Hormonal Contraceptive Use and Women's Sexuality and Well-Being: Estimating Treatment Effects and Their Heterogeneity Based on Longitudinal Data

Programmatic Registered Report Stage 1

Submitted to Peer Community In Registered Reports (PCI RR) on November 30th, 2022

First Revised Version Submitted to PCI RR on July 13th, 2023

Second Revised Version Submitted to PCI RR on September 20th, 2023

Laura J. Botzet[a,b,*], Julia M. Rohrer[c], Lars Penke[a,b], and Ruben C. Arslan[c,d]

[a]Biological Personality Psychology, University of Goettingen, Goettingen, Germany

[b]Leibniz ScienceCampus Primate Cognition, Goettingen, Germany

[c]Department of Psychology, University of Leipzig, Leipzig, Germany

[d]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

*Corresponding author:

Laura J. Botzet

Biological Personality Psychology, Institute of Psychology

Goßlerstraße 14, 37073 Goettingen, Germany

Email: botzet@uni-goettingen.de

Abstract

Different women experience hormonal contraceptives differently, reporting side effects on their sexuality and well-being that range from negative to positive. But research on such causal effects of hormonal contraceptives on psychological outcomes struggles both to identify average causal effects and capture the high heterogeneity in women's treatment responses. In this study, we plan to leverage longitudinal data to improve our ability to separate the causal effects of hormonal contraceptives from other sources of association, including observed and unobserved confounding, reverse causality, and attrition. We will analyze data from up to 6,565 women who participated in PAIRFAM, a German longitudinal panel dataset consisting of 13 waves using Bayesian multilevel regressions. To deal with confounding and probe the robustness of findings, we will implement two analysis approaches: adjusted regression analyses and inverse probability of treatment weighting analyses. Furthermore, to move beyond average treatment effects, we will analyze heterogeneity in treatment responses and test whether interindividual differences can predict such heterogeneity. Lastly, we will investigate whether treatment response predicts women's decisions about which contraceptive method to use in the long run. Our results will help to understand the impact of hormonal contraception on sexuality and well-being in a naturalistic setting in which women adapt their contraception to their own experiences.

Keywords: causal inference, contraception, hormones, longitudinal analyses, sexuality, well-being

This is a programmatic registered report stage 1. Two stage 2 articles investigating different outcome groups will result from this single stage 1 registered report: one stage 2 article focussing on sexuality (desired sexual frequency, reported sexual frequency, and sexual satisfaction), as indicated with a dark-gray background, and one stage 2 article focussing on well-being (depressiveness, life satisfaction, and self-esteem), as indicated with a light-gray background. All other parts of the stage 1 registered report apply for both stage 2 articles.

This manuscript contains supporting information including rmd files and html files for the blind code, the simulation code, and planned analyses online at https://osf.io/u8ntf/?view_only=6d5b0a56a41541249cab38c51847157c.

Estimating Average Treatment Effects and Treatment Heterogeneity of Hormonal Contraceptive Use on Women's Sexuality and Well-Being Based on Longitudinal Analyses

Programmatic Registered Report Stage 1

The impact of hormonal contraceptives on women's sexuality and well-being has been discussed since their approval in 1960. Before their invention, only so-called barrier methods existed, which prevent fertilization by blocking the union of egg and sperm (e.g., condoms, diaphragms, cervical caps, and chemical spermicides). In contrast, hormonal contraceptives (including oral hormonal contraceptives, but also hormonal implants, hormonal shots, skin patches, and vaginal rings) include synthetic hormones (progestins and sometimes synthetic estrogens) that enter the bloodstream and, in most cases, prevent ovulation (Watkins, 2012).

By altering the endocrine system, hormonal contraceptives can have effects on other aspects of the female body and brain—including negative medical and psychological side effects. For instance, two randomized controlled trials reported small negative effects of oral hormonal contraceptives on sexual desire, arousal, and pleasure (Zethraeus et al., 2016) as well as sexual interest (Lundin et al., 2018). But a recent review by Both et al. (2019) found that only a minority of women reported changes in sexual functioning and concluded that the effects of hormonal contraceptives on sexual functioning – and sexual desire in particular – are understudied and therefore poorly understood.

Experiments are considered the gold standard to answer causal research questions, such as the effects of hormonal contraceptives on sexuality and well-being. However, experimental evidence can only partly tell us how these effects affect women's everyday lives. As Graham (2019) points out, women's experiences with hormonal contraceptives are highly heterogeneous – ranging from negative side effects to no effects to positive effects. These heterogeneous responses to hormonal contraceptive use might be due to varying sensitivity to hormones (Kiesner, 2017). Such differences in sensitivity are also supported by evidence that ovulatory cycle shifts with average increases in sexual desire and

self-perceived attractiveness during the fertile phase vary between women (Arslan et al., 2021; Schleifenbaum et al., 2021). Hormonal contraceptives inhibit ovulation, and so hormonal contraceptive users no longer experience the same ovulatory cycle shifts. Heterogeneous effects of hormonal contraceptives might therefore be due to varying sensitivity to ovulatory cycle shifts before starting hormonal contraceptive use, with sensitive women showing stronger effects and insensitive women showing smaller effects on sexuality and well-being.

Such differences in the effects of hormonal contraceptives can be studied in an experimental context, as suggested by Hill and Mengelkoch (2022) who propose a precision medicine approach. They suggest researchers collect detailed information about contraceptive methods, duration of contraceptive use, mental health history, as well as sexual activity and relationship status as important potential moderators of the relationships between hormonal contraceptive use and psychological outcomes (see Box 3 and 4 in Hill & Mengelkoch, 2022).

Carefully isolated experimental settings are valuable to establish the effects of (individual) hormonal contraceptives on women's sexuality and well-being. In contrast, in everyday life, women actively choose between different non-hormonal and hormonal contraceptive methods and often try multiple methods during their lifespan. As women try to find a balance between efficacy, ease-of-use, as well as desirable and undesirable side effects, the causal effects of synthetic hormones are interwoven with confounding, attrition effects, and reverse causality. This poses unique causal inference challenges, but also allows one to investigate additional research questions such as whether side effects determine which contraceptive women eventually choose. Furthermore, the different requirements of observational data collection (as opposed to randomized clinical trials) make it easier to include a broad range of variables such as personality, thus making it possible to more thoroughly investigate potential predictors of women’s heterogeneous responses to hormonal contraceptives.

The current study aims to close the gap between the available experimental and correlational evidence about the relationship between hormonal contraceptives and women's sexuality and well-being. By analyzing the effects of starting and discontinuing hormonal contraceptives on sexuality and well-being in a longitudinal dataset with around 6,500 women, observed over up to 13 yearly waves (years of data collection: 2008–2021), we hope to answer questions about potentially heterogeneous average treatment effects of hormonal contraceptives in real world settings while accounting for (un)observed confounders as well as attrition effects.

This manuscript is a programmatic registered report stage 1. Two stage 2 articles investigating different outcome groups will result from this single stage 1 registered report: one stage 2 article focussing on sexuality (desired sexual frequency, reported sexual frequency, and sexual satisfaction), as indicated with a dark-gray background, and one stage 2 article focussing on well-being (depressiveness, life satisfaction, and self-esteem), as indicated with a light-gray background. All other parts of the stage 1 registered report apply for both stage 2 articles.

Empirical Evidence of Positive and Negative Effects of Hormonal Contraceptives

Hormonal contraceptives contain synthetic versions of progesterone (also called progestin) and sometimes estrogen, which inhibit the natural production of progesterone and estrogens as well as the natural production of pituitary hormones (luteinizing hormone and follicle-stimulating hormone). This reduction of natural hormonal fluctuation across the menstrual cycle prevents the maturation of the ovarian follicle and therefore hinders ovulation (Frye, 2006). In general, women who are using hormonal contraceptives have lower levels of estradiol, progesterone, follicle stimulating hormones, luteinizing hormones, and total and free testosterone as well as higher levels of sex-binding globulins (Gaspard et al., 1983; Zethraeus et al., 2017; Zimmerman et al., 2014). Their endogenous hormone levels remain constantly similar to those found in the early follicular phase of normally cycling women (Mishell et al., 1972).

Sexuality

This intervention into the endocrine system (Stomati et al., 1998) has been hypothesized to negatively affect women's sexuality (Both et al., 2019). Some empirical evidence supports these hypotheses regarding sexual functioning (e.g., Læssøe et al., 2014) and libido (e.g., Lee et al., 2017; Lundin et al., 2018; Zethraeus et al., 2016), as well as sexual activity, arousal, pleasure, orgasm, and lubrication (Smith et al., 2014). While hypotheses and evidence for negative side effects of hormonal contraception exist, the use of hormonal contraception has also been hypothesized to positively affect women's sexuality through several mechanisms, including, for example, overcoming the fear of unwanted pregnancy during sexual activity (Blumenstock & Barber, 2022) and the resolution of painful or troublesome gynecologic disorders (Both et al., 2019). Empirical evidence in support of positive effects on sexuality has been reported concerning sexual functioning (e.g., Oranratanaphan & Taneepanichskul, 2006), libido (McCoy & Matyas, 1996), and, most strongly, sexual frequency (e.g., Caruso et al., 2005; McCoy & Matyas, 1996). In addition, women using hormonal contraceptives reported higher sexual satisfaction (e.g., Caruso et al., 2005) and higher relationship satisfaction (e.g., Taggart et al., 2018)

Well-Being

As for sexuality, the evidence for potential effects of hormonal contraceptives on well-being is mixed. Böttcher et al. (2012) summarized evidence in their meta-analyses showing positive, negative, and null effects of hormonal contraceptives on depressiveness. While there is some empirical evidence for positive effects of hormonal contraceptives on general well-being (e.g., Apter et al., 2003; Caruso et al., 2005) and reduced depressive symptoms (e.g., Toffol et al., 2011, 2012), Zethraeus et al. (2017) reported a negative effect of hormonal contraceptives on general well-being in a double-blind randomized controlled trial with a placebo control group. This effect might be dependent on the specific ovulatory cycle phase, as Lundin et al. (2017) reported small positive effects of hormonal contraceptives on anxiety, irritability, and mood swings during the intermenstrual phase, but

a negative effect on depression during the premenstrual phase based on a double-blind randomized controlled trial. Skovlund et al. (2016) showed that the relative risk of a depression diagnosis was 1.4 times higher six months after starting hormonal contraceptive methods (compared to no or non-hormonal contraceptive methods) in a sample of over one million Danish women (but see Lundin et al. (2022) for somewhat contradictory evidence based on a similar approach applied to a Swedish sample). In addition, the risks for suicide attempts and suicide were increased after starting hormonal contraception (Skovlund et al., 2018).

Obstacles to Estimating Psychological Effects of Hormonal Contraceptives

Taken together, evidence concerning potential psychological effects of hormonal contraceptives remains inconclusive. While randomized-controlled trials provide somewhat consistent evidence of small negative average treatment effects on various aspects of women's sexuality (e.g., Graham et al., 1995; Lundin et al., 2018; Sabatini & Cagiano, 2006; Zethraeus et al., 2016; but see Oranratanaphan & Taneepanichskul (2006) and Strufaldi et al. (2010) for evidence of positive causal effects of certain methods of hormonal contraception), evidence based on correlational data often shows no or even positive relationships between the use of hormonal contraceptives and sexuality (e.g., Caruso et al., 2005; McCoy & Matyas, 1996; but see Wallwiener et al. (2010, 2015) for evidence of a negative relationship). Some reviews about potential effects of hormonal contraceptives conclude that there are negative effects of hormonal contraceptives (Lee et al., 2017) or no effects of hormonal contraceptives (Pastor et al., 2013). However, most reviews conclude that the effects of hormonal contraceptives on sexuality have not been well studied and remain controversial (Both et al., 2019; Burrows et al., 2012; Davis & Castaño, 2004; Schaffir, 2006).

Several explanations for this mixed and inconclusive body of evidence are plausible:

(1) *Contraceptive method and dosage effects*: differing psychological responses are due to differences between hormonal contraceptives (e.g., application

methods or different dosages of synthetic progesterone and estrogen; for supporting empirical evidence see e.g., Boozalis et al., 2016; Læssøe et al., 2014; Sabatini & Cagiano, 2006; Strufaldi et al., 2010)

(2) *Treatment heterogeneity*: differing psychological responses are due to interindividual differences between women (Graham, 2019) and studies systematically vary in sampling procedures (e.g., some only including women with a regular ovulatory cycle)

(3) *Treatment heterogeneity leading to selective attrition:* women who experience negative effects of hormonal contraceptives discontinue them, leaving only women who experience no effects or positive effects in the group of hormonal contraceptive users in correlational studies

(4) *Confounders*: pre-existing differences in women influence the decision what contraceptive method to use *and* affect psychological outcomes, leading to differences between the groups of hormonal contraceptive users and non-hormonal contraceptive users in correlational studies

(5) *Reverse causality*: in some cross-sectional studies, relationships between psychological outcomes and hormonal contraceptive use might occur because the outcome influences the contraceptive choice (e.g., higher frequency of vaginal intercourse might lead to the decision to start using hormonal contraceptives).

Randomized controlled trials with a placebo control group are regarded as the superior approach for estimating the average treatment effect of hormonal contraceptives and their contraceptive efficacy. They can also expand the knowledge about (1) *contraceptive method and dosage effects* and (2) *treatment heterogeneity*. While the estimated effects will not be biased through (4) *confounders* and (5) *reverse causality* as their impacts are nullified by randomization, this also means that the design cannot inform us about the extent to which these two affect correlations between contraceptive usage and outcomes in everyday life. Furthermore, this design is not optimized to inform us about how

(3) *treatment heterogeneity might lead to selective attrition* in everyday life. A related concern is sometimes termed *healthy user bias*: the women who volunteer for a randomized controlled trial will not include, for example, women who, based on previous experience, fear bouts of severe depression if they are assigned to hormonal contraception. By randomly assigning different forms of contraceptives to women, they remove the decision process to start or to discontinue using contraceptives that is inherent to real world settings. In addition, owing to their cost, randomized controlled trials usually have small sample sizes that preclude the rigorous investigation of subgroups, heterogeneity, and uncommon side effects. Finally, trials with a non-hormonal contraceptive control group are uncommon, in part because pharmaceutical trials tend to focus on comparing different formulations and in part because many non-hormonal methods are less efficacious, increasing the risk of unplanned pregnancies. For example, in the randomized trial with a non-hormonal contraceptive control group by Zethraeus et al. (2016, 2017) women were blinded and did not know whether they were using hormonal contraceptives. To avoid unwanted pregnancies, all women were instructed to use additional non-hormonal contraceptive methods during the study and received free condoms (Zethraeus et al., 2017). Therefore, any beneficial effects resulting from knowing that one is using a highly effective birth control method (Both et al., 2019) may be underestimated in such blinded randomized controlled trials.

Observational Cross-Sectional and Longitudinal Designs

In comparison to randomized-controlled trials, observational cross-sectional designs also capture any association induced by the decision process. Therefore, (3) *selective attrition*, (4) *confounders*, and (5) *reverse causality* will often bias the estimated effects. At the same time, they are usually based on larger sample sizes and include users of multiple contraceptive methods as well as those who use no contraceptive method at all. They operate like photographs of the real world. While they only show patterns at one specific time point, they still provide important pieces of the picture (such as the associations between demographic variables and contraceptive method) that could not be obtained based

on randomized controlled trials alone. Going beyond mere associations, we can at least attempt to infer causal effects from cross-sectional data, if we are willing to transparently discuss and defend the necessary strong assumptions and statistical adjustments (e.g., Botzet et al., 2021).

One way to reduce the number of assumptions necessary for causal identification in observational data is examining change over time within individuals, because many of the potential confounding factors that vary between individuals are held constant by design. Longitudinal designs can rule out between-subject confounders by allowing the use of within-subject analyses (Rohrer & Murayama, 2021). Therefore, time-invariant confounders can be ruled out when estimating causal effects based on appropriately specified longitudinal designs.

Such panel studies operate like a series of photos:[1] We can track change, but still have to be cautious not to confuse cause and effect, since multiple events can occur in the interim—a longitudinal design alone is no guarantee of appropriate causal inference. Still, given transparent assumptions and adequate statistical control, we can at least attempt to infer causal effects. Specific statistical models are needed to remove confounders (Hamaker et al., 2015) and all modeling decisions ultimately reflect assumptions about the underlying causal network (Rohrer & Lucas, 2020).

Given the correct modeling decisions, time-invariant confounders are automatically controlled for in longitudinal designs. As they do not vary within a woman, they will not induce spurious correlations between her time-varying predictor and her time-varying outcome. Time-varying confounders on the other hand are not automatically controlled by longitudinal designs, but instead need to be accounted for (Rohrer & Murayama, 2021). A time-varying confounder might affect a woman's choice of contraceptive method as well as the outcome of interest at a given time. For example, an ineffable or at least unmeasured shift from a casual to a more steady exclusive relationship may affect the decision to use

[1] Going a step further, by analogy to movies, we could do even better by having more granular, potentially daily longitudinal data on contraception, which would, for example, allow us to explicitly model the effects of the menstrual cycle.

hormonal contraceptives. In addition, this shift could cause more frequent sexual activity at a later time. In a longitudinal design that only measures hormonal contraceptive use and sexual activity but not this relationship shift, it will appear like there is a positive causal effect of hormonal contraceptives on sexual activity.

Some of these time-varying confounders might not have been observed in the available dataset or might even be completely unobservable – they thus cannot be accounted for in the statistical analysis. Such unobserved confounders bias the estimate no matter what analytic strategy is used, which we analyzed in our simulations reported below. However, additional sensitivity analyses can be conducted to estimate the influence unobserved confounders would need to have to fully account for the remaining observed relationship between treatment and outcome, thus providing at least the opportunity to make an educated guess about the internal validity of the results (for early work on sensitivity analysis for unobserved confounders see Rosenbaum & Rubin, 1983).

Longitudinal designs investigating potential medical effects of hormonal contraception are relatively common (e.g., Eng et al., 2008; Riggs et al., 2007; Wang et al., 2016), although all of these studies implement randomized treatment assignment rather than an observational approach. To our knowledge, only two studies investigated effects of hormonal contraception on sexuality with an observational longitudinal design. Blumenstock and Barber (2022) analyzed data from a weekly survey over 2.5 years from 893 women. They showed that women had a higher sexual frequency when they were using hormonal contraceptives. Frequency of sexual intercourse increased after starting using hormonal contraception, remained high for several months, and then slowly declined. Ott et al. (2008) showed in a 41-month long study with 328 participants that sexual interest based on daily diaries did not change when women started using oral contraceptives. But when women stopped using oral contraceptives, sexual interest decreased. Concerning psychological effects on well-being, Skovlund et al. (2016) used a three-year nationwide prospective cohort study and showed that use of hormonal contraceptives predicted an increased risk of first depression diagnosis under control of time-varying covariates. In a follow-up study, the start

of hormonal contraception predicted increased risks in suicide attempts and suicide and this association was particularly strong after two months of use (Skovlund et al., 2018). Partly contrary to the findings by Skovlund et al. (2016), a recent study by Lundin et al. (2022) used a similar approach to investigate effects of hormonal contraception on depression diagnosis based on a Swedish nationwide prospective cohort study covering seven years and found increased risk of diagnosis of depression after starting all forms of hormonal contraception except for oral hormonal contraceptives.

To summarize, causal inference from longitudinal data is only possible on the basis of assumptions. We strive to make our analysis goal (Lundberg et al., 2021) and the assumptions underlying our causal identification strategy as transparent as possible. In addition, we apply two different analytical approaches with different underlying assumptions.

Heterogeneity in Treatment Responses

While evidence for a negative average treatment effect on sexuality based on randomized controlled trials exist (Zethraeus et al., 2016), self-reports by women indicate that individual treatment effects on sexuality might vary widely (Malmborg et al., 2016). Heterogeneity in treatment responses might be caused by individual differences in responses to steroids (Kiesner, 2017). To our knowledge, treatment heterogeneity of hormonal contraceptives on sexuality or well-being has not been estimated quantitatively. Based on longitudinal data analyses, individual treatment effects on sexuality and well-being for each woman can be estimated and the distribution of individual treatment effects and their uncertainty can be visualized.

Estimating individual treatment effects will allow us to answer further questions about the underlying causal network connecting hormonal contraceptives and sexuality as well as well-being. Is there a large number of women who experience either positive or negative effects? Do women use their own experience with individual effects of hormonal contraceptives on sexuality and well-being to make a decision about their contraceptive method? For example, are women who experience adverse effects of hormonal

contraceptives on sexuality or well-being more likely to stop using them during a specific time span? In addition, we want to answer the question whether interindividual differences like demography and personality predict individual treatment effects. Older women might be more likely to experience beneficial side effects of hormonal contraceptives on sexuality and well-being because they found the method that fits them best. In line with this reasoning, empirical findings suggest that higher age was associated with less negative side effects of hormonal contraceptive use on depression with particularly strong negative effects during adolescents (Skovlund et al., 2016). Nevertheless, these findings might be accountable by other explanations, e.g. a possible decrease in sensitivity to steroid hormones with age or a specifically strong sensitivity to steroid hormones during puberty. Women with higher scores on openness might be more likely to experience beneficial side effects as well because they are more likely to try out different contraceptive methods until they find their perfect method. Other personality dimensions might be related to negative or positive individual treatment effects. For example, women with higher scores on neuroticism may experience more positive psychological effects as their heightened worries about unwanted pregnancies are reduced.

Focusing on individual treatment effects of hormonal contraceptives on sexuality and well-being effects will allow us to broaden our understanding about the individual nature of potential effects of hormonal contraceptives as well as confounding and attrition effects.

The Current Study

In the current study we want to answer the questions whether hormonal contraceptive use influences women's sexuality and well-being (over and above attrition effects, accounting for observed and unobserved confounders) as well as whether and to which extent the effects of hormonal contraceptives on sexuality and well-being vary between users. Sexuality outcomes will include desired sexual frequency in the last three months as a measure for libido, reported sexual frequency in the last three months, and sexual satisfaction. Well-being outcomes will include depressiveness, general life

satisfaction, and self-esteem. By using a longitudinal design we can partly rule out alternative explanations such as reverse causality. Analyses will be based on the German Family Panel (PAIRFAM), a panel dataset containing information about contraceptive use as well as women's sexuality and well-being from more than 6,500 women over 13 waves, starting in 2008 (Brüderl et al., 2021; Huinink et al., 2011).

Conceptual Design and Underlying Assumptions

The conceptual design of the study, including all underlying assumptions, is outlined in Figure 1. These two graphs correspond to the two analytical approaches that we will use to estimate the causal effect of hormonal contraceptives on the four outcomes.

The graph in panel A shows the adjusted regression approach, which estimates the effect of contraceptive method on the outcome while controlling for the respective outcome, contraceptive method, and their interaction at the previous wave, as well as potential observed time-varying confounders (i.e., demography, relationship information). In addition, the potential influence of unobserved (and unobservable) confounders will be estimated.

The graph in panel B corresponds to the conceptual design underlying the inverse probability of treatment weighting approach (IPTW; Thoemmes & Ong, 2016). For this approach, individuals are weighted by their probability to receive a specific treatment, in our case hormonal contraceptive use. This weight for each individual is modeled with effects of the outcome, contraceptive method, and their interaction at the previous wave, as well as potential observed time-varying confounders (i.e., demography, relationship information) on the treatment itself (i.e., hormonal contraceptive use). When estimating the effect of hormonal contraceptives on the respective outcome this weight will be taken into account.

Why implement two approaches instead of only one line of analyses? According to Thoemmes and Ong (2016), the adjusted regression approach has several disadvantages: (1) regressions with different numbers of covariates can be estimated easily and therefore may introduce biases through cherry-picking (Rubin, 2001); (2) the adjusted regression approach relies on the untested key assumption that the relationships between the

covariates and the outcome are modeled appropriately (more narrowly described as the linearity assumption, see Gutman & Rubin, 2017); (3) any comparisons between the treated and the untreated group might be due to extrapolation because there are no treated participants who are comparable to the untreated participants (King & Zeng, 2006).

While we agree that the IPTW approach outperforms adjusted regression analyses in estimating the causal effect of a treatment on an outcome in many possible scenarios (Fuentes et al., 2021), the first two disadvantages of adjusted regression mentioned above can also apply to the IPTW approach: (1) models estimating the individual weights *are* regression models that can be performed as easily with a different number of covariates and therefore potential bias through cherry-picking is not meaningfully precluded, and (2) the IPTW approach relies on the untested key assumption that the relationships between the covariates and the treatment are modeled appropriately (as opposed to the relationships between covariates and outcome, see assumptions of adjusted regression approach).

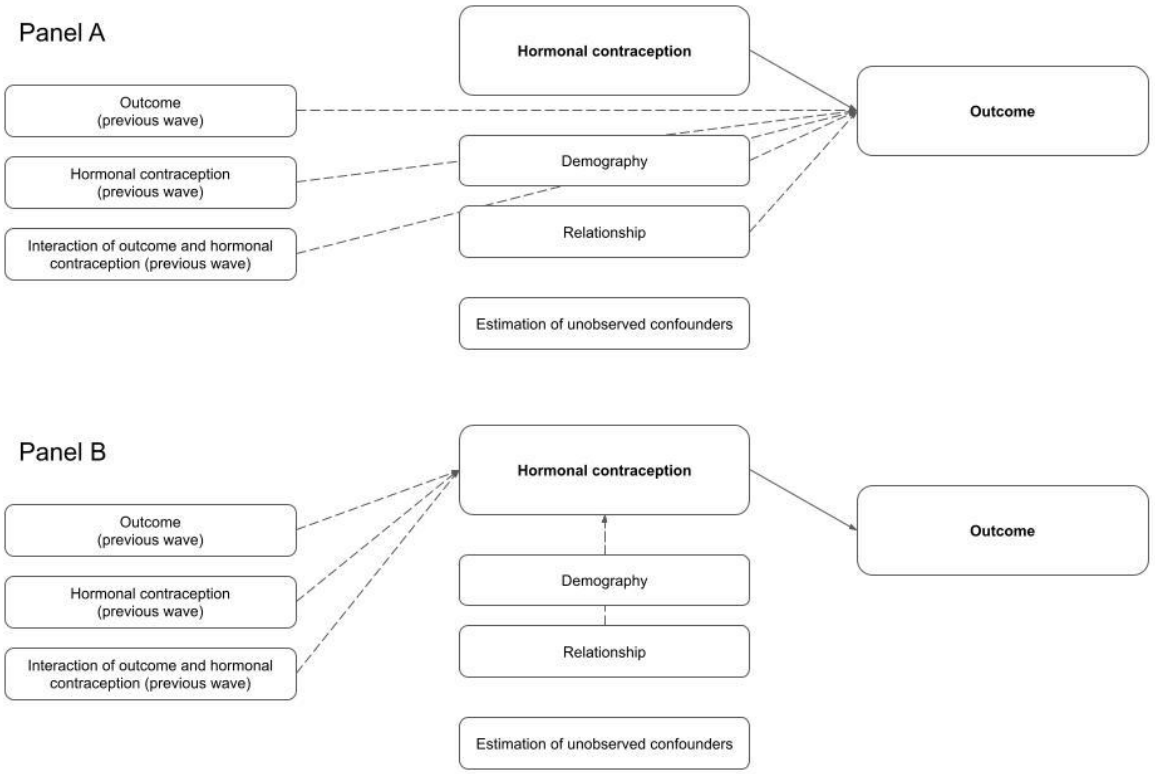
We address the first concern of both approaches (introduction of bias through cherry-picking) by carefully laying out the assumed underlying causal network and preregistering our models in form of a registered report before having access to the data. To address the second concern (nonlinearity between covariates and outcome or treatment, respectively), we decided to perform and compare both approaches to estimate the causal effect of hormonal contraceptives on the outcome robustly under different sets of assumptions. Nevertheless, both approaches still rely on the assumptions of (1) no unobserved confounders; (2) positivity (i.e., every individual having a probability of receiving the treatment that is larger than 0 and smaller than 1); and (3) a correct specification of the underlying models (Thoemmes & Ong, 2016). To estimate the dependency of our analyses on these three underlying assumptions, we tested the proposed models with different specifications based on simulated data with varying data generating mechanisms. The models, simulations, and results are described in the methods section. Given our interest in the immediate effects of hormonal contraceptive use (rather than the lagged effects after one year), and to avoid adding superfluous complexity, we decided against a popular alternative

modeling approach (RI-CLPM, Hamaker et al., 2015) which simultaneously attempts to estimate causal effects pointing into the opposite direction.

In addition, we plan to estimate the potential influence of *unobserved* confounders on the average treatment effect. We will run additional sensitivity analysis to estimate how sensitive the results are to hidden bias. Although a sensitivity analysis does not compensate for unobserved confounding, it quantifies how large the hidden bias would need to be to change the conclusions substantially (see methods section).

Figure 1

Conceptual design of the analyses approaches.



Note. Panel A shows the conceptual design and assumptions underlying the adjusted regression model. Panel B shows the conceptual design and assumptions underlying the inverse probability of treatment weighted (IPTW) regression model.

Estimands

In the context of this study, we are not interested in assessing dichotomous hypotheses (i.e., whether an effect of contraception on sexuality or well-being does or does not exist), but rather in estimating the magnitude and heterogeneity of a range of effects of interest. Thus, instead of formulating hypotheses, we want to specify clear analysis goals and theoretical estimands, define estimation strategies, and specify the corresponding empirical estimands (Lundberg et al., 2021). By precisely defining all target quantities, estimands connect theory with statistical evidence. The study design template in Table 1 based on the template provided by *Peer Community In Registered Report* (https://rr.peercommunityin.org/help/guide_for_authors), therefore includes theoretical estimands and empirical estimands instead of hypotheses.

First of all, we are interested in overall descriptive patterns, including the percentage of hormonal contraceptive users across waves and common patterns in use and switches of hormonal contraceptives. Based on the full sample of all eligible women participating in PAIRFAM, we want to examine descriptives and general trends over the course of the study.

Second, we are interested in why women choose hormonal contraceptive methods. To get a better understanding of potential causes, we will investigate whether time-varying covariates predict contraceptive methods. This will be based on the IPTW model as this approach explicitly models how likely women are to use hormonal contraceptive methods. Our empirical estimands will be quantified as percentage points based on marginal effects.

Third, we want to estimate the average treatment effect of hormonal contraceptive use on all four outcomes. Therefore, adjusted as well as IPTW regression models will be performed to estimate the causal effect, taking into account observed confounders. In addition, the sensitivity of the models to unobserved confounders will be estimated. Our empirical estimand will be the unstandardized mean difference in the outcome between non-hormonal and hormonal contraceptive use. For the reported sexual frequency outcome,

this difference can be seen as a very rough approximation of the percentage change in sexual frequency.[2]

In addition, we are interested in treatment heterogeneity. Therefore, we will investigate individual treatment effects on the outcome based on the adjusted regression models (see section Simulation for an explanation why we do not investigate treatment effect heterogeneity in the context of IPTW regression models).To help interpret this quantity, we will visualize the distribution and uncertainty of individual estimates and report for how many women we estimate negative and positive effects.

Furthermore, we want to explore the correlation between individual treatment effects and age as well as the correlations between individual treatment effects and Big Five personality traits. While these analyses will be less focused on causal identification, they might still provide tentative evidence for substantively plausible causal hypotheses.

In addition, we want to investigate whether women's individual treatment effects on sexuality and well-being inform their decision of which contraceptive method to use by investigating the correlation between estimated individual treatment effects and the number of years using hormonal contraceptives during the course of PAIRFAM. Ideally, we would have sufficient data to instead estimate individual treatment effects (e.g., using all but the last wave of data) to predict individual behavior (e.g., contraceptive method in the very last wave of data). However, in the context of the available data, this would result in very low statistical power, and we thus decided on a different approach which would only provide very rough evidence for potential assortment based on experiences with contraceptive methods. Such an assortment based on experiences would result in the type of selective attrition explained above and may provide a partial explanation for the mixed evidence concerning effects of hormonal contraceptives on sexuality and well-being.

---

[2] This is the case because the response scale of this item is very roughly a log-transformed version of frequency, e.g., on the response scale, the difference between 2 = *once per month and less* and 4 = *once per week* is as large as 4 = *once per week* and 6 = *more than three times a week*. For the full response scale see Table 3.

Table 1

Study design

Theoretical estimand	Quantification of empirical estimand	Sampling plan	Analysis plan / Estimation	Interpretation given different outcomes
Descriptive patterns in hormonal contraceptive use	Percentages of hormonal contraceptive users Probability to switch between hormonal and non-hormonal contraceptive use Average number of switches	All available data from PAIRFAM across 13 waves	Descriptive analyses	—
"Confounding" effects on hormonal contraceptive use	Percentage points based on marginal effects	n = 6,537 women with a mean average of 5.57 waves → 1,950 women reported using both hormonal contraceptives and non-hormonal contraceptives at some point while participating in PAIRFAM → approximately 13,000 switches between contraceptive methods	Linear binomial regression with hormonal contraceptive method as a dichotomous outcome and all treatment predictors as predictors (same model is used for the weights of the inverse probability of treatment weighting approach)	—
Average treatment effects of hormonal contraceptive use on sexuality and well-being	Unstandardized mean difference between non-hormonal and hormonal contraceptive use		Adjusted linear regression analyses Inverse probability of treatment weighting approach	If outcomes based on the two estimations differ, adjusted linear regression analyses will be treated as the main analysis and the inverse probability of treatment weighting approach will be treated as a robustness analysis for identifying the average treatment effect
Heterogeneity in treatment effects of hormonal contraceptive use on sexuality and well-being	Percentage of women with negative estimated effects and positive estimated effects		Extracted individual treatment effects from adjusted linear regression analyses	—
Link between individual treatment effects and predictors of individual treatment effects as well as contraceptive decision	Correlations between individual estimated treatment effects and age, personality traits, as well as years spent on hormonal contraceptives		Extracted individual treatment effects correlated with age, personality traits, as well as years spent on hormonal contraceptives weighted by inverse standard error	—

Note. This table is adapted based on the study design template provided by *Peer Community In Registered Report* here:

https://rr.peercommunityin.org/help/guide_for_authors. PAIRFAM = German Family Panel (Brüderl et al., 2021; Huinink et al., 2011).

Methods

Data

All data will come from a German panel study called PAIRFAM (Brüderl et al., 2021; Huinink et al., 2011). It contains information about contraceptive use and women's sexuality as well as well-being from more than 6,500 women. The longitudinal design consists of annual waves with the first data collection in 2008 and the latest available data from 2021 (wave 13). The ethics committee of the Faculty of Management, Economics, and Social Sciences of the University of Cologne approved PAIRFAM. Huinink et al. (2011) provide a detailed description of the PAIRFAM dataset. In addition, the present manuscript contains supporting information including rmd files and html files for the blind code, the simulation code, and planned analyses online at https://osf.io/u8ntf/?view_only=6d5b0a56a41541249cab38c51847157c.

The data on which our analyses are based were already available and can be used for scientific purposes; the Leibniz Institute for the Social Sciences (GESIS) grants access to the scientific community. Only one of the authors has previously accessed the PAIRFAM data; JMR was granted access to Release 7.0 (waves 1-7) in 2016 within the context of a different research project but never actually worked with the data beyond an initial screening of the included variables to determine suitability for her research question (birth order effects on personality). Thus, some of the data used to answer this research question has been previously downloaded by one of the authors, but we certify that we have not observed any part of the data relevant to the present research question (Level 3 based on the categorization in Table 1 by *Peer Community In Registered Report;* https://rr.peercommunityin.org/help/guide_for_authors).

Exclusion Process and Participants

In order to estimate sample sizes for our planned analyses and investigate potential patterns of missingness, we wrote code prior to data access. This blind code can be found in the form of an rmd file and an html file here:

https://osf.io/u8ntf/?view_only=6d5b0a56a41541249cab38c51847157c. The code was executed on the PAIRFAM dataset by Tita Gonzalez Avilés who is part of the PAIRFAM team on intimate relationships. Gonzalez Avilés is an independent researcher and is not otherwise involved in this manuscript.

We plan to exclude individuals who did not identify as female. Furthermore, once a woman crossed the age of 50 or reported to be (post-)menopausal, her data (including subsequent waves) will be excluded, but previous waves of data collection will remain in the analysis. In addition, we plan to exclude all individual waves of data in which participants indicated being in a homosexual relationship or only reported homosexual relationships in the past, were pregnant, trying to become pregnant, gave birth to a child in the last year, were currently breastfeeding, or indicated using the morning-after-pill or an unknown contraceptive method.

In further separate robustness analyses we plan to additionally exclude waves in which participants indicated that they are sterilized, as well as all subsequent waves of those participants.We will also exclude all waves in which participants indicated that their partner is sterilized and all waves in which women indicated using no contraceptive method[3], an intrauterine device as a contraceptive method, or hormonal methods other than the oral contraceptive pill. In addition, we will exclude all waves in which women indicated that they had never been sexually active. In addition to these robustness analyses, which focus on excluding specific women or waves that might bias the estimates of the main analyses, we would like to conduct exploratory subanalyses based *only* on women who reported being in a homosexual relationship or who have reported only homosexual relationships in the past (otherwise using the same exclusion criteria as in the main analyses). While we hope to gain some initial insight into the potential effects of hormonal contraceptives on sexuality and

[3] In PAIRFAM, women who indicated that they had never been sexually active in their life were not asked about their contraceptive method. These women were coded as using no contraceptive method, i.e., a non-hormonal method (see the section about the variables for more information). This coding may introduce some errors as some women may use hormonal methods without being sexually active; we thus exclude them in an additional robustness analysis to ensure that this coding decision does not systematically affect results.

well-being of homosexual women, the sample size of these exploratory analyses is likely to be too small to draw any definitive conclusions (especially regarding homosexual women using hormonal contraceptives). If, after applying the exclusion rules, fewer than 200 homosexual women reported switching between hormonal and non-hormonal contraception at least once, we will not conduct these additional exploratory subanalyses. All exclusion criteria, reasons for exclusion, and excluded unit(s) are summarized in Table 2.

Overall, n = 6,565 women can be included in our main analyses. They participated in 6 waves on average. Of the full sample, 2,087 women reported using both hormonal contraceptives and non-hormonal contraceptives at some point while participating in PAIRFAM. Approximately 3,000 switches between contraceptive methods were observed, with more switches from hormonal to non-hormonal contraceptives than vice versa. Besides estimating potential sample size and number of switches, information from the blind code allowed us to simulate our planned analyses based on realistic assumptions.

Table 2

Exclusion criteria, reasons for exclusion, and excluded units

Main analyses		
Exclusion criteria	**Reasons for exclusion**	**Excluded unit(s)**
identifying as non-female	potential hormonal influences	current and all subsequent waves
older than 50 years	potential hormonal influences	current and all subsequent waves
(post-)menopausal	potential hormonal influences	current and all subsequent waves
only homosexual relationships	no need to use contraceptives to prevent pregnancy	current wave
pregnant	potential hormonal influences	current wave
trying to become pregnant	no need to use contraceptives to prevent pregnancy	current wave
gave birth in the last year	potential hormonal influences	current wave
breastfeeding	potential hormonal influences	current wave
using the morning-after-pill as a contraceptive method	potential hormonal influences	current wave
using an unknown contraceptive method	not possible to classify method as hormonal or non-hormonal	current wave
Further robustness analyses		
Exclusion criteria	**Reasons for exclusion**	**Excluded unit(s)**
sterilized	no need to use contraceptives to prevent pregnancy	current and all subsequent waves
partner sterilized	no need to use contraceptives to prevent pregnancy	current wave
using no contraceptive method	imprecise classification as non-hormonal in main analyses	current wave
using an intrauterine device as a contraceptive method	imprecise classification as non-hormonal in main analyses	current wave
using other hormonal methods	investigate effects of oral contraceptive pills only	current wave
never sexually active	potentially conditioning on the sexual frequency as an outcome	current and all subsequent waves

Variables

All variables, including the predictor variable, potential time-varying confounders, outcome variables, and variables used to investigate treatment heterogeneity are listed in Table 3. The original German item wording can be found here: https://www.pairfam.de/dokumentation/fragebogen/.

The predictor hormonal contraception will be based on the items about the contraceptive method; participants were able to report multiple contraceptive methods. Hormonal contraception will be coded as 0 if participants indicated that they use no contraceptive method at all. The variable hormonal contraception will also be coded as 0 if participants indicated that they use no hormonal contraceptive method and at least one of the following methods: *condom*; *intrauterine device*[4]; *diaphragm, foam, suppository, gel*; *natural birth control*; *female sterilization*; *male sterilization*; or *withdrawal method, coitus interruptus*. In addition, the variable hormonal contraception will be coded as 0 if participants were never sexually active in their life, as these participants were not asked about their contraceptive method. The variable hormonal contraception will be coded as 1 if participants indicated that they use a *birth control pill*, *mini pill*, or *other hormonal method (implant, patch, ring)*, even if they additionally use non-hormonal methods. Exclusion criteria for main as well as robustness analyses based on the contraceptive method are described above.

[4] Participants were not asked whether they used a hormonal or copper intrauterine device. Therefore, we will code the choice *intrauterine device* as hormonal if participants had indicated earlier in the survey that they use *other hormonal method (implant, patch, ring)*, assuming that women who use a hormonal intrauterine device would classify this as another hormonal method after the option *birth control pill, mini pill*. If participants only indicated that they use an *intrauterine device* but no hormonal method, this was coded as non-hormonal contraception.

Table 3

Item wordings and answer scales for the predictor variable, potential time-varying confounders, outcomes, and potential correlates of individual treatment effects.

Group	Conceptualization in Figure 1	Item	Wave	Variable name	Wording	Answer scale
Predictor variable	Hormonal contraception (at previous wave)	Contraceptive method	T_1-T_{13}	sex5	Have you used a contraceptive method in the last three months?	1 = *yes* 2 = *no*
				sex6_	Which method have you mainly used? You can choose several alternatives. - birth control pill, mini pill - condom - other hormonal method (implant, patch, ring) - intrauterine device - diaphragm, foam, suppository, gel - natural birth control - female sterilization - male sterilization - withdrawal method, coitus interruptus - morning-after-pill - something else	0 = *not mentioned* 1 = *mentioned*
Potential time-varying confounders	Demography	Age	T_1-T_{13}	age[a]	Age of participant (calculated based on dobd, dobm, doby)	Years [number] Months [number]
		Net income	T_1-T_{13}	inc2	And now we would like to ask about your net income. Net income is the amount of money after subtracting taxes and insurance costs for pension, unemployment, and health care.	Income in Euros [number]
		Educational attainment	T_1-T_{13}	yeduc[a]	Years of schooling	Years [number]

Table 3 (continued)

Item wordings and answer scales for the predictor variable, potential time-varying confounders, outcomes, and potential correlates of individual treatment effects.

Group	Conceptualization in Figure 1	Variable	Wave	Variable name	Wording	Answer scale
Potential time-varying confounders (cont.)	Relationship	Relationship status	T_1-T_{13}	hp[a]	Relationship status (single vs. non-single)	0 = *single* 1 = *non-single*
		Relationship duration	T_1-T_{13}	reldur[a]	Relationship duration (calculated based on event history calendar for romantic relationships)	Years [number] Months [number]
		Number of Children	T_1-T_{13}	nkids[a]	Number of children (calculated based on event history calendar for children)	Children [number]
			T_1, T_2	frt6	When you think about your own potential children realistically: How many (further) children will you have?	0 = *no (further) child* 1 = *one (further) child* 2 = *two (further) children* 3 = *three (further) children* 4 = *four (further) children or more* 5 = *I am unsure* 6 = *I have not thought about that*
		Completed fertility plans[c]	T_3-T_{13}	frt26	How many biological or adopted children do you think you will have? (participants without children)	1 = *one child* 2 = *two children* 3 = *three children* 4 = *four children or more* 5 = *I am unsure* 6 = *I have not thought about that* 7 = *no children*
			T_3-T_{13}	frt27	Do you think that you will have further biological or adopted children in addition to your current children or stepchildren? (participants with children)	*1* = yes *2* = *no*

Table 3 (continued)

Item wordings and answer scales for the predictor variable, potential time-varying confounders, outcomes, and potential correlates of individual treatment effects.

Group	Conceptualization in Figure 1	Variable	Wave	Variable name	Wording	Answer scale
Sexuality outcomes	Outcome (at previous wave)	Desired sexual frequency	T_7-T_{13}	sex13	If it was only up to you, would you like to have less or more sexual intercourse compared to the last three months?	1 = *a lot less* 2 = *a little bit less* 3 = *same amount* 4 = *a little bit more* 5 = *a lot more*
		Reported Sexual frequency	T_2-T_{13}	sex8[c]	How many times did you have sexual intercourse during the last three months?	0 = *never* 1 = *no sexual intercourse during the last three months* 2 = *once per month or less* 3 = *twice or three times per month* 4 = *once per week* 5 = *twice to three times per week* 6 = *more than three times per week* 7 = *daily*
		Sexual satisfaction	T_1-T_{13}	sat5	How satisfied are you with your sex life?	Scale from 0 = *very unsatisfied* to 10 = *very satisfied*

Table 3 (continued)

Item wordings and answer scales for the predictor variable, potential time-varying confounders, outcomes, and potential correlates of individual treatment effects.

Group	Conceptualization in Figure 1	Variable	Wave	Variable name	Wording	Answer scale
Well-being outcomes	Outcome (at previous wave)	Depressiveness (10 items)	T_2-T_{13}	per2i1 per2i2 per2i3 per2i4 per2i5 per2i6 per2i7 per2i8 per2i9 per2i10	- My mood is melancholy - I am happy (r) - I am depressed - I am sad - I am in desperation - My mood is gloomy - I feel good (r) - I feel secure (r) - I am calm and composed (r) - I enjoy life (r)	1 = *nearly never* 2 = *sometimes* 3 = *often* 4 = *nearly always*
		Life satisfaction	T_1-T_{13}	sat6	All in all, how satisfied are you at the moment with your life?	Scale from 0 = *very unsatisfied* to 10 = *very satisfied*
		Self-esteem (3 items)	T_1-T_{13}	per1i2 per1i7 per1i13	- Sometimes I believe that I'm worthless - I like myself just the way I am (r) - All in all, I am pleased with myself (r)	Scale from 1 = *does not match at all* to 5 = *matches fully*

Table 3 (continued)

Item wordings and answer scales for the predictor variable, potential time-varying confounders, outcomes, and potential correlates of individual treatment effects.

Group	Conceptualization in Figure 1	Variable	Wave	Variable name		Wording	Answer scale
Correlates of individual treatment effects	—	Big Five Personality Extraversion (4 items)	mean across T_2, T_6, T_{10}, and T_{11}	per3i1	-	I am usually modest and reserved (r)	Scale from 1 = *disagree strongly* to 5 = *agree strongly*
				per3i6	-	I get enthusiastic easily and can motivate others easily	
				per3i11	-	I tend to be the strong and silent type (r)	
				per3i16	-	I am extroverted	
		Big Five Personality Agreeable-ness (4 items)		per3i2	-	I tend to criticize others (r)	
				per3i7	-	I trust others easily and believe that people are inherently good	
				per3i12	-	I can be cold and distanced in my behavior (r)	
				per3i17	-	I can be rude and dismissive with others (r)	
		Big Five Personality Conscien-tiousness (4 items)		per3i3	-	I do a thorough job	
				per3i8	-	I make things comfortable for myself and tend to be lazy (r)	
				per3i13	-	I do things effectively and efficiently	
				per3i18	-	I make plans and carry them out	
		Big Five Personality Neuroticism (4 items)		per3i4	-	I easily become depressed or discouraged	
				per3i9	-	I am relaxed and can handle stress well (r)	
				per3i14	-	I worry a lot	
				per3i19	-	I easily become nervous and insecure	
		Big Five Personality Openness (5 items)		per3i5	-	I am interested in many different kinds of things	
				per3i10	-	I am intellectual and like to contemplate things	
				per3i15	-	I have an active imagination	
				per3i20	-	I value artistic, aesthetic experiences	
				per3i21	-	I am hardly interested in arts (r)	

Table 3 (continued)

Item wordings and answer scales for the predictor variable, potential time-varying confounders, outcomes, and potential correlates of individual treatment effects.

Note. The English translations of German items are provided by PAIRFAM. All sexuality outcomes are indicated with a dark-gray background. All well-being outcomes are indicated with a light-gray background.

[a]Variable is generated by PAIRFAM and included in the dataset.

[b]Variable completed fertility plans will be generated based on frt6 for wave 1 and 2 and based on frt26 and frt27 for the subsequent waves. Values correspond to 0 = *fertility plans not completed* (frt6: values 1 to 6; frt26: values 1 to 6; frt27: 1) and 1 = *fertility plans completed* (frt6: 0; frt26: 7; frt27: 2).

[c]Participants in the first wave and participants from the refreshment sample who indicated that they never had sexual intercourse before (sex1i3; 97 = *did not have sexual intercourse*) were not asked question sex8. For our analyses, they will be coded as 0 = *never had sexual intercourse* for question sex8.

Simulation

In order to contrast our different analytical approaches, we compared the performance of our models (conceptually summarized in Figure 1) on data simulated under different data generating mechanisms. For these models we assumed a true causal effect of treatment on the outcome of -.45 and a standard deviation of the treatment effect of .20 in a sample of n = 6,565 women with a maximum of 13 waves. The size of the underlying causal effect was chosen arbitrarily for the purpose of this simulation. Nevertheless, it is not completely implausible; a recent double-blind, randomized, placebo-controlled trial found effect sizes ranging from *Cohen's d* = -0.41 to -0.22 (Zethraeus et al., 2016, 2017). Sample size and attrition rates were based on actual data and were lifted from the results provided by Gonzalez Avilés after she executed code we provided; both the code used to generate those values as well as the simulation code are available as rmd file and html file: https://osf.io/u8ntf/?view_only=6d5b0a56a41541249cab38c51847157c.

We estimated 1) unadjusted models; 2) adjusted regression models; 3) IPTW models with stabilized, truncated weights at 1%; 4) adjusted regression models accounting for systematic missingness; and 5) IPTW models with stabilized, truncated weights at 1% accounting for systematic missingness. Because of the computationally intensive nature of Bayesian models we decided to perform frequentist models for all simulations.

Considering the data generating mechanisms, we varied the presence of unobserved confounding and systematic missingness. Performance was evaluated by testing whether the confidence interval of the effect size estimate for the effect of treatment on outcome included the true causal effect.

Unsurprisingly, when estimating the causal effect the adjusted and IPTW models outperformed the unadjusted models in all cases. The adjusted and IPTW models did not differ significantly in their performance, except for the simulation based on unobserved confounders and systematic missingness, in which the adjusted regression model performed significantly better. When systematic missingness was part of the data generating mechanism, the IPTW model accounting for systematic missingness performed better than

the IPTW model that did not account for systematic missingness.[5] Predictably, as soon as unobserved confounding was introduced none of the models were able to recover the true causal effect underlying the data generating process. The effect size estimates based on different models and data generating mechanisms are displayed in Figure 2.

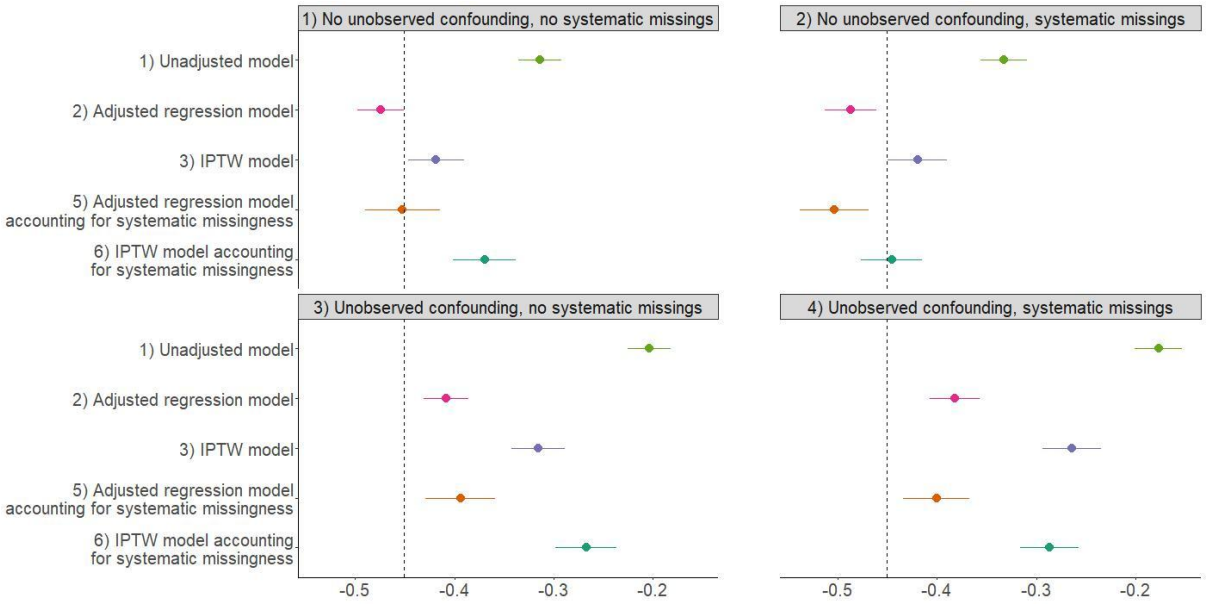
Considering the estimation of treatment effect heterogeneity, Figure 3 displays the estimates of the standard deviation of the effect based on different models and data generating mechanisms. Here, the adjusted regression model clearly outperformed all other models. The inability of models with weights (either for systematic missingness or in the IPTW approach) to recover the true variance of effects is due to a trade-off between bias and variance: by reducing bias, variance is increased resulting in higher estimates for treatment heterogeneity (see Austin, 2016).

Based on these simulation results, to estimate the average treatment effects and treatment effect heterogeneity, we will perform adjusted regression analyses without accounting for systematic missingness. In addition, we will estimate the average treatment effects based on IPTW analyses accounting for systematic missingness.

[5] In the simulation, systematic missingness was partly influenced by treatment-outcome-interaction (women using hormonal contraception, who experienced low values on the outcome score, were more likely to drop out of the simulated study). Figure 2 shows that models which accounted for systematic missingness performed worse in the absence of systematic missingness because they overadjust for systematic missingness.

Figure 2

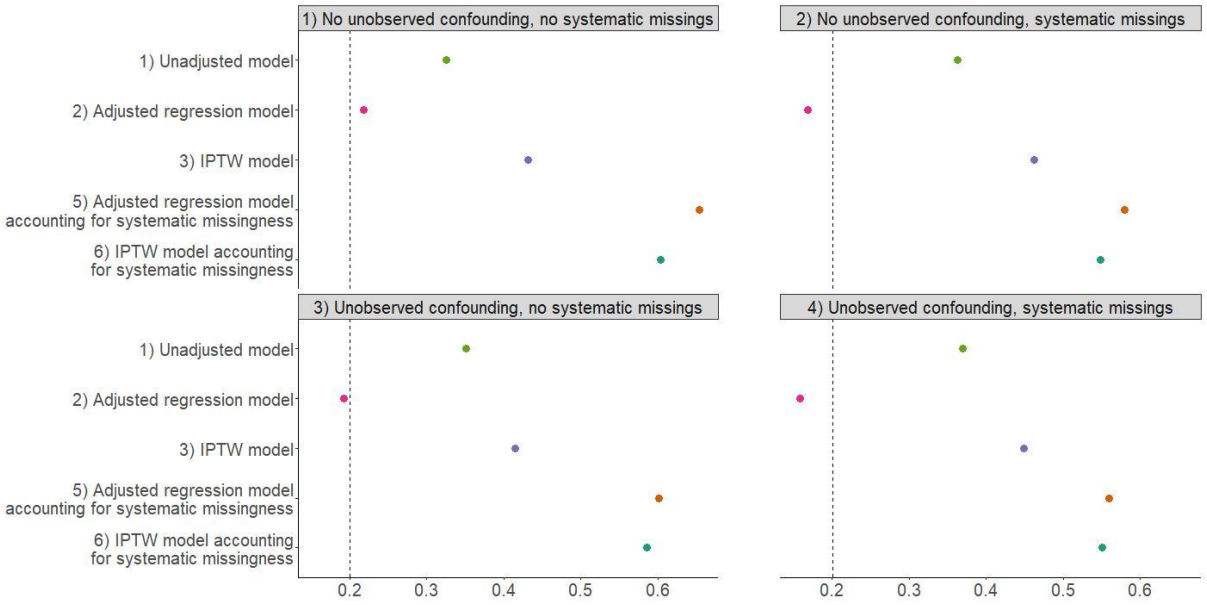
Effect size estimates for different models and data generating mechanisms based on simulated data.



Note. The true causal effect of treatment on outcome was set at -.45 (the dotted line). IPTW = Inverse probability of treatment weighting (Thoemmes & Ong, 2016) with stabilized, truncated weights at 1%.

Figure 3

Standard deviation estimates of effect sizes for different models and data generating mechanisms based on simulated data.



Note. The true standard deviation of the effect of treatment on outcome was set to .20 (the dotted line).

IPTW = Inverse probability of treatment weighting (Thoemmes & Ong, 2016) with stabilized, truncated weights at 1%.

Analysis Plan

To answer the question whether hormonal contraceptive use influences women's sexuality as well as well-being, and to separate these potential causal effects from confounders and attrition effects, we will use two different analytical approaches, as outlined in Figure 1 and described in the section Simulation. First, as outlined above, we want to use the adjusted regression approach. Second, we want to use the IPTW approach with stabilized, truncated weights at 1% (Thoemmes & Ong, 2016). All planned analyses can be found in form of an rmd file and an html file: https://osf.io/u8ntf/?view_only=6d5b0a56a41541249cab38c51847157c.

All Bayesian models will include a random intercept and a random slope for hormonal contraceptive use nested within participants. In addition, each model will include information from the previous wave about the outcome, hormonal contraception, and their interaction as predictors.[6] In order to be able to estimate the causal relationship between the hormonal contraception and the outcome, we will control for individual mean levels of hormonal contraceptive use across waves (see Bafumi & Gelman, 2007, and Hamaker & Muthen, 2020, for further information); this approach effectively controls for stable confounding influences that work between women (time-invariant confounders). For both models, potential time-varying confounders will include linear effects for log transformed net income, educational attainment, and fertility plans; a thin-plate spline effect (Wood, 2003) for age; and a categorical effect for number of children (no children, one child, two children, three or more children).

Furthermore, relationship duration will be included as a nested variable. This allows us to model a linear association with relationship duration which is only informed by women who are in a relationship, while simultaneously including those who are not in the analysis. Technically, we achieve this by including a dummy coded variable for current relationship status (single vs. non-single) and its interaction with log transformed relationship duration as a predictor. No main effect of relationship duration will be included in the model. Relationship duration for singles will be set to -1; this value is arbitrary and does not affect the resulting estimates because when multiplied with the relationship status dummy, relationship duration for singles is dropped from the analysis. In addition, we will include two dummy coded variables: one indicating whether a woman started a relationship between the previous wave and the current wave and one indicating whether a woman became single between the previous wave and the current wave.

---

[6] We decided to include the interaction term in the IPTW approach to model the possibility that certain outcomes might have stronger effects in hormonal contraceptive users than in non-hormonal contraceptive users on the contraceptive choice (e.g., strong negative side effects on sexuality and well-being might be more likely to be attributed to the contraceptive choice in hormonal contraceptive users leading to the decision to stop using this method). To keep both approaches parallel, we also included this interaction term in the adjusted regression analyses.

In the IPTW approach the outcome in the first model will be the contraceptive method. The first model results in an estimated weight which is then included in the second model. In the IPTW approach, the effects will be additionally weighted for systematic missingness based on weights provided by PAIRFAM[7]. Separate analyses for sexuality outcomes will include desired sexual frequency, reported sexual frequency, and sexual satisfaction and separate analyses for well-being outcomes will include depressiveness, general life satisfaction, and self-esteem. All included variables are listed in Table 3.

To answer the question whether interindividual differences predict individual treatment effects, we will extract individual treatment effect estimates from the adjusted regression analysis and subsequently correlate them with age (continuous) and the Big Five personality traits. These correlation analyses will be weighted by the inverse of the standard error of the individual treatment effect estimates to propagate uncertainties in their estimation.

To answer the question whether women guide their contraceptive method choices by deciding against hormonal contraceptive methods after experiencing adverse effects, we will again use individual treatment effect estimates from the adjusted regression analysis, this time correlating them with the proportion of years using hormonal contraceptives (waves in which hormonal contraceptives were used divided by total number of waves participating in PAIRFAM). This correlation analysis will again be weighted by the inverse of the standard error of the individual treatment effect estimates. This analysis can potentially provide tentative evidence for assortment based on experiences with contraceptive methods.

Additionally, given the possibility of unobserved confounding, we will run sensitivity analysis to estimate how sensitive our results are to hidden bias. We will calculate E-values for the effect of hormonal contraception on all outcomes (VanderWeele & Ding, 2017). As VanderWeele and Ding (2017) write “The E-value is defined as the minimum strength of association, on the risk ratio scale, that an unobserved confounder would need to have with

[7] We will use the calibration weights which adjust for differences between the population and the sample on the following characteristics: gender, federal state, education level, migration background, settlement structure, family status, number of children in household.

both the treatment and the outcome to fully explain away a specific treatment-outcome association, conditional on the measured covariates.“ A large E-value implies that unobserved confounding would need to be relatively substantial to explain away an effect. Conversely, a small E-value implies that even just a little unobserved confounding would be able to explain away the estimated effect. E-values are one of the few approaches to unobserved confounding that can be applied to longitudinal designs (VanderWeele et al., 2020).

Models

All planned analyses can be found in form of an rmd file and an html file: https://osf.io/u8ntf/?view_only=6d5b0a56a41541249cab38c51847157c. In addition, all models are outlined below using simplified readable notation. Code and the notation provided below use the same names for all variables. For the variable *outcome*, sexuality measures will include desired sexual frequency in the last three months, reported sexual frequency in the last three months, and sexual satisfaction. Well-being measures will include depressiveness, general life satisfaction, and self-esteem. These models are all multilevel models with a random intercept u_{0i} and in some cases a random slope for hormonal contraceptive use $u_{1i}hc_{wi}$ across waves *w* nested within participants *i*. $s(age_{wi})$ refers to a function of the spline constructed variable for *age*. Some models are weighted multilevel models indicated by the sign | followed by the name of the respective weight after the outcome variable. w_miss_i refers to a weight for systematic missingness reported by PAIRFAM. w_iptw_i refers to the weight calculated based on the first step of the IPTW approach.

Models to Gauge Confounding Effects on Hormonal Contraceptive Use

$$hc_{wi} = b_0 + b_1 * mean\_hc_i + b_2 * hc_{(w-1)i} + b_3 * outcome_{(w-1)t} + b_4 * hc_{(w-1)i} * outcome_{(w-1)t} + b_5 * s(age) + b_6 * log(net\_income_{wi}) + b_7 * years\_edu_{wi} + b_8 * rel\_stat_{wi} + b_9 * rel\_stat_{wi} * rel\_dur_{wi} + b_{10} * rel\_start_{wi} + b_{11} * single\_start_{wi} + b_{12} * n\_kids_{wi} + b_{13} * comp\_fert\_plans_{wi} + u_{0i} + \varepsilon_{wi}$$

with

$$u_{0i} \sim N(0, \sigma_u^2)$$

$$\varepsilon_{wi} \sim N(0, \sigma_\varepsilon^2)$$

Adjusted Regression Analyses

$$outcome_{wi} = b_0 + b_1 * mean\_hc_i + b_2 * hc_{wi} + b_3 * hc_{(w-1)i} + b_4 * outcome_{(w-1)t} + b_5 * hc_{(w-1)i} * outcome_{(w-1)t} + b_6 * s(age) + b_7 * log(net\_income_{wi}) + b_8 * years\_edu_{wi} + b_9 * rel\_stat_{wi} + b_{10} * rel\_stat_{wi} * rel\_dur_{wi} + b_{11} * rel\_start_{wi} + b_{12} * single\_start_{wi} + b_{13} * n\_kids_{wi} + b_{14} * comp\_fert\_plans_{wi} + u_{0i} + u_{1i} hc_{wi} + \varepsilon_{wi}$$

with

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N(0, \Omega_u)$$

$$\Omega_u = \begin{bmatrix} \sigma^2_{u_0} & \\ \sigma_{u_{01}} & \sigma^2_{u_1} \end{bmatrix}$$

$$\varepsilon_{wi} \sim N(0, \sigma_\varepsilon^2)$$

Inverse Probability of Treatment Weighting Approach

Model to Compute Weights w_iptw.

$$hc_{wi} = b_0 + b_1 * mean\_hc_i + b_2 * hc_{(w-1)i} + b_3 * outcome_{(w-1)t} + b_4 * hc_{(w-1)i} * outcome_{(w-1)t} + b_5 * s(age) + b_6 * log(net\_income_{wi}) + b_7 * years\_edu_{wi} + b_8 * rel\_stat_{wi} + b_9 * rel\_stat_{wi} * rel\_dur_{wi} + b_{10} * rel\_start_{wi} + b_{11} * single\_start_{wi} + b_{12} * n\_kids_{wi} + b_{13} * comp\_fert\_plans_{wi} + u_{0i} + \varepsilon_{wi}$$

with

$$u_i \sim N(0, \sigma_u^2)$$

$$\varepsilon_{wi} \sim N(0, \sigma_\varepsilon^2)$$

Outcome Models.

$$outcome_{wi} \mid (w\_miss_i * w\_iptw_i) = b_0 + b_1 * hc_{wi} + u_{0i} + u_{1i} hc_{wi} + \varepsilon_{wi}$$

with

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N(0, \Omega_u)$$

$$\Omega_u = \begin{bmatrix} \sigma^2_{u_0} & \\ \sigma_{u_{01}} & \sigma^2_{u_1} \end{bmatrix}$$

$$\varepsilon_{wi} \sim N(0, \sigma_\varepsilon^2)$$

References

Apter, D., Borsos, A., Baumgärtner, W., Melis, G. B., Vexiau-Robert, D., Colligs-Hakert, A., Palmer, M., & Kelly, S. (2003). Effect of an oral contraceptive containing drospirenone and ethinylestradiol on general well-being and fluid-related symptoms. *The European Journal of Contraception & Reproductive Health Care*, *8*(1), 37–51. https://doi.org/10.1080/ejc.8.1.37.51

Arslan, R. C., Schilling, K. M., Gerlach, T. M., & Penke, L. (2021). Using 26,000 diary entries to show ovulatory changes in sexual desire and behavior. *Journal of Personality and Social Psychology*, *121*(2), 410–431. https://doi.org/10.1037/pspp0000208

Austin, P. C. (2016). Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. *Statistics in Medicine*, *35*(30), 5642–5655. https://doi.org/10.1002/sim.7084

Bafumi, J., & Gelman, A. (2007). Fitting multilevel models when predictors and group effects correlate. *Available at SSRN*. http://dx.doi.org/10.2139/ssrn.1010095

Böttcher, B., Radenbach, K., Wildt, L., & Hinney, B. (2012). Hormonal contraception and depression: A survey of the present state of knowledge. Archives of Gynecology and Obstetrics, 286(1), 231–236. https://doi.org/10.1007/s00404-012-2298-2

Boozalis, M. A., Tutlam, N. T., Robbins, C. C., & Peipert, J. F. (2016). Sexual desire and hormonal contraception. *Obstetrics and Gynecology*, *127*(3), 563–572. https://doi.org/10.1097/AOG.0000000000001286

Both, S., Lew-Starowicz, M., Luria, M., Sartorius, G., Maseroli, E., Tripodi, F., Lowenstein, L., Nappi, R. E., Corona, G., Reisman, Y. & Vignozzi, L. (2019). Hormonal contraception and female sexuality: position statements from the European Society of Sexual

Medicine (ESSM). *The Journal of Sexual Medicine*, *16*(11), 1681–1695. https://doi.org/10.1016/j.jsxm.2019.08.005

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). ttps://doi.org/10.1525/collabra.29039

Brüderl, J., Drobnič, S., Hank, K., Neyer, F. J., Walper, S., Alt, P., Borschel, E., Bozoyan, C., Garrett, M., Geissler, S., Gonzalez Avilés, T., Gröpler, N., Hajek, K., Herzig, M., Huyer-May, B., Lenke, R., Lorenz, R., Lutz, K., Minkus, L.,... Wetzel, M. (2021). *The German Family Panel (pairfam)*. GESIS Data Archive, Cologne. ZA5678 Data file Version 12.0.0, https://doi.org/10.4232/pairfam.5678.12.0.0

Burrows, L. J., Basha, M., & Goldstein, A. T. (2012). The effects of hormonal contraceptives on female sexuality: A review. *The Journal of Sexual Medicine*, *9*(9), 2213–2223. https://doi.org/10.1111/j.1743-6109.2012.02848.x

Blumenstock, S. M., & Barber, J. S. (2022). Hormonal Contraception Use and Sexual Frequency across Young Women's Intimate Relationships. *The Journal of Sex Research*, 1–14. https://doi.org/10.1080/00224499.2022.2059649

Caruso, S., Agnello, C., Intelisano, G., Farina, M., Di Mari, L., Sparacino, L., & Cianci, A. (2005). Prospective study on sexual behavior of women using 30 μg ethinylestradiol and 3 mg drospirenone oral contraceptive. *Contraception*, *72*(1), 19–23. https://doi.org/10.1016/j.contraception.2005.02.002

Davis, A. R., & Castaño, P. M. (2004). Oral contraceptives and libido in women. *Annual Review of Sex Research*, *15*(1), 297–320. https://doi.org/10.1080/10532528.2004.10559822

Eng, P. M., Seeger, J. D., Loughlin, J., Clifford, C. R., Mentor Ba, S., & Walker, A. M. (2008). Supplementary data collection with case-cohort analysis to address potential confounding in a cohort study of thromboembolism in oral contraceptive initiators matched on claims-based propensity scores. *Pharmacoepidemiology and Drug Safety*, *17*(3), 297–305. https://doi.org/10.1002/pds.1554

Frye, C. A. (2006). An overview of oral contraceptives: mechanism of action and clinical use. *Neurology*, *66*, 29–36. https://doi.org/10.1212/WNL.66.66_suppl_3.S29

Fuentes, A., Lüdtke, O., & Robitzsch, A. (2021). Causal inference with multilevel data: A comparison of different propensity score weighting approaches. *Multivariate Behavioral Research*, 1–24. https://doi.org/10.1080/00273171.2021.1925521

Gaspard, U. J., Romus, M. A., Gillain, D., Duvivier, J., Demey-Ponsart, E., & Franchimont, P. (1983). Plasma hormone levels in women receiving new oral contraceptives containing ethinyl estradiol plus levonorgestrel or desogestrel. *Contraception*, *27*(6), 577–590. https://doi.org/10.1016/0010-7824(83)90023-9

Graham, C. A. (2019). The pill and women's sexuality. *BMJ*, *364*. https://doi.org/10.1136/bmj.l335

Graham, C. A., Ramos, R., Bancroft, J., Maglaya, C., & Farley, T. M. (1995). The effects of steroidal contraceptives on the well-being and sexuality of women: a double-blind, placebo controlled, two-centre study of combined and progestogen-only methods. *Contraception*, *52*(6), 363–369. https://doi.org/10.1016/0010-7824(95)00226-X

Gutman, R., & Rubin, D. B. (2017). Estimation of causal effects of binary treatments in unconfounded studies with one continuous covariate. *Statistical Methods in Medical Research*, *26*(3), 1199–1215. https://doi.org/10.1177/096228021557072

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, *25*(3), 365–379. https://doi.org/10.1037/met0000239

Hill, S. E., & Mengelkoch, S. (2022). Moving Beyond the Mean: Promising Research Pathways to Support a Precision Medicine Approach to Hormonal Contraception. *Frontiers in Neuroendocrinology*, *68*. https://doi.org/10.1016/j.yfrne.2022.101042

Huinink, J.; Brüderl, J.; Nauck, B.; Walper, S.; Castiglioni, L.; Feldhaus, M (2011). Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. *Zeitschrift für Familienforschung - Journal of Family Research*, *23*, 77–101. Retrieved from http://www.pairfam.de/fileadmin/user_upload/redakteur/publis/Dokumentation/Manuals/Concept%20Paper%20by%20Huinink%20et%20al_en%2C%20ZfF%202011.pdf

Kiesner, J. (2017). The menstrual cycle-response and developmental affective-risk model: A multilevel and integrative model of influence. *Psychological Review*, *124* (*2*), 215–244. https://doi.org/10.1037/rev0000058

King, G., & Zeng, L. (2006). The dangers of extreme counterfactuals. *Political Analysis*, *14*(2), 131–159. https://doi.org/10.1093/pan/mpj004

Lee, J. J. M., Low, L. L., & Ang, S. B. (2017). Oral contraception and female sexual dysfunction in reproductive women. *Sexual Medicine Reviews*, *5*(1), 31–44. https://doi.org/10.1016/j.sxmr.2016.06.001

Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. *American Sociological Review*, *86*(3), 532–565. https://doi.org/10.1177/00031224211004187

Lundin, C., Danielsson, K. G., Bixo, M., Moby, L., Bengtsdotter, H., Jawad, I., Marrions, L., Brynhildsen, J., Malmborg, A., Lindh, I., & Poromaa, I. S. (2017). Combined oral contraceptive use is associated with both improvement and worsening of mood in the different phases of the treatment cycle—a double-blind, placebo-controlled randomized trial. *Psychoneuroendocrinology*, *76*, 135–143. https://doi.org/10.1016/j.psyneuen.2016.11.033

Lundin, C., Malmborg, A., Slezak, J., Danielsson, K. G., Bixo, M., Bengtsdotter, H., Marions, L., Lindh, I., Theodorsson, E., Hammar, M., & Sundström-Poromaa, I. (2018). Sexual function and combined oral contraceptives: a randomised, placebo-controlled trial. *Endocrine Connections*, *7*(11), 1208--1216. https://doi.org/10.1530/EC-18-0384

Lundin, C., Wikman, A., Lampa, E., Bixo, M., Gemzell‐Danielsson, K., Wikman, P., Ljung, R., & Sundström Poromaa, I. (2022). There is no association between combined oral hormonal contraceptives and depression: a Swedish register‐based cohort study. *BJOG: An International Journal of Obstetrics & Gynaecology*, *129*(6), 917–925. https://doi.org/10.1111/1471-0528.17028

Læssøe, N. C., Wåhlin, S., Pedersen, A. T., Kristensen, E., & Giraldi, A. G. (2014). Combined hormonal contraception and women's sexual function: a cross-sectional pilot study in a cohort of Danish women. *Obstetrics and Gynecology: An International Journal*, *2014*, 1–13. https://doi.org/10.5171/2014.616630

Malmborg, A., Persson, E., Brynhildsen, J., & Hammar, M. (2016). Hormonal contraception and sexual desire: A questionnaire-based study of young Swedish women. *The European Journal of Contraception & Reproductive Health Care*, *21*(2), 158–167. https://doi.org/10.3109/13625187.2015.1079609

McCoy, N. L., & Matyas, J. R. (1996). Oral contraceptives and sexuality in university women. *Archives of Sexual Behavior*, *25*(1). https://doi.org/10.1007/BF02437907

Mishell, D. R., Thorneycroft, I. H., Nakamura, R. M., Nagata, Y., & Stone, S. C. (1972). Serum estradiol in women ingesting combination oral contraceptive steroids. *American Journal of Obstetrics and Gynecology*, *114*(7), 923–928. https://doi.org/10.1016/0002-9378(72)90098-1

Oranratanaphan, S., & Taneepanichskul, S. (2006). A double blind randomized control trial, comparing effect of drospirenone and gestodene to sexual desire and libido. *Journal of the Medical Association of Thailand*, *89*(4), 17–22. Retrieved from https://pdfs.semanticscholar.org/f426/e6cac8a255632df21482b25e09a7a6350876.pdf

Ott, M. A., Shew, M. L., Ofner, S., Tu, W., & Fortenberry, J. D. (2008). The influence of hormonal contraception on mood and sexual interest among adolescents. *Archives of Sexual Behavior*, *37*(4), 605–613. https://doi.org/10.1007/s10508-007-9302-0

Pastor, Z., Holla, K., & Chmel, R. (2013). The influence of combined oral contraceptives on female sexual desire: a systematic review. *The European Journal of Contraception & Reproductive Health Care*, *18*(1), 27–43. https://doi.org/10.3109/13625187.2012.728643

Riggs, M., Klebanoff, M., Nansel, T., Zhang, J., Schwebke, J., & Andrews, W. (2007). Longitudinal association between hormonal contraceptives and bacterial vaginosis in women of reproductive age. *Sexually Transmitted Diseases*, *34*(12), 954–959. https://doi.org/10.1097/OLQ.0b013e31811ed0e4

Rohrer, J. M., & Lucas, R. E. (2020). Causal effects of well-being on health: It's complicated. *Preprint on PsyArXiv*. https://doi.org/10.31234/osf.io/wgbe4

Rohrer, J. M., & Murayama, K. (2021). These are not the effects you are looking for: Causality and the within-/between-person distinction in longitudinal data analysis. *Preprint on PsyArXiv*. https://doi.org/10.31234/osf.io/tg4vj

Rosenbaum, P. R., & Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, *45*(2), 212–218. https://doi.org/10.1111/j.2517-6161.1983.tb01242.x

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*(3), 169–188. https://doi.org/10.1023/A:1020363010465

Sabatini, R., & Cagiano, R. (2006). Comparison profiles of cycle control, side effects and sexual satisfaction of three hormonal contraceptives. *Contraception*, *74*(3), 220–223. https://doi.org/10.1016/j.contraception.2006.03.022

Schaffir, J. (2006). Hormonal contraception and sexual desire: a critical review. *Journal of Sex & Marital Therapy*, *32*(4), 305–314. https://doi.org/10.1080/00926230600666311

Schleifenbaum, L., Driebe, J. C., Gerlach, T. M., Penke, L., & Arslan, R. C. (2021). Women feel more attractive before ovulation: evidence from a large-scale online diary study. *Evolutionary Human Sciences*, *3*. https://doi.org/10.1017/ehs.2021.44

Skovlund, C. W., Mørch, L. S., Kessing, L. V., Lange, T., & Lidegaard, Ø. (2018). Association of hormonal contraception with suicide attempts and suicides. *American Journal of Psychiatry*, *175*(4), 336–342. https://doi.org/10.1176/appi.ajp.2017.17060616

Skovlund, C. W., Mørch, L. S., Kessing, L. V., & Lidegaard, Ø. (2016). Association of hormonal contraception with depression. *JAMA Psychiatry*, *73*(11), 1154–1162. https://doi.org/10.1001/jamapsychiatry.2016.2387

Smith, N. K., Jozkowski, K. N., & Sanders, S. A. (2014). Hormonal contraception and female pain, orgasm and sexual pleasure. *The Journal of Sexual Medicine*, *11*(2), 462–470. https://doi.org/10.1111/jsm.12409

Stomati, M., Genazzani, A. D., Petraglia, F., & Genazzani, A. R. (1998). Contraception as prevention and therapy: sex steroids and the brain. *The European Journal of Contraception & Reproductive Health Care*, *3*(1), 21–28. https://doi.org/10.3109/13625189809167481

Strufaldi, R., Pompei, L. M., Steiner, M. L., Cunha, E. P., Ferreira, J. A., Peixoto, S., & Fernandes, C. E. (2010). Effects of two combined hormonal contraceptives with the same composition and different doses on female sexual function and plasma androgen levels. *Contraception*, *82*(2), 147–154. https://doi.org/10.1016/j.contraception.2010.02.016

Taggart, T. C., Eaton, N. R., Keyes, K. M., Hammett, J. F., & Ulloa, E. C. (2018). Oral contraceptive use is associated with greater mood stability and higher relationship

satisfaction. *Neurology, Psychiatry and Brain Research*, *30*, 154–1162. https://doi.org/10.1016/j.npbr.2018.10.004

Thoemmes, F., & Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, *4*(1), 40–59. https://doi.org/10.1177/2167696815621645

Toffol, E., Heikinheimo, O., Koponen, P., Luoto, R., & Partonen, T. A. (2011). Hormonal contraception and mental health: Results of a population-based study. *Human Reproduction*, *26*(11), 3085–3093. https://doi.org/10.1093/humrep/der269

Toffol, E., Heikinheimo, O., Koponen, P., Luoto, R., & Partonen, T. (2012). Further evidence for lack of negative associations between hormonal contraception and mental health. *Contraception*, *86*(5), 470–480. https://doi.org/10.1016/j.contraception.2012.02.014

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine*, *167*(4), 268–274. https://doi.org/10.7326/M16-2607

VanderWeele, T. J., Mathur, M. B., & Chen, Y. (2020). Outcome-wide longitudinal designs for causal inference: a new template for empirical studies. *Statistical Science*, *35*(3), 437–466. https://doi.org/10.1214/19-STS728

Wallwiener, C. W., Wallwiener, L. M., Seeger, H., Schönfisch, B., Mueck, A. O., Bitzer, J., Zipfel, S., Brucker, S. Y., Taran, F.-A., & Wallwiener, M. (2015). Are hormonal components of oral contraceptives associated with impaired female sexual function? A questionnaire-based online survey of medical students in Germany, Austria, and Switzerland. *Archives of Gynecology and Obstetrics*, *292*(4), 883–890. https://doi.org/10.1007/s00404-015-3726-x

Wallwiener, M., Wallwiener, L. M., Seeger, H., Mueck, A. O., Zipfel, S., Bitzer, J., & Wallwiener, C. W. (2010). Effects of sex hormones in oral contraceptives on the female sexual function score: a study in German female medical students. *Contraception*, *82*(2), 155–159. https://doi.org/10.1016/j.contraception.2009.12.022

Wang, Q., Würtz, P., Auro, K., Morin-Papunen, L., Kangas, A. J., Soininen, P., Tiainen, M., Tynkkynen, T., Joensuu, A., Havulinna, A. S., Aalto, K., Salmi, M., Blankenberg, S., Zeller, T., Viikari, J., Kähönen, M., Lehtimäki, T., Salomaa, V., Jalkanen, S.,... Ala-Korpela, M. (2016). Effects of hormonal contraception on systemic metabolism: cross-sectional and longitudinal evidence. *International Journal of Epidemiology*, *45*(5), 1445–1457. https://doi.org/10.1093/ije/dyw147

Watkins, E. S. (2012). How the pill became a lifestyle drug: the pharmaceutical industry and birth control in the United States since 1960. *American Journal of Public Health*, *102*(8), 1462–1472. https://doi.org/10.2105/AJPH.2012.300706

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Zethraeus, N., Dreber, A., Ranehill, E., Blomberg, L., Labrie, F., von Schoultz, B., Johannesson, M., & Hirschberg, A. L. (2016). Combined oral contraceptives and sexual function in women - a double-blind, randomized, placebo-controlled trial. *The Journal of Clinical Endocrinology & Metabolism*, *101*(11), 4046–4053. https://doi.org/10.1210/jc.2016-2032

Zethraeus, N., Dreber, A., Ranehill, E., Blomberg, L., Labrie, F., von Schoultz, B., Johannesson, M., & Hirschberg, A. L. (2017). A first-choice combined oral

contraceptive influences general well-being in healthy women: a double-blind, randomized, placebo-controlled trial. *Fertility and Sterility*, *107*(5), 1238–1245. https://doi.org/10.1016/j.fertnstert.2017.02.120

Zimmerman, Y., Eijkemans, M. J. C., Coelingh Bennink, H. J. T., Blankenstein, M. A., & Fauser, B. C. J. M. (2014). The effect of combined oral contraception on testosterone levels in healthy women: a systematic review and meta-analysis. *Human Reproduction Update*, *20*(1), 76–105. https://doi.org/10.1093/humupd/dmt038