Åbo Akademi

OPEN ACCESS

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

A high-powered replication study finds no effect of starting or stopping hormonal contraceptive use on relationship quality

Jern, Patrik; Kärnä, Antti; Hujanen, Janna; Erlin, Tatu; Gunst, Annika; Rautaheimo, Helmi; Öhman, Emilia; Roberts, Craig; Zietsch, Brendan

Published in:
Evolution and Human Behavior

DOI:
10.1016/j.evolhumbehav.2018.02.008

Published: 01/01/2018

Link to publication

Please cite the original version:
Jern, P., Kärnä, A., Hujanen, J., Erlin, T., Gunst, A., Rautaheimo, H., Öhman, E., Roberts, C., & Zietsch, B. (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

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This document is downloaded from the Research Information Portal of ÅAU: 25. Apr. 2025

Evolution and Human Behavior 39 (2018) 373–379

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

A high-powered replication study finds no effect of starting or stopping hormonal contraceptive use on relationship quality

Patrick Jern [a,b,*], Antti Kärnä [c], Janna Hujanen [b], Tatu Erlin [b], Annika Gunst [b], Helmi Rautaheimo [a], Emilia Öhman [a], S. Craig Roberts [d], Brendan P. Zietsch [e]

[a] Department of Psychology, Åbo Akademi University, FIN-20500 Turku, Finland
[b] Department of Psychology, University of Turku, FIN-20014 Turku, Finland
[c] Independent researcher, Turku, Finland
[d] Division of Psychology, University of Stirling, Stirling FK9 4LA, Scotland, UK
[e] School of Psychology, University of Queensland, Sir Fred Schonell Dr., St Lucia, QLD 4072, Australia

ARTICLE INFO

Article history:
Initial receipt 16 November 2016
26 February 2018
Final revision received 26 February 2018

Keywords:
Hormonal contraceptives
Relationship satisfaction
Jealousy
Sexual satisfaction
Physical attractiveness

ABSTRACT

A number of recent studies have implicated that incongruent use of hormonal contraceptives (HCs) negatively affects various aspects of women's romantic relationships. It has been suggested that women with incongruent HC use (a discrepancy in HC use status between when they first met their current partner and the time of study participation) report less sexual satisfaction and higher jealousy scores compared to women with congruent HC use. A similar effect has also been hypothesized for general relationship satisfaction, and recent findings suggest that the association between HC incongruency and women's general relationship satisfaction is moderated by third-party ratings of facial attractiveness of the women's male partners. Using a large convenience sample ($N = 948$) of Finnish women, we attempted to replicate previously reported findings but found no support for the HC congruency hypothesis, despite excellent statistical power (≥98.7%) to detect previously reported effect sizes. Instead, after dividing our sample into four groups based on HC congruency/incongruency, we found that the largest differences in jealousy, sexual satisfaction, and relationship satisfaction scores tended to be found between women who were consistent HC users and consistent non-users (i.e., between women with different kinds of congruent HC use). We also detected a significant main effect of current HC use on jealousy. We conclude that HC congruency effects reported in previous studies may have spuriously arisen from unequal distributions of current HC users within congruent and incongruent HC user groups.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Based on the premise that women's hormone levels affect their mate preferences (Gildersleeve, Haselton, & Fales, 2014; though see Jones et al., 2018a, in press), a recent line of research suggests that a change in the status of hormonal contraceptive (HC) use after initiation of a romantic relationship can have negative effects on the relationship. This is known as the HC congruency hypothesis (Roberts, Cobey, Klapilová, & Havlíček, 2014). The rationale of this hypothesis is that if HC use is congruent (i.e., the same at and after relationship initiation), women's desires and preferences remain more consistent than if HC use is incongruent (i.e. starting or discontinuing HC use after relationship initiation). The shift in preferences caused by incongruent HC use (and thus hormonal change) is thought to cause mismatch between preferences and the partner, in turn leading to relationship dissatisfaction.

In an early test the HC congruency hypothesis, women with congruent HC use were found to report lower levels of jealousy (Cobey, Roberts, & Buunk, 2013). This association was robust after controlling for main effects of HC use both at the time of study participation and at the beginning of the relationship, as well as potential confounds such as relationship satisfaction and duration. Subsequent studies have found support for the HC congruency hypothesis in other contexts. For example, women with congruent HC use have been shown to report higher levels of sexual satisfaction compared to women with incongruent HC use (Roberts et al., 2012; Roberts, Little, et al., 2014; Roberts, Cobey, et al., 2014; Russell, McNulty, Baker, & Meltzer, 2014). Furthermore, third-party ratings of the facial attractiveness of women's male spouses seem to predict a reduction of the negative impact of HC incongruency on general relationship satisfaction (Russell et al., 2014).

However, empirical tests of the HC congruency hypothesis that have been conducted to date have some limitations. Firstly, the sample sizes in studies directly testing the HC congruency hypothesis have been relatively small. For example, the largest published study involved 365 participants (Roberts, Little, et al., 2014); the first-ever study directly

* Corresponding author at: Department of Psychology, Åbo Akademi University, FIN-20500 Turku, Finland.
E-mail address: pjern@abo.fi (P. Jern).

https://doi.org/10.1016/j.evolhumbehav.2018.02.008
1090-5138/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

testing this hypothesis involved 121 participants (Cobey et al., 2013). This is a slight cause for concern, as the size of the congruency effect has been relatively small in all published studies. For example, in Cobey et al. (2013), the congruency effect was relatively modest at $R^2 = 0.03$. (Note: Cobey et al., 2013 did not report an effect size, but this can be calculated based on the *F*-statistic [$F = 3.96$] and the sample size [incongruent users, $n = 34$; congruent users, $n = 87$; total $N = 121$]). Secondly, there are actually two types of incongruent users (women using HCs at the beginning of the relationship, but not at present, and women not using HCs at the beginning of the relationship, but using at present) and, likewise, two types of congruent users (women who neither used HCs at the beginning of the relationship nor at present; and women using HCs at the beginning of the relationship and at present). It is possible that two subgroups of the same category, such as never- and always-users of HCs (who would both be considered congruent) could be markedly different from one another if exogenous hormones exert a main effect on the behavior in question. This potential confound would be more pronounced if one of these subgroups is more prevalent in one of the congruent/incongruent groups (e.g., if the incongruent but not the congruent group consists mainly of women who are currently using HCs). Thus far, the studies testing the HC congruency hypothesis have studied only the differences between two groups: congruent and incongruent users (Cobey et al., 2013; Roberts, Little, et al., 2014; Russell et al., 2014). In one study (Roberts et al., 2012), all participants were non-users of HCs at the time of study participation, and thus statistical comparisons were conducted based on HC use status at the initiation of the relationship only. Thirdly, in some instances (e.g., Roberts, Little, et al., 2014), non-validated measures have been used to quantify behaviors and attitudes that have been hypothesized to be associated with HC congruency status.

Recently, investigators and scientific journals have raised concerns regarding the low reproducibility of novel research findings (e.g., Open Science Collaboration, 2015), and many have called for thorough direct replication studies in order to avoid publication bias and correct erroneous novel reports (e.g., Ioannidis, 2012; Pashler & Harris, 2012). Indeed, recent studies using large samples and robust methodology have not replicated the effect of hormonal changes on mate preferences (e.g., Jones et al., in press), thus undermining the very premise of the HC congruency hypothesis. Thus, the aim of the present study was to conduct a thorough and high-powered test of the HC congruency hypothesis, while addressing some of the limitations of previous research. Furthermore, recent evidence suggests that the association between hormonal fluctuations and women's mate preferences is not as pronounced as previously thought (Jones et al., 2018a, in press). Therefore, we investigate alternative explanations for previous findings relating to the HC incongruency hypothesis in which we compare not only congruent and incongruent HC users, but also take into account possible main effects of present HC use vs. non-use.

2. Materials and methods

2.1. Participants

Statistical analyses were based on $N = 948$ Finnish women ($M_{age} = 23.95$ years, $SD = 4.16$). In the spring of 2015, we targeted women aged 18 or above who were in steady relationships to participate in the present study. Recruitment was done through university email lists at five different universities or polytechnic schools in Turku and Helsinki, Finland, as well as through social media (Facebook, Twitter). All data were collected through a secure, online questionnaire. We kept the survey open for a period of four weeks (the time frame was decided before commencing the data collection). In total, 1590 individuals logged on to the survey, and of these, 468 did not answer any of the questions relevant to the present study.

The online survey was constructed so that participation would terminate if the participant stated that 1) his/her gender was not female (30 individuals); 2) she was not in a relationship at present (33 individuals); 3) she was in a non-heterosexual relationship (29 participants); or 4) she was undergoing any sort of hormonal treatment, such as fertility treatment (10 individuals). Finally, we excluded women who were pregnant, trying to conceive, or uncertain if they were pregnant or not (66 individuals). Six individuals were uncertain whether they had used HCs or not when they first met their partner (and could thus not be allocated to congruent or incongruent HC user groups for statistical analyses), leaving us with a final study sample of 948 women. All participants provided written, informed consent. The research plan was approved by the Ethics Committee of Åbo Akademi University (Turku, Finland) in accordance with the Helsinki Declaration.

2.2. Measures

All measures were back-translated into Finnish and Swedish. The final translations were reviewed by a panel of seven individuals with excellent command of English as well as Finnish and/or Swedish.

2.2.1. Use of hormonal contraceptives

Participants responded to two dichotomous queries: *"Did you use any hormone-based contraceptives when you first began dating your current partner?"* [yes/no], and *"Do you use any hormone-based contraceptives at the moment?"* [yes/no]. In addition, we inquired about type and brand of HC, and for how long they had been using their current contraceptive method (if any).

2.2.2. General relationship satisfaction

In order to measure general relationship satisfaction, we used two different measures. The first was the same measure used by Roberts, Little, et al. (2014), a composite variable consisting of two questions (*"How satisfied are you with your partner's financial provision/intelligence?"*), with response alternatives on a 9-point Likert scale with the anchors "my partner does not satisfy me at all on this condition" and "my partner completely satisfies me on this condition". The internal consistency of this composite variable was modest (Cronbach's $\alpha = 0.419$). The second was the *Perceived Relationship Quality Components* (PRQC; Fletcher, Simpson, & Thomas, 2000a), which has been shown to have good internal consistency and predictive validity in previous studies (Fletcher et al., 2000a; Fletcher, Simpson, & Thomas, 2000b). The PRQC consists of six questions (e.g., *"How satisfied are you with your relationship?"*) with response alternatives on a 7-point Likert scale with the anchors "not at all" and "extremely". In the present study, the internal consistency of this measure was good (Cronbach's $\alpha = 0.858$). The (Roberts, Little, et al., 2014) measures of general relationship satisfaction had not previously been validated. The correlation between this measure and the PRQC was modest ($r = 0.342$, $p < 0.001$).

2.2.3. Sexual satisfaction

In similar fashion, we used two different measures of sexual satisfaction: first, a composite variable consisting of a question used by Roberts, Little, et al. (2014) (e.g. *"How satisfied are you with your partner's sexual adventurousness/ability to arouse you sexually?"*), with the same response alternatives as above. The internal consistency of this composite variable was fairly good (Cronbach's $\alpha = 0.770$). Second, we used the sexual satisfaction subscale of the *Female Sexual Function Inventory* (FSFI; Rosen et al., 2000), which consists of three items (e.g., *"Over the past four weeks, how satisfied have you been with your sexual relationship with your partner?"*) responded to on a five-point Likert scale with the anchors "very dissatisfied" and "very satisfied". The FSFI has been found to possess good validity in previous studies (Rosen et al., 2000). The FSFI measures sexual function and satisfaction based on the past four weeks. Individuals who had not engaged in partnered sexual activity in the past four weeks were excluded from analyses using the FSFI sexual satisfaction domain (50 individuals; 5.3%). This composite variable had good internal consistency (Cronbach's $\alpha = 0.860$). Again, the

(Roberts, Little, et al., 2014) measures of sexual satisfaction had not previously been validated. The correlation between this measure and the sexual satisfaction subdomain of the FSFI was moderate ($r = 0.496$, $p < .001$).

2.2.4. Jealousy

To measure jealousy, we used the same basic measure that was used by Cobey et al. (2013) (adapted from Buunk, 1997). This measure includes 15 questions (e.g., *"How would you feel if your partner would dance intimately with someone of the opposite sex?"*) with responses recorded on a 9-point scale, with higher scores indicating higher levels of jealousy. The internal consistency of this composite variable was good (Cronbach's $\alpha = 0.876$).

2.2.5. Partner attractiveness

We used the following two questions to form a composite variable measuring how attractive the participants perceived their male partners to be: *"Compared to other men, how attractive do you think your partner's face and body is?"* and *"How satisfied are you with your partner's physical attractiveness?"* Both items were responded to on a 7-point scale, with higher scores indicating higher perceived attractiveness. The internal consistency of this composite variable was fairly good (Cronbach's $\alpha = 0.787$).

2.2.6. Control measures

Based on the studies conducted by Roberts, Little, et al. (2014) and Cobey et al. (2013), we collected data for a number of covariates to be able to check that any detected associations would remain robust after controlling for these. These were: household income (9-point scale with the anchors "no income" and "more than €100,000 per year"), relationship duration, age, and whether the participants had children with their current partner.

2.3. Statistical analyses

The SPSS 21.0 software package was used for all analyses. Post hoc statistical power was computed based on previously reported effect sizes, the number of predictors in the model, the sample size for the present analysis, and the alpha level (set at 0.05 in all cases). The General Linear Model module was used to compute main effects of congruency on general relationship satisfaction, sexual satisfaction, and jealousy; and interaction effects of perceived facial attractiveness on the association between HC congruency and general relationship satisfaction. Congruency was inserted as a fixed factor, and perceived facial attractiveness as a covariate in the interaction analysis. Confidence intervals for effect sizes were calculated using R2 (R2 user manual retrieved from http://www.statpower.net/Content/R2/R2%20Manual.pdf, see also Steiger & Fouladi, 1997).

3. Results

3.1. Descriptive statistics

In the present study, 334 women (35.2%) reported that they had been using HCs both when they first started dating their partner and at the time of participating in the study, and 217 women (22.9%) reported that they did not use HCs at either time point. These women formed the HC congruent group ($n = 551$, 58.1%). Seventy-seven women (8.1%) reported that they had used HCs when they first started dating but no longer did so at the time of study participation, and 320 women (33.8%) reported that they had not been using HCs when they first started dating their current partner, but were using HCs currently. Thus, the HC incongruent group consisted of 397 women (41.9%). Of the women who had used HCs when they first started dating their current partner, 326 (79.3%) reported having used combined oral contraceptives at the time, while 29 (7.1%) reported having used progestin-only pills. Fifty-six women (13.6%) reported having used various non-oral HCs such as implants, patches, intrauterine devices or vaginal rings. Women who were currently using HCs reported their use as follows: combined pills (432 women, 65.6%); progestin-only pills (65 women, 9.9%); non-oral HCs (162 women, 24.6%). The average relationship duration was 3.46 years ($SD = 3.20$). Fifty women (5.2%) had children with their current partner, and the average household income was €25,000–€37,499.

3.1.1. Statistical power

We chose to base the power calculation for our replication of the association between sexual satisfaction and HC congruency on Roberts, Little, et al. (2014) study, as this study had a considerably larger sample size than that of Russell et al. (2014). The previously reported effect size for this association was $\eta^2_{partial} = 0.018$. Our sample size of 948 women thus provided 98.74% power to detect an association with this effect size (with one predictor in the model and the α level set at 0.05). In the study investigating associations between HC congruency and jealousy (Cobey et al., 2013), no effect sizes were reported. Therefore, we calculated the effect size from the reported F statistic ($F = 3.96$) and the sample size ($N = 129$), giving $d = 0.35$ (95% CI: 0.003–0.698). Next, we converted the effect size d to R^2 (Rosenthal, 1994), giving $R^2 = 0.03$. In this case, we had 99.97% power to detect this effect size in our sample ($n = 944$ for these items), again with one predictor in the model and α set at 0.05. For the moderating effect of partner's facial attractiveness on the association between marital satisfaction and HC congruency, Russell et al. (2014) reported an effect size of $r = 0.14$ for the combined sample of two studies (this effect size was smaller than the effect sizes reported for either of the individual studies), giving $R^2 = 0.02$. Thus, we had 99.22% power to detect an effect of this magnitude in our sample.

3.1.2. Main effects of HC contraceptive use

First, we report results from comparisons between users and non-users of HCs at the two time points (past use refers to retrospective recollections of when the participants first met their partners). Both present ($F = 4.993$, $p = .027$, $\eta^2_{partial} = 0.005$) and past ($F = 4.927$, $p = 0.026$, $\eta^2_{partial} = 0.005$) HC use predicted higher sexual satisfaction compared to non-use using Roberts, Little, et al. (2014) measure, but no effects of HCs on sexual satisfaction were detected for the sexual satisfaction dimension of the Female Sexual Function Index (FSFI) measure ($ps > .074$, $\eta^2_{partials} < 0.003$). Present ($F = 5.651$, $p = .018$, $\eta^2_{partial} = 0.006$), but not past HC use ($p = .435$, $\eta^2_{partial} = 0.001$) predicted higher general relationship satisfaction as measured with the Perceived Relationship Quality Components (PRQC) questionnaire, however, this effect could not be replicated with Roberts, Little, et al. (2014) measure of general relationship satisfaction ($ps > .452$, $\eta^2_{partials} < 0.002$). Present ($F = 13.798$, $p < .001$, $\eta^2_{partial} = 0.014$), but not past use of HCs ($p = .178$, $\eta^2_{partial} = 0.002$) predicted higher jealousy scores, suggesting a main effect of HCs on jealousy.

3.2. Replications of previously reported effects of HC congruency on sexual satisfaction, general relationship satisfaction, and jealousy

3.2.1. Testing the HC congruency hypothesis

Note that in these analyses, the HC congruent groups included women who had used HCs both when they first began dating their partner and at the time of participation in the study, as well as women who had not used HCs at either time point. The HC incongruent group included women who were using HCs when they began dating their partner but did not use HCs at the time of participation in the study, and vice versa. Contrary to previous studies' findings, no main effects of HC congruency status were detected for any of the dependent variables (Fig. 1), and effect sizes were invariably close to zero (Table 1). Despite using exactly the same measures as used in all previous studies except one (Russell et al., 2014) as well as additional, well-validated alternative measures, we were unable

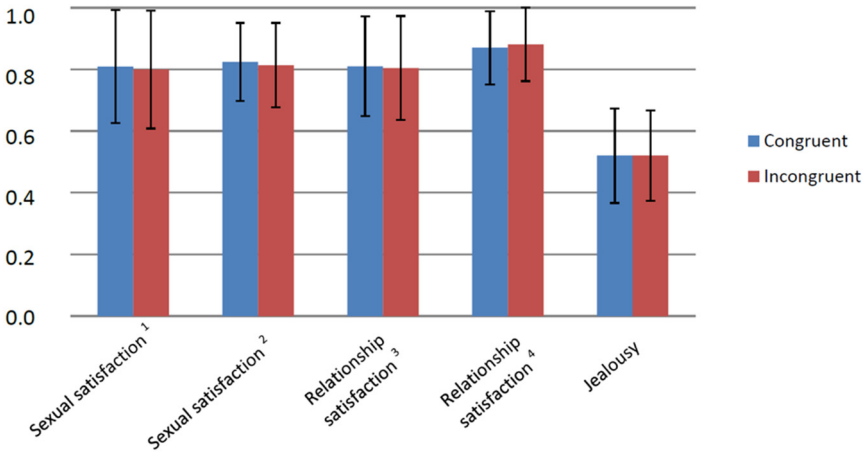

Fig. 1. Means and standard deviations (error bars) for the congruent and incongruent hormonal contraceptive use (incongruent users have either began or stopped using contraceptives after the initiation of the relationship). There were no significant group differences for any of the five measures. All measures have been converted to a scale from 0 to 1 for clarity. [1] Roberts, Little, et al. (2014) measure of sexual satisfaction; [2] the sexual satisfaction domain of the Female Sexual Function Inventory (Rosen et al., 2000); [3] Roberts, Little, et al. (2014) measure of general relationship satisfaction; [4] Perceived Relationship Quality Components (Fletcher et al., 2000a).

to confirm the results concerning sexual satisfaction reported by Roberts, Little, et al. (2014) and Russell et al. (2014), or the results concerning jealousy reported by Cobey et al. (2013).

3.3. Replication of previously reported moderating effects of perceived partner facial attractiveness on the association between HC congruency and general relationship satisfaction

Russell et al. (2014) reported a moderating effect of perceived facial attractiveness of male partners on the association between HC congruency and marital satisfaction. In the present study, no data were available regarding facial attractiveness specifically. Instead, we used scores of the partner's overall physical (face/body) attractiveness as reported by the women participating in the present study. Reported partner attractiveness had a main effect of small-to-moderate size on general relationship satisfaction (PRQC: $F = 211.821$, $p < .001$, $\eta^2_{partial} = 0.189$; Roberts, Little, et al. (2014) measure: $F = 74.023$, $p < .001$, $\eta^2_{partial} = 0.075$). We could not, however, replicate the significant moderating effect of the partner's attractiveness on the association between HC congruency and relationship satisfaction (PRQC: $F = 0.482$, $p = .488$; $\eta^2_{partial} = 0.001$, 95% C.I. = 0.000–0.008). The result was similar when Roberts, Little, et al. (2014) measure of general relationship satisfaction was used as the dependent variable ($F = 0.033$, $p = .855$; $\eta^2_{partial} < 0.001$, 95% C.I. = 0.000–0.004).

3.4. Comparisons between four different subgroups based on HC congruency/incongruency status

Next, we reran all analyses (i.e., two measures of sexual satisfaction, two measures of general relationship satisfaction, and jealousy), now comparing the four possible groups of HC congruency: women who 1) used HCs both when they first started dating their partners and currently (congruent); 2) used HCs when they first started dating their partners but not currently (incongruent); 3) did not use HCs when they first started dating their partners but used HCs currently (incongruent); and 4) did not use HCs at either time point (congruent). That is, one variable with the aforementioned four levels of HC congruency was inserted as a predictor in the model, with pairwise comparisons between all variable levels (i.e., so that each of the four possible subgroups were tested for mean differences against the other three groups). As can be seen in Table 2 and Fig. 2, at least one congruent group was statistically indistinguishable from at least one incongruent group in all analyses, and both incongruent groups were indistinguishable from one another. However, for all dependent variables except two (Roberts, Little, et al. (2014) measure of general relationship satisfaction and the sexual satisfaction domain of the FSFI, for which there were no significant differences between any of the four groups), the two congruent groups were significantly different from one another (*p*s ranging from .006–.00004). In most cases, the differences were such that the two congruent group means represented the highest and lowest of the four group means. Women who were consistent HC users reported the highest scores on both measures of sexual satisfaction, and also the highest score on the jealousy measure.

3.5. Additional analyses

As a control analysis for the replication of the results of Roberts, Little, et al. (2014) results regarding sexual satisfaction, we reran the analyses by controlling for the effects of the following variables by including them as covariates: general relationship satisfaction, household

Table 1
Associations between hormonal contraceptive congruency and sexual satisfaction, general relationship satisfaction, and jealousy.

Measure	M[a]	SD	ΔM[b]	95% C.I. (M)	B	SE	95% C.I. (B)	p	$\eta^2_{partial}$	95% C.I. ($\eta^2_{partial}$)[c]
Sexual satisfaction (Roberts, Little, et al., 2014)	7.244	1.681	0.08	7.136, 7.353	0.083	0.111	−0.134, 0.300	.453	0.001	0.000, 0.008
Sexual satisfaction (FSFI)	3.918	0.913	0.10	3.859, 3.977	0.097	0.060	−0.021, 0.214	.108	0.003	0.000, 0.013
General relationship satisfaction (Roberts, Little, et al., 2014)	7.265	1.480	0.05	7.170, 7.361	0.050	0.097	−0.141, 0.241	.609	<0.001	0.000, 0.006
General relationship satisfaction (PRQC)	6.132	0.833	−0.08	6.078, 6.185	−0.079	0.055	−0.186, 0.029	.152	0.002	0.000, 0.012
Jealousy (Cobey et al., 2013)	61.924	17.940	−0.01	60.762, 63.087	−0.009	1.184	−2.333, 2.315	.994	<0.001	0.000, 0.005

Note. FSFI = Female Sexual Function Index (Rosen et al., 2000); PRQC = Perceived Relationship Quality Components (Fletcher et al., 2000a); *SD* = standard deviation; *SE* = standard error; *M* = mean.
[a] Grand mean.
[b] Difference in means between congruent and incongruent hormonal contraceptive user groups. Positive values indicate that the mean was higher in the group of congruent users.
[c] Lower bound of effect size confidence intervals is zero.

Table 2
Means, standard deviations and 95% confidence intervals for four groups of hormonal contraceptive congruency.

Measure	Congruency group	*M*	*SD*	95% C.I.
Sexual satisfaction (Roberts, Little, et al., 2014)	Congruent_A	7.488	1.562	7.308, 7.668
	Incongruent_A	7.156	1.887	6.782, 7.530
	Incongruent_B	7.214	1.687	7.031, 7.398
	Congruent_B	6.975	1.735	6.752, 7.197
Sexual satisfaction (FSFI)	Congruent_A	4.049	0.730	3.952, 4.147
	Incongruent_A	3.796	1.125	3.593, 4.000
	Incongruent_B	3.887	0.948	3.787, 3.987
	Congruent_B	3.838	1.010	3.717, 3.959
Relationship satisfaction (Roberts, Little, et al., 2014)	Congruent_A	7.328	1.412	7.169, 7.487
	Incongruent_A	7.260	1.708	6.928, 7.591
	Incongruent_B	7.236	1.470	7.073, 7.398
	Congruent_B	7.233	1.517	7.035, 7.430
Relationship satisfaction (PRQC)	Congruent_A	6.171	0.786	6.081, 6.260
	Incongruent_A	6.154	0.866	5.968, 6.339
	Incongruent_B	6.175	0.827	6.084, 6.266
	Congruent_B	5.972	0.889	5.861, 6.082
Jealousy (Cobey et al., 2013)	Congruent_A	64.428	18.277	62.512, 66.344
	Incongruent_A	59.105	18.091	55.101, 63.110
	Incongruent_B	62.602	17.257	60.647, 64.556
	Congruent_B	58.083	17.696	55.713, 60.453

Note. Congruent_A = congruent users, used hormonal contraception (HC) both when first meeting their partner and at present; Incongruent_A = incongruent users, used HC when first meeting their partner but not at present; Incongruent_B = incongruent users, did not use HC when first meeting their partner but used at present; Congruent_B = congruent users; did not use HC at either time point. *M* = mean; *SD* = standard deviation. FSFI = Female Sexual Function Inventory (Rosen et al., 2000); PRQC = Perceived Relationship Quality Components (Fletcher et al., 2000a).

income, relationship duration, age, and having children together with the partner (the latter was inserted as a factor in the model because the variable was dichotomous). In Roberts, Little, et al. (2014) study, the effect size of the association between HC congruency and sexual satisfaction dropped slightly after controlling for these potential confounders ($\Delta\eta^2_{partial} = -0.006$). Hence, we did not expect that controlling for these potential confounders would improve our chances to detect a significant association between HC congruency and sexual satisfaction, and this was indeed the case ($p = .571$).

Cobey et al. (2013) controlled for age, relationship duration, and general relationship satisfaction when testing the robustness of the association between HC congruency and jealousy. In the present study, the association between HC congruency and jealousy remained nonsignificant after controlling for the above mentioned covariates, regardless of which measure of general relationship satisfaction was used (both *ps* > .898).

As suggested by Cobey, Pollet, Roberts, and Buunk (2011) and Welling, Puts, Roberts, Little, and Burriss (2012), combined contraceptive pills that contain estrogen may have a stronger effect on behavioral phenotypes compared to progesterone-only contraceptives. In order to test this hypothesis, we reran the analyses including only women who had used combination pills. However, no significant group differences between congruent and incongruent users were detected (*ps* ranging from .052–.946). The lowest *p* value (.052) was detected for general relationship satisfaction as measured with the PRQC, and in this case, incongruent users had slightly higher mean satisfaction, contradicting the HC congruency hypothesis.

3.6. Data availability

The data associated with this research are available at: https://data.mendeley.com/datasets/8frmgnz9hx/draft?a=8bfdc2de-0cf5-45a9-affe-e623e389c109.

4. Discussion

Despite excellent statistical power (≥98.7%), and close procedural replications of two out of three previous studies (Cobey et al., 2013; Roberts, Little, et al., 2014), we did not find support for the HC congruency hypothesis. Results from a recent, large-scale investigation into the reproducibility of psychological studies showed that only 36–47% of the replication studies had significant results, with mean effect sizes of the replications of around half of that in the original studies (Open Science Collaboration, 2015). We were unable to perform a direct replication of the study conducted by Russell et al. (2014) as we did not have exactly the same psychometric data. Russell et al. (2014) had third-party measures of perceived facial attractiveness of the partner, whereas in

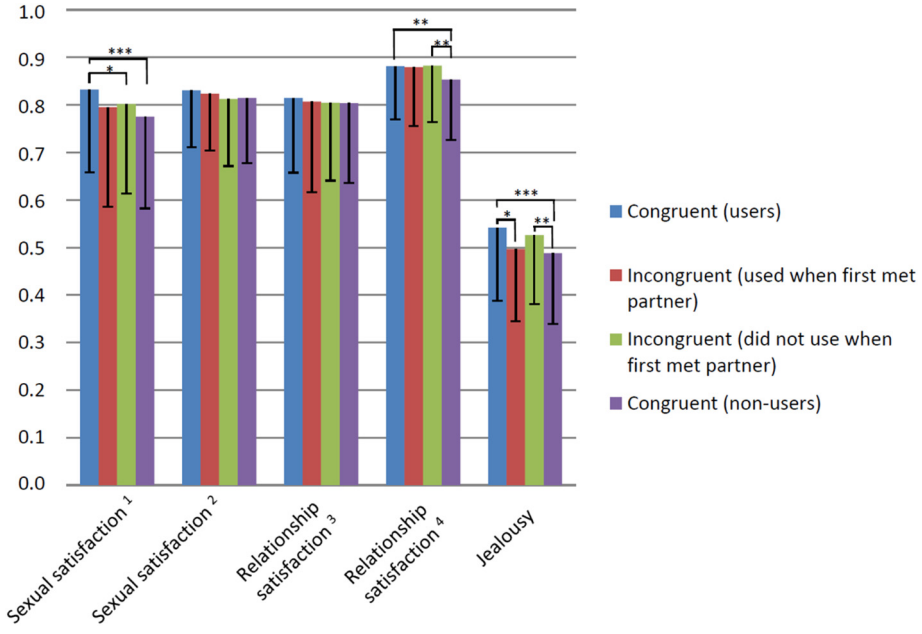

Fig. 2. Means and standard deviations (error bars) for the four different groups of HC congruency. All scales have been converted to a scale from 0 to 1 for clarity. [1] Roberts, Little, et al. (2014) measure of sexual satisfaction; [2] the sexual satisfaction domain of the Female Sexual Function Inventory (Rosen et al., 2000); [3] Roberts, Little, et al. (2014) measure of general relationship satisfaction; [4] Perceived Relationship Quality Components (Fletcher et al., 2000a). * = $p < .05$. ** = $p < .01$. *** = $p < .001$.

the present study, female participants were asked to give an overall appraisal of their partners' physical attractiveness. In addition, we did not have longitudinal observations as in Russell et al. (2014), nor did we have information regarding perceived relationship satisfaction of the women's partners. We were, however, able to test whether the women's perceived attractiveness of their partners moderated the association between HC congruency and general relationship satisfaction, but found no such moderating effect. We did, though, find a significant main effect of HC use on jealousy, such that current HC users were more jealous than their abstaining counterparts (though it should be noted that this has no relevance for the HC congruency hypothesis.)

The differences in sample sizes between the present study and those we aimed to replicate became pronounced when participants were divided into four HC congruency groups. For example, the smallest group in Cobey et al.'s study (2013) (incongruent users who had begun using HCs after meeting their spouse) consisted of only four participants. Thus, in Cobey et al. (2013) the majority (87%) of the incongruent HC users were in fact currently using HCs, whereas the majority (again 87%) of the congruent HC users were not using HCs at the time of study participation. Cobey et al.'s (2013) results might therefore be explained by a main effect of HCs on jealousy – as also detected in the present study – rather than an effect of HC (in)congruency. In Roberts, Little, et al. (2014) study, the group of congruent HC users contained a relatively equal distribution of current users and non-users (>100 individuals in each subgroup). However, the majority ($n = 115$) of incongruent users were current non-users of HCs (77% of incongruent users), with only 35 individuals (23% of incongruent users) having initiated HCs following relationship formation. That is, the congruent group's responses might have reflected a main effect of HCs to a greater extent than in the incongruent group. Russell et al. (2014) collected longitudinal data (a baseline measure with up to 7 follow-ups over 4 years, most reporting data from 3 follow-ups over ~1 year) from two relatively small samples of newlywed wives. The authors focused primarily on effects of HC discontinuation (i.e., incongruent users who quitted using HCs), and thus did not report exact frequencies for all possible combinations of congruency/incongruency. However, the number of wives who had not quit using HCs at any measurement point was relatively low (six individuals in the first sample, 20 in the other). In summary, all of the studies we attempted to replicate had unequal distributions of current users and non-users in at least one of the congruent and incongruent groups, often so that at least one of these groups was almost entirely based on either current users or non-users of HCs. Thus, bearing in mind the small expected effect size of the HC congruency effect, a main effect of HCs could give rise to a spurious effect of incongruency if one of the congruent or incongruent groups has an unequal distribution of current users and non-users.

Our findings of main effects of HC use on relationship satisfaction, sexual satisfaction, and jealousy could reflect a causal effect of HCs on these traits. Research with robust methodology has shown that hormone levels do affect general sexual desire (Jones et al., 2018a), which could in turn influence sexual satisfaction, relationship satisfaction, and jealousy. However, the HC associations in our data could alternatively be due to unmeasured confounding variables, such as personality or sociosexual orientation differences between women who do and do not use HCs (e.g., women with a restricted sociosexual orientation are less likely to enjoy casual sex, and may thus be less likely to use HCs).

There were some discrepancies in terms of sample characteristics between the present study and those we set out to replicate. For example, the average relationship duration was around 30 months in the present study (it should, however, be noted that relationship duration was controlled for in the analyses in the present study, as well as that of Cobey et al., 2013, Roberts, Little, et al. (2014). In Roberts, Little, et al. (2014) the median relationship duration was considerably longer, at 104 months (data not reported; personal communication). Roberts et al. (2012) found that women who were using HCs when they met their partner were less likely to reject sex or undergo compliant sex in shorter relationships, but more likely to do so in longer relationships. Thus, it is conceivable that HC congruency effects on sexual satisfaction may take time before they manifest. However, the average relationship duration in the present study was still nearly three and a half years, which arguably is enough time for a relationship to settle: for example, a recent survey found that more than half of couples decide to get married within the first two years of dating (Francis-Tan & Mialon, 2015). Furthermore, HC congruency effects were detected in the studies of Cobey et al. (2013) and Russell et al. (2014) even though the average relationship durations were comparable to the present study, or shorter.

With regard to the analyses concerning jealousy, a general limitation of both the current study and that of Cobey et al. (2013) was that participants were asked to evaluate their feelings towards a hypothesized situation, although forecasts of emotional reactions to negative events have been shown to be inaccurate (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998). While this is perhaps a valid point in terms of ecological validity, it cannot be viewed as a limitation in the context of the replication effort: the current study used the same questionnaire as employed by Cobey et al. (2013), in which participants also responded to questions regarding hypothetical scenarios.

In summary, the present study found no support for the HC congruency hypothesis, despite being the best-powered study conducted to date. Previous studies testing the HC congruency hypothesis have received substantial attention from the academic community (two out of three studies replicated in the present study have been published in leading journals), as well as coverage in leading mainstream media such as *Time Magazine* (Alter, 2014), *Women's Health* (Goldman, 2014) and *The Guardian* (Gray, 2014). Therefore, it is possible that women exposed to these reports may choose not to use HCs out of fear of choosing an unsuitable partner or father for their children. The present study, which casts serious doubt on the HC congruency effect, could therefore have considerable real-world implications.

Funding

This work was supported by the Academy of Finland [grant numbers: 274521 and 284385].

Open Practices

The raw data associated with this research are available at: https://data.mendeley.com/datasets/8frmgnz9hx/draft?a=8bfdc2de-0cf5-45a9-affe-e623e389c109. All variables, including composite variables, have been clearly labeled in the .sav format. Missing data are indicated as "-9" or empty fields. All analyses should be reproducible by following the instructions in the Methods section.

References

Alter, C. (2014). Going off the pill could affect who you're attracted to, study finds. *Time Magazine* Available: http://time.com/3596014/attraction-sex-birth-control/, Accessed date: 15 June 2016.

Buunk, B. P. (1997). Personality, birth order and attachment styles as related to various types of jealousy. *Personality and Individual Differences*, *23*, 997–1006.

Cobey, K. D., Pollet, T. V., Roberts, S. C., & Buunk, A. P. (2011). Hormonal birth control use and relationship jealousy: Evidence for estrogen dosage effects. *Personality and Individual Differences*, *50*, 315–317.

Cobey, K. D., Roberts, S. C., & Buunk, A. P. (2013). Hormonal contraceptive congruency: Implications for relationship jealousy. *Personality and Inivididual Differences*, *55*, 569–573.

Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000a). The measurement of perceived relationship quality components: A confirmatory factor analytic approach. *Personality and Social Psychology Bulletin*, *26*, 340–354.

Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000b). Ideals, perceptions and evaluations in early relationship development. *Journal of Personality and Social Psychology*, *79*, 933–940.

Francis-Tan, A., & Mialon, H. M. (2015). "A diamond is forever" and other fairy tales: The relationship between wedding expenses and marriage duration. *Economic Inquiry*, *53*, 1919–1930.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*, 617–638.

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*, 1205–1259.

Goldman, A. (2014). The surprising way the pill can impact your sexual satisfaction. *Women's Health Magazine* Available: http://www.womenshealthmag.com/sex-and-love/birth-control-pill-side-effects (Accessed on June 15, 2016).

Gray, R. (2014). The pill affects women's satisfaction with their relationships, research finds. *The Guardian* Available: http://www.theguardian.com/society/2014/nov/17/pill-affects-womens-satisfaction-relationship-contraception (Accessed on June 15, 2016).

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, *7*, 645–654.

Jones, B. C., Hahn, A. C., Fisher, C., Wang, H., Kandrik, M., & DeBruine, L. M. (2018a). General sexual desire, but not desire for uncommitted sexual relationships, tracks changes in women's hormonal status. *Psychoneuroendocrinology*, *88*, 153–157.

Jones, B. C., Hahn, A. C., Fisher, C., Wang, H., Kandrik, M., Han, C., ... DeBruine, L. M. (2018b). No compelling evidence that preferences for facial masculinity track changes in women's hormonal status. *Psychological Science*. https://doi.org/10.1101/136549 (in press).

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349* (aac4716).

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments explained. *Perspectives on Psychological Science*, *7*, 531–536.

Roberts, S. C., Cobey, K. D., Klapilová, K., & Havlíček, J. (2014). Oral contraception and romantic relationships – From the lab to the real world. *Human Ethology Bulletin*, *29*, 4–13.

Roberts, S. C., Klapilová, K., Little, A. C., Burriss, R. P., Jones, B. C., DeBruine, L. M., et al. (2012). Relationship satisfaction and outcome in women who meet their partner while using oral contraception. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*, 1430–1436.

Roberts, S. C., Little, A. C., Burriss, R. P., Cobey, K. D., Klapilová, K., Havlíček, J., et al. (2014). Partner choice, relationship satisfaction, and oral contraception: The congruency hypothesis. *Psychological Science*, *25*, 1497–1503.

Rosen, R., Brown, C., Heiman, J., Leiblum, S., Meston, C., Shabsigh, R., et al. (2000). The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *Journal of Sex & Marital Therapy*, *26*, 191–208.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York, NY: Sage.

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. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 17081–17086.

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?*. Mahwah, NJ: Lawrence Erlbaum Associates.

Welling, L. L., Puts, D. A., Roberts, S. C., Little, A. C., & Burriss, R. P. (2012). Hormonal contraceptive use and mate retention behavior in women and their male partners. *Hormones and Behavior*, *61*, 114–120.