in our initial analyses.

3.2. Analyses of within-subjects effects

We analyzed the four within-subjects effects in the same way we had analyzed the 13 between-subjects effects. The random-effects meta-analysis here showed a statistically significant pooled correlation in line with the Congruency Hypothesis ($r = 0.18$; 95 % CI [0.07–0.29]; 95 %

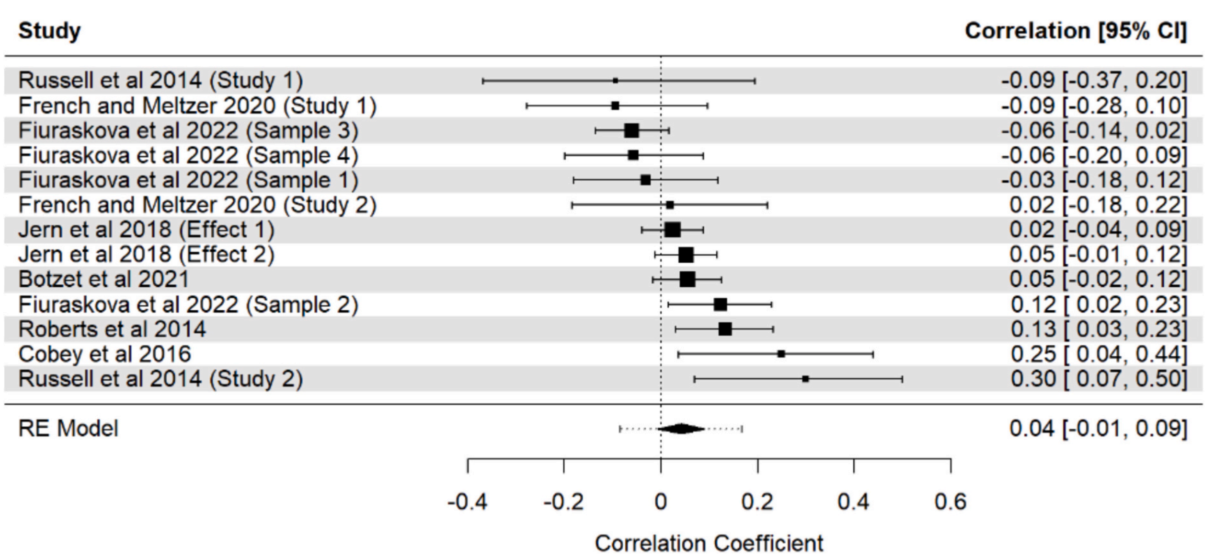

Fig. 2. Forest plot of the effect size estimates for the meta-analysis of the 13 between-subjects effects. The study's effect size is depicted by a square in the forest plot, with the square's size reflecting the weighting assigned to that effect in the meta-analysis. The whiskers illustrate the 95 % confidence interval for the effect size. At the bottom of the forest plot, the diamond indicates the estimated weighted mean effect size of the model, with the tips of the diamond indicating the 95 % confidence interval and whiskers representing the 95 % prediction interval.

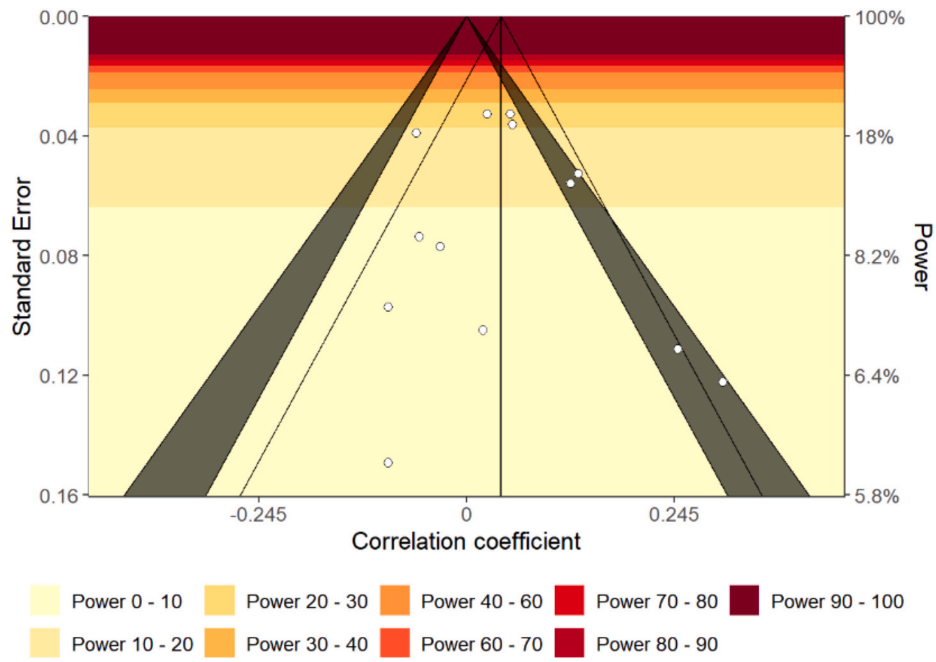

Fig. 3. Power-enhanced funnel plot of effect sizes, their standard errors, and the statistical power of primary studies to detect the underlying effect. Each dot represents one effect. The black zones indicate effects with *p*-values between.05 and 0.01. Different colors represent different ranges of statistical power.

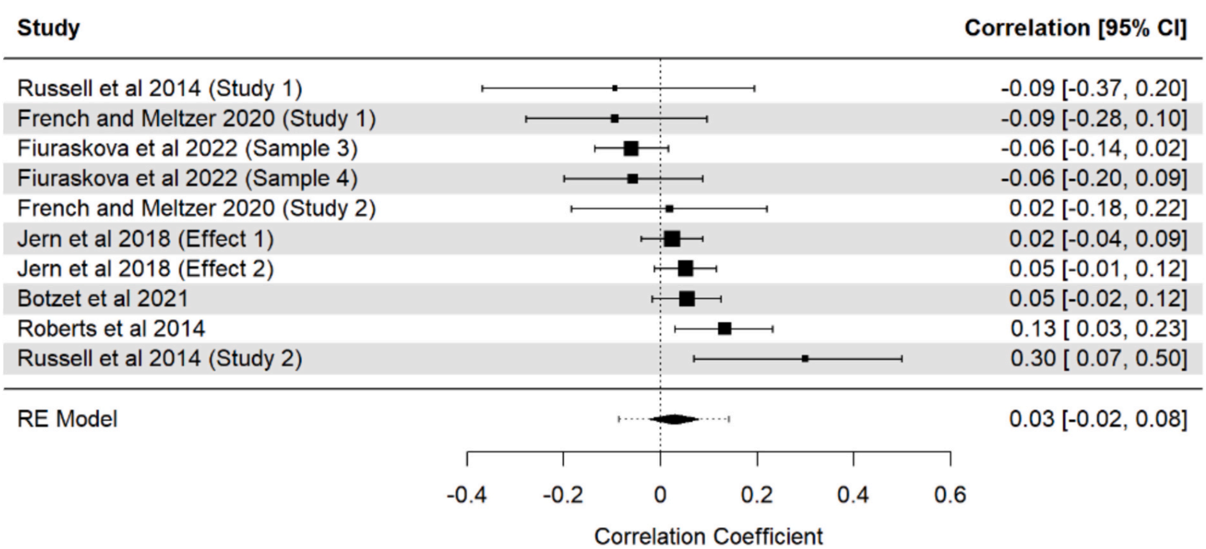

Fig. 4. Forest plot of the effect size estimates for the meta-analysis of the 10 between-subjects effects (i.e., the meta-analysis of between-subjects effects when the 3 studies including pregnant were excluded). The study's effect size is depicted by a square in the forest plot, with the square's size reflecting the weighting assigned to that effect in the meta-analysis. The whiskers illustrate the 95 % confidence interval for the effect size. At the bottom of the forest plot, the diamond indicates the estimated weighted mean effect size of the model, with the tips of the diamond indicating the 95 % confidence interval and whiskers representing the 95 % prediction interval.

prediction interval [0.07–0.29], $p = .001$) and there was no statistically significant heterogeneity across effect sizes ($Q(3) = 1.39$, $p = .708$, $\tau_z = 0$). Fig. 5 shows a forest plot of the individual effect sizes and the pooled effect size. The proportion of heterogeneity not attributable to sampling error was 0 %. A power-enhanced funnel plot did not reveal asymmetry in the distribution of effects (Fig. 6). The Egger's regression test did not find evidence of publication bias ($z = -0.907$, $p = .364$). The median power of the included studies to detect the underlying effect was 37.8 %. The RoBMA found weak evidence in favour of the congruency effect ($BF_{10} = 1.55$, $r = 0.093$, 95 % CI [< 0.001, 0.265]), weak evidence against heterogeneity ($BF_{rf} = 0.694$, $\tau_z = 0.038$, 95 % CI [< 0.001, 0.215]), and weak evidence in favour of publication bias ($BF_{pb} = 1.62$).

4. Discussion

We tested the hypothesized effect of the congruency between the current use of oral contraceptives and their use during relationship formation on women's sexual satisfaction using a meta-analytical approach. Meta-analyses of the effects from between-subjects studies found no evidence (frequentist approach) and moderate evidence of absence (Bayesian approach) for the congruency effect of oral contraceptive use on women's sexual satisfaction. Meta-analyses of the effects from within-subjects studies showed weak evidence for the Congruency Hypothesis, suggesting that more evidence is needed for forming a reliable conclusion. Collectively, these results challenge both the

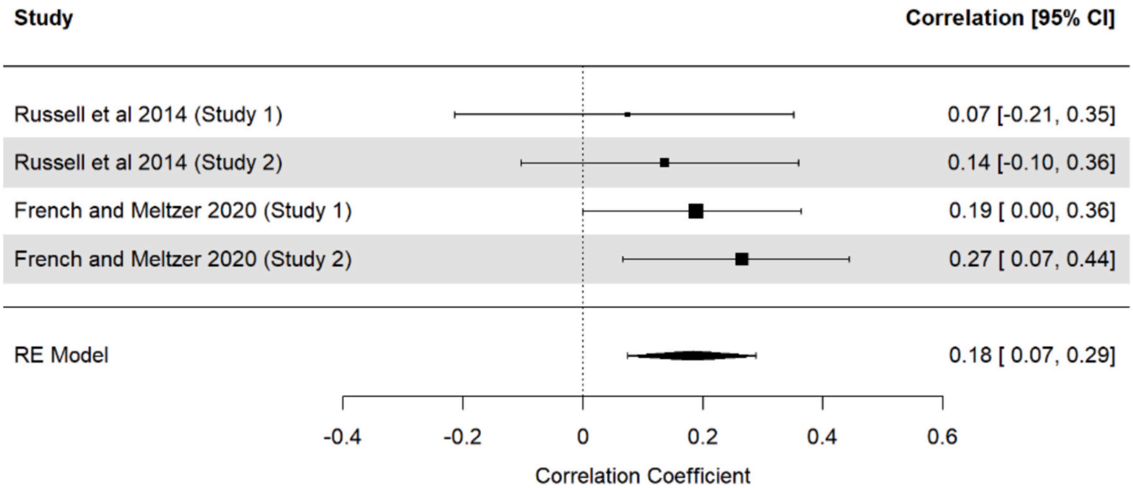

Fig. 5. Forest plot of the effect size estimates for the meta-analysis of the 4 within-subjects effects. The study's effect size is depicted by a square in the forest plot, with the square's size reflecting the weighting assigned to that effect in the meta-analysis. The branches illustrate the 95 % confidence interval for the effect size. At the bottom of the forest plot, the diamond indicates the estimated weighted mean effect size of the model, with the tip of the diamond indicating the 95 % confidence interval and whiskers represent the 95 % prediction interval.

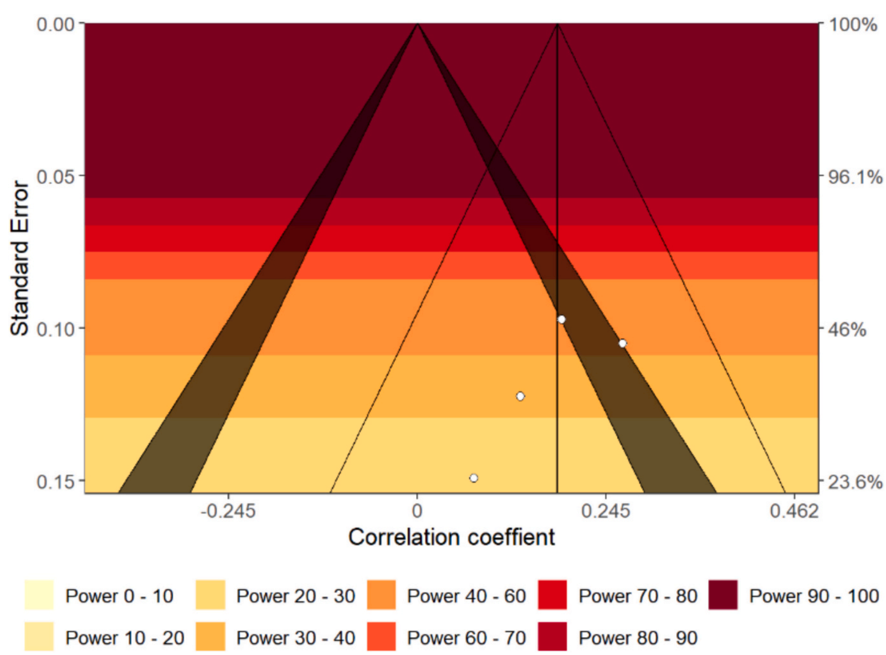

Fig. 6. Power-enhanced funnel plot of effect sizes, their standard errors, and the statistical power of primary studies to detect the underlying effect. Each dot represents one effect. The black zones indicate effects with *p*-values between.05 and 0.01. Different colors represent different ranges of statistical power.

Congruency Hypothesis and its theoretical assumptions grounded on the hormonal mechanisms underlying women's mating preferences, while also highlighting the need for sufficiently powered, longitudinal studies investigating the Congruency Hypothesis.

Results from the meta-analyses of the between-subject effects are in line with recent large studies testing the Congruency Hypothesis (Botzet et al., 2021; Jern et al., 2018), which were well-powered to detect the effect size reported by Roberts et al. (2014) ($r = 0.13$). However, when looking at the median power of the studies included in the current meta-analysis to detect the pooled effect size ($r = 0.04$), all studies were substantially underpowered (median power = 9 %) to find a small effect size that might not be of practical relevance (i.e., have a substantial effect on romantic relationships). More precisely, an $r = 0.04$ translates into Cohen's d = 0.08, meaning that 53.2 % of the women in the HC congruent state group will be above the mean of the incongruent state group, 96.8 % of the two groups will overlap, and there is a 52.2 % chance that a person selected at random from the HC congruent state group will have a higher score than a person selected at random from the HC incongruent state group. Therefore, previous studies reporting evidence supporting the Congruency Hypothesis may be due to randomness, publication bias (indeed, RoBMA results showed moderate evidence in favour of publication bias), or *p*-hacking, especially given that none of these studies was preregistered (Table 1). Nonetheless, we acknowledge that how effect sizes in survey studies translate into measurable real-world effects is not necessarily straightforward. Importantly, however, is possible that small effects of congruency actually reflect a main effect of HC use caused simply by an unbalanced distribution of participants across the four possible combinations of HC use at relationship formation and time of testing (as discussed in Jern et al., 2018). Consequently, we encourage researchers testing the congruency effect to carefully consider this issue (see Botzet et al., 2021 for a study that carefully controlled for this possibility).

We also found heterogeneity across effect sizes in the between-subjects effects, and most of this heterogeneity could be attributed to between-study differences ($I^2 = 55$ %). This could be partially explained by the measures of sexual satisfaction employed across studies, which varied from single items to different standardized questionnaires (Table 1). Therefore, future studies should focus on using questionnaires that have strong validity evidence to avoid questionable measurement practices (Flake and Fried, 2020). Rather than employing single measures of sexual satisfaction, future studies should consider combining different measures, submit them to a principal component analysis (PCA), and use the resulting scores in subsequent analysis. However, please note that quite large samples are needed for this approach.

The frequentist meta-analysis including the within-subjects effects showed a small and statistically significant pooled effect ($r = 0.18$) of the congruency between the current use of oral contraceptives and their use during relationship formation on women's sexual satisfaction. In the Bayesian analyses, the mean model-averaged estimate correlation was supported by weak evidence and shrank towards the null hypothesis ($r = 0.09$) after accounting for model uncertainty with RoBMA. Although there is some evidence in favour of the Congruency Hypothesis, the median power of the within-subject studies to detect the pooled effect size also suggested these analyses were underpowered (median power = 38 %). On the one hand, this means that the evidence for the Congruency Hypothesis is weak because previous studies did not have sufficient statistical power to detect the effect. On the other hand, it is possible that the true effect size is even smaller, as underpowered studies often overestimate effect sizes, leading to difficulties in replication, and undermining the reliability of findings (Button et al., 2013). Besides, an $r = 0.18$ translates into Cohen's d = 0.37, meaning that 64.3 % of the women in the HC congruent state group will be above the mean of the incongruent state group, 85.5 % of the two groups will overlap, and there is a 60.2 % chance that a person selected at random from the HC congruent state group will have a higher score than a person selected at random from the HC incongruent state group. Unfortunately, the within-subjects studies did not provide sufficient information on sample sizes of the different (in-) congruent groups to interpret the possibility that the reported effects are driven by main effects of hormonal contraceptives in general, rather than (in-) congruent use, as suggested by Jern et al. (2018). Given the small number of (non-preregistered) studies and the low power, it is unclear whether this proposed small effect is supported.

The causal assumption that hormonal contraceptives change or generally influence mate choice and sexual satisfaction could not be tested in previous studies based on correlational, often cross-sectional designs (as argued by e.g., Botzet et al., 2021). Longitudinal, within-subject designs are the designs of choice to investigate a research question about within-subject effects, such as the Congruency Hypothesis. Between-subject designs are likely to encounter more confounders (e.g., genetic confounding or recall error regarding their contraceptive method when meeting their partner). They further need much larger sample sizes to reduce unsystematic errors and achieve sufficient test power. Still, most studies to date investigating the Congruency Hypothesis are cross-sectional studies, and even the longitudinal ones cannot solve issues with causality (Rohrer, 2018), as there are still potential confounders that need to be considered (e.g., selection effects in congruency or reverse causality in that low sexual satisfaction leads to changing the contraceptive method). These issues should be kept in mind when interpreting results from studies investigating the Congruency Hypothesis. As highlighted in Table 1, none of the studies included in our meta-analysis were preregistered. Future studies with a within-subjects design are needed and, if such future studies are conducted, they should be adequately powered to detect the proposed small effect. Such studies would also benefit from pre-registration (Hardwicke and Wagenmakers, 2023; Nosek et al., 2018). As pointed out by Botzet et al. (2021), randomized controlled trials (RCTs) would be the most direct method for examining the causal effects of oral contraceptives on women's sexual satisfaction. Future research could also consider that there might be strong individual differences in the congruency effect. Although it remains unclear whether the effect is meaningful for the average women, some women might experience stronger and other weaker effects. Exploring factors that contribute to these individual differences might be an interesting direction for future research.

Overall, based on the current meta-analyses, the evidence to date suggests it is unlikely that contraceptive congruency has strong effects on sexual satisfaction. Still, it remains unclear whether the effect exists. Furthermore, we want to urge that there are currently no studies that allow for causal inference, although the Congruency Hypothesis clearly makes causal assumptions, and effects are often interpreted and communicated in a causal manner, including in popular science books and on social media.

5. Conclusion

Our results found little evidence for a congruence effect of oral contraceptives on women's sexual satisfaction. The evidence for the between-subjects analysis did not show a significant effect. Although the analysis of the within-subjects studies showed a statistically significant effect, the effect size was small. Thus, we suggest that it is unlikely that changing oral contraceptive use has a substantial (i.e., large) impact on women's sexual satisfaction with their romantic partner, given the small effect observed. Longitudinal studies controlling for confounders, and ideally randomized controlled trials, are needed to find a definite answer regarding the many research questions arising from the Congruency Hypothesis.

CRediT authorship contribution statement

Victor Shiramizu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Julia Stern:** Writing – review & editing, Writing – original draft, Validation, Conceptualization. **František Bartoš:** Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis. **Yasaman Rafiee:** Writing – review & editing, Writing – original draft. **Thomas V. Pollet:** Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis. **Benedict C. Jones:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Data availability

Data is available at https://osf.io/28pku/

References

Alvergne, A., Lummaa, V., 2010. Does the contraceptive pill alter mate choice in humans? Trends Ecol. Evol. 25 (3), 171–179. https://doi.org/10.1016/j.tree.2009.08.003.

Bartoš, F., Maier, M. (2020). "RoBMA: an R package for robust Bayesian Meta-analyses." R package version 3.1.0, https://CRAN.R-project.org/package=RoBMA.

Bartoš, F., Maier, M., Wagenmakers, E.J., Doucouliagos, H., Stanley, T.D., 2023. Robust Bayesian meta-analysis: model-averaging across complementary publication bias adjustment methods. Res. Synth. Methods 14 (1), 99–116. https://doi.org/10.1002/jrsm.1594.

Ben-Shachar, M., Lüdecke, D., Makowski, D., 2020. Effectsize: estimation of effect size indices and standardized parameters. Journal of Open-Source Software 5 (56), 2815. https://doi.org/10.21105/joss.02815.

Birnbaum, S., Birnbaum, G.E., Ein-Dor, T., 2017. Can contraceptive pill affect future offspring's health? The implications of using hormonal birth control for human evolution. Evol. Psychol. Sci. 3, 89–96. https://doi.org/10.1007/s40806-016-0074-4.

Blake, K.R., Dixson, B.J., O'Dean, S.M., Denson, T.F., 2016. Standardized protocols for characterizing women's fertility: a data-driven approach. Horm. Behav. 81, 74–83. https://doi.org/10.1016/j.yhbeh.2016.03.004.

Botzet, L.J., Gerlach, T.M., Driebe, J.C., Penke, L., Arslan, R.C., 2021. Hormonal contraception and sexuality: causal effects, unobserved selection, or reverse

causality?. *Collabra*. Psychology 7 (1), 29039. https://doi.org/10.1525/collabra.29039.
Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14 (5), 365–376. https://doi.org/10.1038/nrn3475.
Cobey, K.D., Havlíček, J., Klapilová, K., Roberts, S.C., 2016. Hormonal contraceptive use during relationship formation and sexual desire during pregnancy. Arch. Sex. Behav. 45, 2117–2122. https://doi.org/10.1007/s10508-015-0662-6.
Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315 (7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629.
Feinberg, D.R., Jones, B.C., Smith, M.L., Moore, F.R., DeBruine, L.M., Cornwell, R.E., Perrett, D.I., 2006. Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. Horm. Behav. 49 (2), 215–222. https://doi.org/10.1016/j.yhbeh.2005.07.004.
Feinberg, D.R., DeBruine, L.M., Jones, B.C., Little, A.C., 2008. Correlated preferences for men's facial and vocal masculinity. Evolution and Human Behavior 29 (4), 233–241. https://doi.org/10.1016/j.evolhumbehav.2007.12.008.
Fiurašková, K., Roberts, S.C., Kaňková, Š., Hlaváčová, J., Calda, P., Havlíček, J., 2022. Oral contraceptive use during relationship formation and current relationship satisfaction: testing the congruency hypothesis in couples attending pregnancy and fertility clinics. Psychoneuroendocrinology 135, 105451. https://doi.org/10.1016/j.psyneuen.2021.105451.
Flake, J.K., Fried, E.I., 2020. Measurement schmeasurement: questionable measurement practices and how to avoid them. Adv. Methods Pract. Psychol. Sci. 3 (4), 456–465. https://doi.org/10.1177/2515245920952393.
French, J.E., Meltzer, A.L., 2020. The implications of changing hormonal contraceptive use after relationship formation. Evol. Hum. Behav. 41 (4), 274–283. https://doi.org/10.1016/j.evolhumbehav.2020.04.003.
Gangestad, S.W., Simpson, J.A., Cousins, A.J., Garver-Apgar, C.E., Christensen, P.N., 2004. Women's preferences for male behavioral displays change across the menstrual cycle. Psychol. Sci. 15 (3), 203–207. https://doi.org/10.1111/j.0956-7976.2004.01503010.x.
Gangestad, S.W., Haselton, M.G., Welling, L.L., Gildersleeve, K., Pillsworth, E.G., Burriss, R.P., Puts, D.A., 2016. How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. Evolution and Human Behavior 37 (2), 85–96. https://doi.org/10.1016/j.evolhumbehav.2015.09.001.
Gildersleeve, K., Haselton, M.G., Fales, M.R., 2014. Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. Psychological Bulletin 140 (5), 1205–1259. https://doi.org/10.1037/a0035438.
Gurvich, C., Nicholls, I., Lavale, A., Kulkarni, J., 2023. Oral contraceptives and cognition: a systematic review. Front. Neuroendocrinol. 69, 101052. https://doi.org/10.1016/j.yfrne.2022.101052.
Hardwicke, T.E., Wagenmakers, E.-J., 2023. Reducing bias, increasing transparency and calibrating confidence with preregistration. Nat. Hum. Behav. 7 (1), 15–26. https://doi.org/10.1038/s41562-022-01497-2.
Hedges, L.V., Vevea, J.L., 1998. Fixed-and random-effects models in meta-analysis. Psychol. Methods 3 (4), 486–504. https://doi.org/10.1037/1082-989X.3.4.486.
Jern, P., Kärnä, A., Hujanen, J., Erlin, T., Gunst, A., Rautaheimo, H., Zietsch, B.P., 2018. A high-powered replication study finds no effect of starting or stopping hormonal contraceptive use on relationship quality. Evolution and Human Behavior 39 (4), 373–379. https://doi.org/10.1016/j.evolhumbehav.2018.02.008.
Jones, B.C., Hahn, A.C., Fisher, C.I., Wang, H., Kandrik, M., Han, C., DeBruine, L.M., 2018. No compelling evidence that preferences for facial masculinity track changes in women's hormonal status. Psychol. Sci. 29 (6), 996–1005. https://doi.org/10.1177/0956797618760197.
Jones, B.C., Hahn, A.C., DeBruine, L.M., 2019. Ovulation, sex hormones, and women's mating psychology. Trends Cogn. Sci. 23 (1), 51–62. https://doi.org/10.1016/j.tics.2018.10.008.
Jünger, J., Motta-Mena, N.V., Cardenas, R., Bailey, D., Rosenfield, K.A., Schild, C., Puts, D.A., 2018a. Do women's preferences for masculine voices shift across the ovulatory cycle? Horm. Behav. 106, 122–134. https://doi.org/10.1016/j.yhbeh.2018.10.008.
Jünger, J., Kordsmeyer, T.L., Gerlach, T.M., Penke, L., 2018b. Fertile women evaluate male bodies as more attractive, regardless of masculinity. Evolution and Human Behavior 39 (4), 412–423. https://doi.org/10.1016/j.evolhumbehav.2018.03.007.
Kossmeier, M., Tran, U.S., Voracek, M., 2020. Power-enhanced funnel plots for meta-analysis. Z. Psychol. 228 (1). https://doi.org/10.1027/2151-2604/a000392, 43-39.
Lee, M., Wagenmakers, E.J., 2013. Bayesian Data Analysis for Cognitive Science: A Practical Course. Cambridge University Press, New York, NY, USA.
Little, A.C., Jones, B.C., 2012. Variation in facial masculinity and symmetry preferences across the menstrual cycle is moderated by relationship context. Psychoneuroendocrinology 37 (7), 999–1008. https://doi.org/10.1016/j.psyneuen.2011.11.007.
Little, A.C., Jones, B.C., Penton-Voak, I.S., Burt, D.M., Perrett, D.I., 2002. Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. Proc. R. Soc. Lond. Ser. B Biol. Sci. 269 (1496), 1095–1100. https://doi.org/10.1098/rspb.2002.1984.
Little, A.C., Jones, B.C., Burriss, R.P., 2007. Preferences for masculinity in male bodies change across the menstrual cycle. Horm. Behav. 51 (5), 633–639.
Little, A.C., Burriss, R.P., Petrie, M., Jones, B.C., Roberts, S.C., 2013. Oral contraceptive use in women changes preferences for male facial masculinity and is associated with partner facial masculinity. Psychoneuroendocrinology 38 (9), 1777–1785. https://doi.org/10.1016/j.psyneuen.2013.02.014.
Lüdecke, D. (2022). Esc: effect size computation for Meta analysis (Version 0.5.1). doi: https://doi.org/10.5281/zenodo.1249218, https://CRAN.R-project.org/package=esc.
Maier, M., Bartoš, F., Wagenmakers, E.-J., 2023. Robust Bayesian meta-analysis: addressing publication bias with model-averaging. Psychol. Methods 28 (1), 107–122. https://doi.org/10.1037/met0000405.
Marcinkowska, U.M., Galbarczyk, A., Jasienska, G., 2018. La donna è mobile? Lack of cyclical shifts in facial symmetry, and facial and body masculinity preferences— A hormone based study. Psychoneuroendocrinology 88, 47–53.
Marcinkowska, U.M., Hahn, A.C., Little, A.C., DeBruine, L.M., Jones, B.C., 2019. No evidence that women using oral contraceptives have weaker preferences for masculine characteristics in men's faces. PloS One 14 (1), e0210162. https://doi.org/10.1371/journal.pone.0210162.
Montoya, E.R., Bos, P.A., 2017. How oral contraceptives impact social-emotional behavior and brain function. Trends Cogn. Sci. 21 (2), 125–136. https://doi.org/10.1016/j.tics.2016.11.005.
Niemann, J., Wicherski, L., Glaum, L., Schenk, L., Stadler, G., Richter, M., 2024. YouTube and the implementation and discontinuation of the oral contraceptive pill: a mixed-method content analysis. PloS One 19 (5), e0302316. https://doi.org/10.1371/journal.pone.0302316.
Nosek, B.A., Ebersole, C.R., DeHaven, A.C., Mellor, D.T., 2018. The preregistration revolution. Proc. Natl. Acad. Sci. 115 (11), 2600–2606. https://doi.org/10.1073/pnas.1708274114.
Penton-Voak, I.S., Perrett, D.I., Castles, D.L., Kobayashi, T., Burt, D.M., Murray, L.K., Minamisawa, R., 1999. Menstrual cycle alters face preference. Nature 399 (6738), 741–742. https://doi.org/10.1038/21557.
Puts, D.A., 2006. Cyclic variation in women's preferences for masculine traits: potential hormonal causes. Hum. Nat. 17, 114–127. https://doi.org/10.1007/s12110-006-1023-x.
Quintana, D.S., 2023. A guide for calculating study-level statistical power for Meta-analyses. *Advances in methods and practices*. Psychol. Sci. 6 (1). https://doi.org/10.1177/25152459221147260.
R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/>.
Roberts, S.C., Gosling, L.M., Carter, V., Petrie, M., 2008. MHC-correlated odour preferences in humans and the use of oral contraceptives. Proc. R. Soc. B Biol. Sci. 275 (1652), 2715–2722. https://doi.org/10.1098/rspb.2008.0825.
Roberts, S.C., Klapilová, K., Little, A.C., Burriss, R.P., Jones, B.C., DeBruine, L.M., Havlíček, J., 2012. Relationship satisfaction and outcome in women who meet their partner while using oral contraception. Proceedings of the Royal Society B: Biological Sciences 279 (1732), 1430–1436. https://doi.org/10.1098/rspb.2011.1647.
Roberts, S.C., Little, A.C., Burriss, R.P., Cobey, K.D., Klapilová, K., Havlíček, J., Petrie, M., 2014. Partner choice, relationship satisfaction, and oral contraception: the congruency hypothesis. Psychol. Sci. 25 (7), 1497–1503. https://doi.org/10.1177/0956797614532295.
Rohrer, J.M., 2018. Thinking clearly about correlations and causation: graphical causal models for observational data. Adv. Methods Pract. Psychol. Sci. 1 (1), 27–42. https://doi.org/10.1177/2515245917745629.
Russell, V.M., McNulty, J.K., Baker, L.R., Meltzer, A.L., 2014. The association between discontinuing hormonal contraceptives and wives' marital satisfaction depends on husbands' facial attractiveness. Proc. Natl. Acad. Sci. 111 (48), 17081–17086. https://doi.org/10.1073/pnas.141478411.
Shackleford, M., Horvath, A., Repetto, M., Thi, A., Twells, R., Sanders, M., Free, L., 2024. An analysis of oral contraceptive related videos on TikTok. AJOG Global Reports 4 (3), 100364. https://doi.org/10.1016/j.xagr.2024.100364.
Stern, J., Gerlach, T.M., Penke, L., 2020. Probing ovulatory-cycle shifts in women's preferences for men's behaviors. Psychol. Sci. 31 (4), 424–436. https://doi.org/10.1177/0956797619882022.
Stern, J., Kordsmeyer, T.L., Penke, L., 2021. A longitudinal evaluation of ovulatory cycle shifts in women's mate attraction and preferences. Horm. Behav. 128, 104916. https://doi.org/10.1016/j.yhbeh.2020.104916.
United Nations Department of Economic and Social Affairs, Population Division (2022). *World Family Planning 2022: Meeting the changing needs for family planning: Contraceptive use by age and method*. UN DESA/POP/2022/TR/NO. 4.
van Stein, K.R., Strauss, B., Brenk-Franz, K., 2019. Ovulatory shifts in sexual desire but not mate preferences: an LH-test-confirmed, longitudinal study. Evol. Psychol. 17, 1–10. https://doi.org/10.1177/1474704919848116.
Veroniki, A.A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Salanti, G., 2016. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res. Synth. Methods 7 (1), 55–79. https://doi.org/10.1002/jrsm.1164.
Viechtbauer, W., 2023. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36 (3), 1–48. https://doi.org/10.18637/jss.v036.i03.